Dynamical Systems and Ergodic Theory

Lecture notes by Prof. Corinna Ulcigrai

2.1 Ergodic Theory

In this short course will introduce the main ideas and concepts in *ergodic theory*. Ergodic theory is a branch of dynamical systems which has strict connections with analysis and probability theory. The discrete dynamical systems $f : X \to X$ studied in topological dynamics were continuous maps f on metric spaces X (or more in general, topological spaces). In ergodic theory, $f : X \to X$ will be a *measure-preserving* map on a *measure space* X (we will see the corresponding definitions below). While the focus in topological dynamics was to understand the qualitative behavior (for example, periodicity or density) of *all* orbits, in ergodic theory we will not study *all* orbits, but only *typical*[1] *orbits*, but will investigate more *quantitative* dynamical properties, as frequencies of visits, equidistribution and mixing.

An example of a basic question studied in ergodic theory is the following. Let $A \subset X$ be a subset of the space X. Consider the visits of an orbit $\mathcal{O}_f^+(x)$ to the set A. If we consider a finite orbit segment $\{x, f(x), \dots, f^{n-1}(x)\}$, the number of visits to A up to time n is given by

$$Card\ \{\ 0 \le k \le n-1, \qquad f^k(x) \in A\ \}. \tag{2.1}$$

A convenient way to write this quantity is the following. Let χ_A be the characteristic function of the set A, that is a function $\chi_A : X \to \mathbb{R}$ given by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Consider the following sum along the orbit

$$\sum_{k=0}^{n-1} \chi_A(f^k(x)). \tag{2.2}$$

This sum gives exactly the number (2.1) of visits to A up to time n. This is because $\chi_A(f^k(x)) = 1$ if and only if $f^k(x) \in A$ and it is zero otherwise, so that there are as many ones in the sum in (2.2) than visits up to time n and summing them all up one gets the total number of visits up to time n.

If we divide the number of visits up to time n by the time n, we get the *frequency* of visits up to time n, that is

$$\frac{Card\{0 \le k < n, \quad \text{such that } f^k(x) \in X\}}{n} = \frac{1}{n}\sum_{k=0}^{n-1} \chi_A(f^k(x)).$$

The frequency is a number between 0 and 1.

Q1 Does the frequency of visits converge to a limit as n tends to infinity? (for all points? for a typical point?)

Q2 If the limit exists, what does the frequency tend to?

A useful notion to consider for dynamical systems on the circle (or on the unit interval) is that of uniform distribution.

Definition 2.1.1. *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ (or $\mathbb{R}/\mathbb{Z}$). We say that $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed if for all intervals $I \subset [0,1]$ we have that*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \chi_I(x_k) = |I|.$$

Where $|I|$ denotes the length of the interval $|I|$. An equivalent definition is for all continuous functions $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ we should have

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x_k) = \int_0^1 f(x)dx.$$

So if we have a dynamical system $T : [0,1] \to [0,1]$ (or $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$) we can ask whether orbits $\{x, T(x), T^2(x), \dots\}$ are uniformly distributed or not.

[1] Typical will become precise when we introduce measures: by typical orbit we mean the orbit of *almost every* point, that is all orbits of points in a set of full measure.

Example 2.1.1. Let $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the rotation given by $R_\alpha(x) = x + \alpha$. If $\alpha \in \mathbb{Q}$ then for all $x \in \mathbb{R}/\mathbb{Z}$ the orbit of R_α will be periodic, so cannot be dense and thus cannot be uniformly distributed (why?). On the otherhand if $\alpha \notin \mathbb{Q}$ it will turn out for all $x \in \mathbb{R}/\mathbb{Z}$ the orbit of R_α will be uniformly distributed (this is often thought of as the 1st ergodic theorem to have been proved, it was proved independently in 1909 and 1910 by Bohl, Sierpiński and Weyl.)

