Lecture notes for Tsingua Workshop

Alex Blumenthal

References for this course:

- Lai-Sang Young's lecture notes: "Ergodic theory of differentiable dynamical systems". Chapter in book *Real and Complex Dynamical Systems*, pages 293-336.
- Brin & Stuck: "Introduction to dynamical systems", book.

1 Lecture 1: Motivation, examples

1.1 Thought experiment: a 'typical' experimental setup

Let us write X for the phase space of our experiment, typically a high-dimensional (or sometimes infinite-dimensional) manifold. Starting at an initial condition $x_0 \in X$, we will write $x_n \in X$ for the state of our system at discrete times $n = 1, 2, \cdots$.

For many systems, especially those for which the phase space is extremely high-dimensional, the only way to 'probe' phase space is to measure the state $\phi(x_n)$ of an observable $\phi : X \to \mathbb{R}$ at time n, e.g., temperature, or the pressure exerted on a wall, or the concentration of a tracer chemical at a point.

Imagine that you run the system and take measurements of the observable ϕ at each integer timestep, resulting in collecting the time series data $\{\phi(x_n)\}_{n\geq 0}$. We are interested describing the time-asymptotic behavior of the system, i.e., what happens to the system for large times n.

One scenario is as follows: the time series $\phi(x_n)$ settles down to a single fixed value, or settles down into a finite cycle of repeating values. This hints that the dynamics (x_n) is settling down into an asymptotically stable equilibrium or periodic orbit. If you can describe this equilibrium or periodic orbit, then you have more-or-less described all asymptotic dynamics of the system.

Another scenario is that the time series $\phi(x_n)$ never settles down to a fixed value or a periodic cycle, and instead fluctuates indefinitely for all large times. How can you describe the asymptotic dynamics in this case? For systems with such 'complicated' time series, physicists often make the following assumptions:

(1) The asymptotic or 'equilibrium' state of the system can be described *statistically* by a probability measure μ on the phase space X. That is, if you let the system run for a long time, then the state of the system can be thought of as an X-valued random variable with empirical law μ.

(2) We can obtain a description of the statistic μ by estimating the expected values $\int \phi(x) d\mu(x)$. These can be estimated from time series by the limit

$$\int \phi d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \phi(x_n) \tag{1}$$

The limiting value on the LHS does not depend on the initial condition $x_0 \in X$.

In (1), the RHS is an asymptotic *time average*, while the LHS is a *spatial average* over the phase space X. The formula (1) 'exchanges' time and space averages.

Implicit in (1) and (2) above is the idea that there is some 'randomizing' mechanism in our system. In (1), no matter where we initiate the system, after a long time the system has 'forgotten' where it started and is distributed according to the law μ. That is, the initial state x_0 and the future state $x_N, N \gg 1$ have *de-correlated* from each other. There is something mysterious about this: the system in question is possibly entirely deterministic, so the initial state should entirely determine the final state.

Let us consider briefly the situation when the states $x_1, x_2, \cdots$ are entirely independent of each other, distributed in X with law μ.

1.2 Simplified model: IID time dependence

Let us assume that X is a measurable space. For our purposes, we will assume that $x_1, x_2, \cdots$ are IID X-valued random variables, $x_i : \Omega \to X$, on a probability space $(\Omega, \mathbb{P})$. Let us write μ for the empirical law of the x_i, i.e.,

$$\mu(A) = \mathbb{P}\{x_1 \in A\}$$

for measurable $A \subset X$.

Suppose we don't know the law μ of the x_i, but you can sample as many times as you want. In this case, we can use time averages to determine μ as follows. Let $\phi : X \to \mathbb{R}$ be a bounded, measurable observable, so that $Y_i = \phi(x_i), i = 1, 2, \cdots$ are a sequence of IID $\mathbb{R}$-valued random variables. Observe that

$$\mathbb{E}(Y_1) = \int \phi(x) d\mu(x) \, .$$

Theorem 1 (Strong Law of Large Numbers)**.** *With probability one, we have that* $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N Y_i$ *converges to the (deterministic) value* $\mathbb{E}(Y_1)$.