A more complicated example is the following

Example 2.1.2. Let $T : [0, 1) \to [0, 1)$ be the doubling map given by $T(x) = 2x \mod 1$. We know that there is a dense set of x for which the orbit of T will be periodic and hence not uniformly distributed. However it will turn out the for 'almost all' x the orbit of T will be uniformly distributed (where almost all can be thought of as meaning except for a set of length 0.

We may also have maps $T : [0, 1) \to [0, 1)$ where for 'typical' x orbits are not equidistributed

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_A(T^i(x)) = \int_A f(x)dx$$

for some suitable function f (we will see that the Gauss map is an example of such a map). To make these notions precise we need to introduce some measure theory which will have the additional advantage of introducing a theory of integration which is more suited to our purposes.

2.2 Measures and Measure Spaces

Intuitively, a *measure* μ on a space X is a function from a collection of subsets of X, called *measurable sets*, which assigns to each measurable set A its measure $\mu(A)$, that is a positive number (possibly infinity). You already know at least two natural examples of measures.

Example 2.2.1. Let $X = \mathbb{R}$. The 1-dimensional Lebesgue measure λ on $\mathbb{R}$ assigns to each interval $[a, b] \in \mathbb{R}$ its *length*:

$$\lambda([a, b]) = b - a, \qquad a, b \in \mathbb{R}.$$

Let $X = \mathbb{R}^2$. The 2-dimensional Lebesgue measure, that we will still call λ, assigns to each measurable set[2] $A \subset \mathbb{R}^2$ its *area*, which is given by the integral[3]

$$\lambda(A) = \text{Area}(A) = \int_A \mathrm{d}x\mathrm{d}y.$$

Measurable spaces

One might hope to assign a measure to *all* subsets of X. Unfortunately, if we want the measure to have the reasonable and useful properties of a measure (listed in the definition of measure below), this leads to a contradiction (see Extra if you are curious). So, we are forced to assign a measure only to a sub-collection all subsets of X. We ask that the collection of measurable subsets is closed under the operation of taking countable unions in the following sense.

Definition 2.2.1. A collection $\mathscr{A}$ of subsets of a space X is called an *algebra* of subsets if

(i) The empty set $\emptyset \in \mathscr{A}$;

(ii) $\mathscr{A}$ is closed under complements, that is if $A \in \mathscr{A}$, then its complement $A^c = X \backslash A$ also belongs to $\mathscr{A}$;

(iii) $\mathscr{A}$ is closed under *finite* unions, that is if $A_1, \ldots, A_n \in \mathscr{A}$, then

$$\bigcup_{i=1}^{n} A_i \in \mathscr{A}.$$

[2] We will precisely define what are the measurable sets for the Lebesgue measure in what follows.

[3] If A is such that χ_A is integrable in the sense of Riemann, this integral is the usual Riemannian integral. More in general, we will need the notion of Lebesgue integral, which we will introduce in the following lectures.

Example 2.2.2. If $X = \mathbb{R}$ an example of algebra is given by the collection $\mathscr{A}$ of all possible finite unions of subintervals of $\mathbb{R}$.

Exercise 2.2.1. Check that the collection $\mathscr{A}$ of all possible finite unions of subintervals of $\mathbb{R}$ is an algebra.

Definition 2.2.2. A collection $\mathscr{A}$ of subsets of a space X is called a $\sigma-$*algebra* of subsets if

(i) The empty set $\emptyset \in \mathscr{A}$;

(ii) $\mathscr{A}$ is closed under complements, that is if $A \in \mathscr{A}$, then its complement $A^c = X \backslash A$ also belongs to $\mathscr{A}$;

(iii) $\mathscr{A}$ is closed under *countable* unions, that is if $\{A_n, \quad n \in \mathbb{N}\} \subset \mathscr{A}$, then

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}.$$

Thus, a $\sigma-$algebra is an algebra which in addition is closed under the operation of taking countable unions. The easiest way to define a $\sigma-$algebra is to start from any collection of sets, and take the closure under the operation of taking complements and countable unions:

Definition 2.2.3. If $\mathcal{S}$ is a collection of subsets, we denote by $\mathscr{A}(\mathcal{S})$ the smallest $\sigma-$algebra which contains $\mathcal{S}$. The *smallest* means that if $\mathscr{B}$ is another $\sigma-$algebra which contains $\mathcal{S}$, then $\mathscr{A}(\mathcal{S}) \subset \mathscr{B}$. We say that $\mathscr{A}(\mathcal{S})$ is the $\sigma-$algebra *generated by* $\mathcal{S}$.