In particular, we have that

$$\int \phi(x) d\mu(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \phi(x_i)$$

holds with probability 1. That is, the SLLN justifies exchanging time and spatial averages under the simplifying assumption that the states $x_1, x_2, \cdots$ are IID.

1.3 Dynamics and basic ergodic theory

The case when (x_i) are IID is instructive, but by no means indicative of what happens in real systems. More realistically, the state of the experiment evolves in time according to a *map* $f : X \to X$, so that given an initial state $x_0 \in X$ the state of the system x_n at time n is given by

$$x_n = f(x_{n-1}) = f^n(x_0) = \underbrace{(f \circ \cdots \circ f)}_{n \text{ times}}(x_0) \, .$$

Let us assume that $(X, \mathcal{F})$ is a measurable space. Our aim is to describe asymptotic dynamics by probability measures which are 'invariant' under the time evolution proscribed by f.

Definition 2. A probability measure μ on $(X, \mathcal{F})$ is called *invariant* if for all $A \in \mathcal{F}$, we have $\mu(f^{-1}A) = \mu(A)$.

We can think of invariant measures as *equilibrium statistics* for the dynamical system $f : X \to X$, in the sense that if x_0 is distributed according to μ, then $x_1 = f(x_0), x_2 = f^2(x_0), \cdots, x_n = f^n(x_0)$ are each distributed according to μ.

Example 3. Let $x_{per} \in X$, and assume $f^p(x_{per}) = x_{per}$ for some $p \geq 1$. We call $\{x_{per}, f x_{per}, \cdots, f^{p-1} x_{per}\}$ a periodic orbit of f. In this case, one can check that

$$\frac{1}{p} \sum_{i=0}^{p-1} \delta_{f^i x_{per}}$$

is an invariant measure for f. Here, δ_x denotes the Dirac delta-mass at $x \in X$.

1.3.1 Ergodicity

Definition 4. An invariant measure μ is called *ergodic* if, for some $A \in \mathcal{F}$, the property $f^{-1}A = A$ implies that $\mu(A) = 0$ or 1.

A set $A \in \mathcal{F}$ for which $f^{-1}A = A$ is sometimes itself called *invariant*, for the reason that trajectories initiated in A do not leave A. What ergodicity means as above is that if you could 'partition' X into two invariant pieces, then μ only 'sees' one such piece.

Ergodicity has many equivalent characterizations, a few of which we list here:

Lemma 5. *The following are equivalent.*

(a) μ *is ergodic.*

(b) If $A \in \mathcal{F}$ *is such that* $f^{-1}A \subset A$*, then* $\mu(A) = 0$ *or* 1*.*

(c) If $A \in \mathcal{F}$ *is such that* $f^{-1}A \supset A$*, then* $\mu(A) = 0$ *or* 1*.*

(d) If $A \in \mathcal{F}$ *is such that* $\mu(A) > 0$*, then* $\mu(\cup_{n \geq 0} f^{-n}A) = 1$*.*

Property (d) is equivalent to the following: if $\mu(A) > 0$, then μ-almost every $x \in X$ has the property that $f^n x \in A$ for some $n \geq 0$.

Remark 6. The old name for ergodicity is *metric transitivity*, and is perhaps more instructive as to what the concept means. In this context, *metric* refers to 'measure-theoretical', and *transitivity* refers to the way in which the μ-generic dynamics spreads over phase space. Property (d) in Lemma 5 is a kind of 'transitivity' property following from ergodicity.

1.4 The Birkhoff Ergodic Theorem

As it turns out, ergodicity is precisely the condition necessary needed to allow exchanging time and space averages with respect to a given invariant measure μ.

Theorem 7 (Birkhoff Ergodic Theorem)**.** *Let* μ *be an ergodic invariant measure for* f*. Let* $\phi \in L^1(\mu)$*. Then,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \phi \circ f^n(x) = \int \phi \, d\mu$$

for μ*-almost all* $x \in X$*.*

This theorem provides a number of useful insights about asymptotic dynamics with respect to an invariant measure. Below, for $A \in \mathcal{F}$, we write

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

Corollary 8. *Assume μ is ergodic. Let $A \in \mathcal{F}, \mu(A) > 0$. Then, for μ-almost all $x \in X$, we have that*

$$\lim_{N \to \infty} \frac{1}{N} \#\{1 \leq n \leq N : f^n x \in A\} = \mu(A) \, .$$

Proof. Apply the Birkhoff Ergodic Theorem to $\phi = \chi_A$. □

Remark 9. The Birkhoff Ergodic Theorem can be stated for non-ergodic measures μ. In that case, we still have that the limit $f^*(x) = \lim_N N^{-1} \sum_{n=1}^N \phi \circ f^n(x)$ exists for μ-almost all x, but the value of $f^*(x)$ may be non-constant in x. Nevertheless, $f^* \in L^1(\mu)$ does hold and $\int f^* d\mu = \int f d\mu$.

Exercise 10. Prove the following enhancement of Lemma 5: If μ is ergodic and $\mu(A) > 0$, then

$$\mu\{x \in X : f^n x \in A \text{ for infinitely many } n \geq 0\} = 1 \, .$$

1.4.1 Caveats

Just as important as what a theorem says is what it does *not* say. You might be tempted to think that the Birkhoff Ergodic Theorem justifies exchanging time and space averages in the sense described in Lecture 1, but this is not quite true for the following two reasons.

- The BET only entitles you to exchange time and space averages for 'μ-almost all' intiial conditions, while the support of μ could in fact be quite small, or even singular, with respect to a natural reference measure (in our case, Lebesgue). This is the case, for instance, for the Smale Solenoid described in Example 18, for which all invariant measures are singular with respect to Lebsegue.

- Even when there is an invariant measure μ which serves as a natural reference measure, e.g., Lebesgue measure is preserved (as is the case for, say, Hamiltonian systems), ergodicity does not necessarily follow. Indeed, it is possible to construct, using KAM theory, a large variety of Hamiltonian systems (preserving Liouville measure / Lebesgue measure on phase space) for which ergodicity is false in a dramatic way.

Remark 11. The rate at which time averages converge to space averages has not been specified, and can in general be quite slow. There are general results in this direction (see, e.g., upcrossing inequalities) but they are typically quite weak.

1.5 Goals of this course

Measures μ for which (1) converges for a 'large' set of initial states x_0 are called *physical*, meaning that they are physically relevant for the 'bulk' dynamics of the experiment / system. Typically, X is a manifold or just R^d for some d, in which case 'large' is meant to mean 'positive Lebesgue volume'.

Our focus in this mini-course are the following two items:

(a) What mechanisms are responsible for a deterministic dynamical system 'forgetting' its initial state?

(b) In particular, when can we mathematically justify the existence of a *physical measure* for a dynamical system?

1.6 Examples

Example 12. Let $Y_1, Y_2, \cdots$ be an IID sequence of $\mathbb{R}$-valued random variables with empirical law ν_Y on $\mathbb{R}$. We can encode this into a 'deterministic' dynamical system as follows. Define $X = [0,1]^{\otimes \mathbb{N}}$, writing $x = (x_1, x_2, \cdots)$ for $x \in X$. Define $\vartheta : X \to X, (\vartheta x)_n = x_{n+1}$ to be the leftward shift, so $\vartheta x = (x_2, x_3, \cdots)$. Define the observable $\phi : X \to \mathbb{R}$, $\phi(x) = x_1$. Then,

$$\frac{1}{N}\sum_{n=1}^{N} \phi(\vartheta^n x) = \frac{1}{N}\sum_{n=1}^{N} x_n$$

In particular, the SLLN for the IID sequence (Y_i) is the same as the Birkhoff ergodic theorem applied to the measure $\mu = \nu_Y^{\otimes \mathbb{N}}$ on X.

Exercise 13. Show that μ is invariant and ergodic for $\vartheta : X \to X$.

Smooth examples

Here we list a number of instructive, low-dimensional examples of dynamical systems to give a concrete illustration of the infinitesimal mechanisms which produce seemingly 'random' or 'decorrelated' behavior in asymptotic dynamics.

Example 14. Let $X = S^1$, the unit circle, parametrized by the unit interval $[0,1]$ with endpoints identified. Define $f : X \to X$ by

$$fx = 2x \,(\mathrm{mod}\ 1)\,.$$

This is called the *doubling map*.

Exercise 15.