The following example/definition is the main example that we will consider.

Definition 2.2.4. If (X, d) is a metric space[4], the *Borel* $\sigma-$*algebra* $\mathscr{B}(X)$ (or simply $\mathscr{B}$) is the smallest $\sigma-$algebra which contains all *open sets* of X. The subsets $B \in \mathscr{B}(X)$ are called *Borel sets*.

Borel σ-algebras are the natural collections of subsets to take as measurable sets. In virtually all of our examples, the measurable sets will be Borel subsets.

Definition 2.2.5. A *measurable space* $(X, \mathscr{A})$ is a space X together with a $\sigma-$algebra $\mathscr{A}$ of sets. The sets in $\mathscr{A}$ are called *measurable sets* and $\mathscr{A}$ is called the $\sigma-$algebra of measurable sets.

Example 2.2.3. If (X, d) is a metric space, $(X, \mathscr{B}(X))$ is a measurable space, where $\mathscr{B}(X)$ is the Borel $\sigma-$algebra.

Measures

We can now give the formal definition of measure.

Definition 2.2.6. Let $(X, \mathscr{A})$ be a measurable space. A measure μ is a function $\mu : \mathscr{A} \to \mathbb{R}^+ \cup \{+\infty\}$ such that

(i) $\mu(\emptyset) = 0$;

(ii) If $\{A_n, \quad n \in \mathbb{N}\} \subset \mathscr{A}$ is a countable collection of *pairwise disjoint* measurable subsets, that is if $A_n \cap A_m = \emptyset$ for all $n \neq m$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{2.3}$$

We say that the measure μ is *finite* if $\mu(X) < \infty$.

Remark that to have (2.3) we need to assume that A_n are *disjoint*.

Example 2.2.4. You can check that both length and area have this natural property: for example the area of the union of disjoint sets is the sum of the areas. If $X = \mathbb{R}$, the Lebesgue measure on $\mathbb{R}$ is *not* finite, since $\lambda(\mathbb{R}) = +\infty$. On the contrary, the Lebesgue measure restricted to an interval $X = [a, b] \subset \mathbb{R}$ is finite since $\lambda([a, b]) < \infty$. Similarly, if $\mathbb{T}^2$ is the torus and we consider the area λ, $Area(\mathbb{T}^2) = \lambda(\mathbb{T}^2) = 1 < \infty$, so λ is a finite measure on $\mathbb{T}^2$.

Definition 2.2.7. A *measure space* $(X, \mathscr{A}, \mu)$ is a measurable space $(X, \mathscr{A})$ and a measure $\mu : \mathscr{A} \to \mathbb{R}^+ \cup \{+\infty\}$. If $\mu(X) = 1$, we say that $(X, \mathscr{A}, \mu)$ is a *probability space*.

[4] The same definition of Borel $\sigma-$algebra holds more in general if X is a topological space, so that we know what are the open sets.

If we just work directly from the definition of a measure it is hard to produce examples of measures. One simple example is the following (as well as being simple it also turns out to be extremely useful).

Example 2.2.5. Let X a space and $x \in X$ a point. In this example we can take $\mathscr{A}$ to be the collection of all subsets of X. The measure δ_x, called *Dirac measure at* x, is defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Thus, the measure δ_x takes only two values, 0 and 1, and assigns measure 1 only to the sets which contain the point x.