(a) Show that Lebesgue measure Leb on the circle S^1 is an invariant measure for f.

(b) Let $I \subset S^1$ be a small interval of length $\epsilon > 0$. Show that $f^n(I) = S^1$ for all sufficiently large n. How do you interpret this in the context of a dynamical system 'randomizing' its initial condition?

Example 16. Let $X = \mathbb{T}^2 = S^1 \times S^1$, parametrized as the unit box $[0,1]^2$ with opposite sides identified. Regard $x \in X$ as column vectors $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $x_i \in S^1 \cong [0,1), i = 1,2$ and define

$$fx = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} x \,(\mathrm{mod}\ 1)\,,$$

where 'mod 1' is taken on both coordinates individually. The resulting map of $\mathbb{T}^2$ is called the CAT map (short for 'continuous automorphism of the torus').

Exercise 17.

(a) Show that f is actually a continuous map on $\mathbb{T}^2$.

(b) Show that Lebesgue measure on $\mathbb{T}^2$ is invariant for f.

(c) Evaluate the differential $D_p f, p \in \mathbb{T}^2$ by identifying $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and identifying tangent spaces $\mathbb{T}^2$ with the same copy of $\mathbb{R}^2$.

(c) Let v^u, v^s denote unit eigenvectors for the matrix defining f above, where v^u corresponds to the eigenvalue $\lambda^u = \frac{1}{2}(3+\sqrt{5})$ and v^s to $\lambda^s = \frac{1}{2}(3-\sqrt{5})$.

Let $p, q \in \mathbb{T}^2, d(p,q) \leq 1/10$ and assume that the length of the displacement vector $p - q$ coincides with $d(p,q)$. What happens to $d(f^n p, f^n q)$ if $p - q$ is parallel to (i) v^u, (ii) v^s?

Example 18. Let $X = S^1 \times D$, where $D \subset \mathbb{C}$ is the closed unit disk in the complex plane. As usual, S^1 is parametrized by the unit interval $[0,1]$ with endpoints identified. We can visualize X as the solid or filled-in torus (i.e., donut) in $\mathbb{R}^3$. Define $f : X \to X$ by

$$f(t,z) = (2t \,(\mathrm{mod}\ 1), \frac{1}{4}z + \frac{1}{2}e^{2\pi i t})$$

This map is called the *Smale solenoid map.*

Exercise 19. Show that if μ is an invariant measure f, then it is singular with respect to Lebesgue mesaure on X. Specifically, show that there is a set $\mathcal{A} \subset X$ such that $\mu(\mathcal{A}) = 1$ for all invariant μ, while $\mathrm{Leb}(A) = 0$. *Hint: Define* $\mathcal{A}_n = \cap_{i=0}^n f^i(\mathcal{A})$, $\mathcal{A} = \cap_{n\geq 1}\mathcal{A}_n$ *and show that* $\mathrm{Leb}(\mathcal{A}_n) = c\,\mathrm{Leb}(\mathcal{A}_{n-1})$, *where* $c \in (0,1)$ *is a constant. Conclude that* $Leb(\mathcal{A}) = 0$. *Conversely, argue that* $\mu(\mathcal{A}) = 1$.

Example 20. The mechanism of stretching/contracting in the above examples is *uniform.* Real-world systems might exhibit these mechanisms, but usually not uniformly in phase space. One famous (albeit notorious) example is the Chirikov standard map

$$F_L(x,y) = (2x + L\sin(2\pi x) - y, x)$$

defined on $\mathbb{T}^2$. Clearly for $L \gg 1$ this map exhibits very strong expansion in the horizontal (x) direction; since it preserves volume, it contracts along some roughly vertical direction on most of phase space. At the same time, this expansion/contraction is not uniformly expressed in phase space, as it fails along neighborhoods of the lines $\{x = 1/4\}, \{x = 3/4\}$. The following is an open question against which many graduate students and famous dynamicsts have failed:

Problem 21 (Standard map conjecture; attempt at your own peril!)**.** Show that for any value of L that the map F_L has the property that a positive Leb-measure set of points $x \in \mathbb{T}^2$ satisfy

$$\liminf_{n\to\infty} \frac{1}{n} \log \|D_x F_L^n\| > 0\,.$$