It is also straight forward to see that if $(X, \mathscr{A})$ is a measurable space and μ_1, μ_2 are measures on $(X, \mathscr{A})$ them $\mu_1, \mu_2 : \mathscr{A} \to \mathbb{R}^+ \cup \{+\infty\}$ given by

$$\mu_1 + \mu_2(A) = \mu_1(A) + \mu_2(A) \text{ for all } A \in \mathscr{A}$$

is also a measure.

In very few examples (like the Dirac measure) it is possible to define a measure by explicitly saying which values it assigns to *all* measurable sets. The following theorem shows that it is not necessary to do this and one can define the measure only on a smaller collection of sets.

Theorem 2.2.1. *[Carathéodory Extension Theorem] Let $\mathscr{A}$ be an* algebra *of subsets of X. If $\mu^* : \mathscr{A} \to \mathbb{R}^+$ satisfies*

(i) $\mu^*(\emptyset) = 0$; $\mu^*(X) < \infty$;

(ii) If $\{A_n, \quad n \in \mathbb{N}\} \subset \mathscr{A}$ is a countable collection of pairwise disjoiint sets in the algebra $\mathscr{A}$ and

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{A} \qquad \Rightarrow \qquad \mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

Then there exists a unique *measure $\mu : \mathscr{B}(\mathscr{A}) \to \mathbb{R}^+$ on the $\sigma-$algebra $\mathscr{B}(\mathscr{A})$ generated by $\mathscr{A}$ which extends μ^* (in the sense that it has the prescribed values on the sets of $\mathscr{A}$). We will refer to μ^* as a premeasure.*

Remark that since $\mathscr{A}$ is only an algebra and not a $\sigma-$algebra, $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$ does not have to belong to $\mathscr{A}$. Thus, (ii) has to hold only for collections of sets A_n for which the countable union happens to still belong to $\mathscr{A}$. Thus, the theorem states that if we have a function μ^* that *behaves like* a measure on an algebra, it can indeed be extended (and uniquely) to a measure.

This theorem will be used mostly in the following two ways:

1. To define a measure on the $\sigma-$algebra $\mathscr{B}(\mathcal{S})$ generated by $\mathcal{S}$, it is enough to define μ on $\mathcal{S}$ in such a way that it satisfies the assumptions of the Theorem on the *algebra* generated by $\mathcal{S}$. This automatically defines a measure on the whole $\sigma-$algebra $\mathscr{B}(\mathcal{S})$.

2. If we have two measures μ, ν and we want to show that $\mu = \nu$, it is enough to check that $\mu(A) = \nu(A)$ for all $A \in \mathscr{A}$ where $\mathscr{A}$ is an algebra that generates the $\sigma-$algebra of all measurable sets. Then, by the uniqueness part of the Theorem, the measures μ and ν are the same measure.

We can now formally define the following measures

Example 2.2.6. Let $a, b \in \mathbb{R}$ with $a < b$ consider the interval $[a, b] \subset \mathbb{R}$. We can define Lebesgue meausre λ on intervals $(c, d) \subset [a, b]$, by setting $\lambda((c, d)) = d - c$. This clearly defines by additivity also a premeasure λ on the algebra consisting of finite unions of intervals. If $A = \cup_{i=1}^{n}(a_i, b_i)$ and (a_i, b_i) are *disjoint* intervals, we just define

$$\mu(A) = \sum_{i=1}^{n} b_i - a_i.$$

Since the condition (*ii*) of the theorem holds, this automatically defines the Lebesgue measure on the $\sigma-$algebra generated by all intervals, that is on all Borel subsets of $[a, b]$. The same method works to define Lebesgue measure on the whole of $\mathbb{R}$ however as stated the Carathéodory Extension Theorem only holds when the premeasure is finite. However in fact it holds with a slightly weaker assumption (σ-finiteness) which allows us to define Lebesgue measure on the whole of $\mathbb{R}$, see remark 2.2.2.

Example 2.2.7. Let $X = \mathbb{T}^2$. Consider sets of the form $[a, b] \times [c, d]$, that we call rectangles. Define a measure λ by setting

$$\lambda([a, b] \times [c, d]) = (b - a)(d - c).$$

The collection of all finite unions of rectangles is an algebra. Extending the definition of λ to union of rectangles by additivity, condition (ii) of the theorem automatically holds. Thus, the Theorem guarantees that we defined a Lebesgue measure on the $\sigma-$algebra generated by all rectangles, which coincides with all Borel subsets. This is again the $2-$dimensional Lebesgue measure λ on $\mathbb{T}^2$.

Example 2.2.8. If we have a non-negative Riemann integrable (shortly we will extend this to Lebesgue integrable) function $f : \mathbb{R} \to \mathbb{R}^+$ then for any subinterval $A \subset \mathbb{R}$ we can define

$$\mu_f(A) = \int_A f(x)\mathrm{dx}.$$

Now if we consider a disjoint finite union of subintervals $A_1, \ldots, A_n \subset \mathbb{R}$ we can write

$$\mu_f\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu_f(A_i).$$

Now any finite union of subintervals can be rewritten as a disjoint finite union of subintervals, the set of finite unions of subintervals forms an algebra and μ_f satisfies the conditions to apply Thoerem 2.2.1. Thus we can extend μ_f to a measure on $(\mathbb{R}, \mathscr{B})$, since the σ-algebra generated by our algebra is the Borel σ-algebra.

Extras: Remarks

Remark 2.2.1. Condition (iii) in the definition of algebra, that is

(iii) $\mathscr{A}$ is closed under finite unions, that is if $A_1, \ldots, A_n \in \mathscr{A}$, then

$$\bigcup_{i=1}^n A_i \in \mathscr{A};$$

can be equivalently replaced by the following condition

(iii)' $\mathscr{A}$ is closed under intersections, that is if $A, B \in \mathscr{A}$, then $A \cap B \in \mathscr{A}$;

In some books, the definition of algebra is given using $(i), (ii), (iii)'$.

Exercise 2.2.2. Show that a set satisfies conditions $(i), (ii), (iii)$ if and only if it satisfies $(i), (ii), (iii)'$.

Remark 2.2.2. The condition $\mu^*(X) < \infty$ in the Extension theorem can be relaxed. It is enough that $X = \cup_n X_n$ where each X_n is such that $\mu^*(X_n) < \infty$. We say in this case that the resulting measure μ is $\sigma-$finite. For example, the Lebesgue measure on $\mathbb{R}$ is $\sigma-$finite since

$$\mathbb{R} = \cup_n[-n, n], \quad \text{and} \quad \lambda([-n, n]) = 2n < \infty.$$

Extras: necessity of restricting the class of measurable sets

One might wonder why we need to use $\sigma-$algebras of measurable sets in the definition of measure and why we cannot ask that a measure is defined on the *whole* collection of subsets of X. Let $X = \mathbb{R}^n$ and consider the Lebesgue measure λ. It is natural to ask that the Lebesgue measure, that intuitively represents the concept of length, or area or volume ..., has the following properties:

(i) λ has the countable additivity property in the Definition of measure, that is if $A_1, \ldots, A_n, \ldots$ are disjoint subsets of X for which λ is defined, then

$$\lambda\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \lambda(A_n);$$

(ii) If two sets $A, B \subset \mathbb{R}^n$ are congruent, in the sense that A and B are mapped to each other by translations, rotations or reflections, then they should have the same length (or area or volume...). In particular, if $A = B + \underline{c}$, where $\underline{c} \in \mathbb{R}^n$, then $\lambda(A) = \lambda(B)$;

(iii) $\lambda([0,1)^n) = 1$ (this is simply a renormalization requirement).

Let us show that unfortunately these three properties *and* the requirement that λ is defined on *all* subsets of X are incompatible.

For simplicity, let us take $n = 1$ and $X = \mathbb{R}$. Consider $[0,1) \subset \mathbb{R}$ and let α be an irrational. Consider the rotation $R_\alpha : [0,1) \to [0,1)$ and consider all the orbits $\mathscr{O}^+_{R_\alpha}(x)$ of the rotation R_α. Let us pick[5] a representative x for each orbit and so that, if $\mathcal{R}$ is the set of representative, we can write the whole interval as union over all the orbits of the representatives:

$$[0,1] = \bigcup_{x \in \mathcal{R}} \mathcal{O}^+_{R_\alpha}(x).$$

Since $\mathcal{O}^+_{R_\alpha}(x) = \bigcup_{n \in \mathbb{N}} \{R^n_\alpha(x)\}$, we can rewrite

$$[0,1] = \bigcup_{x \in \mathcal{R}} \bigcup_{n \in \mathbb{N}} \{R^n_\alpha(x)\} = \bigcup_{n \in \mathbb{N}} \bigcup_{x \in \mathcal{R}} \{R^n_\alpha(x)\} = \bigcup_{n \in \mathbb{N}} A_n \qquad \text{where} \quad A_n = \bigcup_{x \in \mathcal{R}} \{R^n_\alpha(x)\}. \tag{2.4}$$

Let assume that λ is defined on all subsets of X and has the Properties $(i), (ii)$ and (iii) above. In particular, $\lambda(A_n)$ is defined for each $n \in \mathbb{N}$. The sets A_n are such that $A_{n+1} = R_\alpha(A_n)$, they are all obtained from each other by translations, so by Property (i)

$$\lambda(A_n) = \lambda(A_m), \qquad \text{for all } n, n \in \mathbb{N}.$$

By Property (iii), $\lambda([0,1]) = 1$. Moreover, $\{A_n\}_{n \in \mathbb{N}}$ are clearly disjoint sets, so by Property (ii), we have

$$\sum_{n \in \mathbb{N}} \lambda(A_n) = \lambda\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lambda([0,1]) = 1.$$

But since $\lambda(A_n) = \lambda(A_m)$ for all m, n, if $\mu(A_n) > 0$, this gives a contradiction (since a series with equal positive terms diverges), but if $\mu(A_n) = 0$, this also gives a contradiction (since a series with all terms equal to zero is zero). This shows that requiring all these three condition *and* asking that λ is defined on all subsets of $\mathbb{R}$ gives a contradiction.

This problem is solved when we consider the Lebesgue measure λ only on the collection of Borel subsets. It turns out that the sets A_n which lead us to a contradiction are *not measurable* (they do not belong to the $\sigma-$algebra $\mathscr{B}$ generated by open sets). Thus, $\lambda(A_n)$ is simply *not defined*.

One could consider weakening the countable additivity condition, even if countable additivity is very useful for the theory of limits and continuity. Nevertheless, even if one substitutes Property (i) with finite additivity

(i)' λ has the *finite* additivity property in the Definition of measure, that is if $A_1, \dots, A_n$ are disjoint subsets of X for which λ is defined, then

$$\lambda\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \lambda(A_k);$$

and asks that a measure is defined on *all* subsets *and* has Properties $(i)', (ii), (iii)$ one still gets paradoxical results for $n \geq 3$, as was by Banach and Tarski in 1924 in the famous:

Theorem 2.2.2 (Banach-Tarski paradox)**.** *Let U, V be any two bounded open sets in $\mathbb{R}^n$, $n \geq 3$. One can decompose each of them in finitely many disjoint pieces*

$$U = \bigcup_{k=1,\dots,n} A_k, \qquad V = \bigcup_{k=1,\dots,n} B_k,$$

such that each A_k is congruent to each B_k.

For example, if we take U to be a sphere of volume one and V to be a sphere of volume two, one can cut up the smaller sphere in finitely many pieces, move them by translations and reflections and recompose the bigger sphere. If each of these pieces had a well defined volume, we would get a contradiction: each pair of congruent pieces has the same volume, but the volume of their union has doubled! In conclusion, it is better to give up the hope that all subsets are measurable and accept that there are non measurable subsets.

[5]To pick a representative for each orbit, we are implicitly using the *Axiom of choice*.