

Introduction to Ergodic Theory - Lecture Notes

Professor Omri Sarig - Gulbenkian Summer School 2015

Francisco Machado

July 15, 2015

Based on mine and Sagar Pratapsi's notes

1 Lecture 1

GOAL: How to understand *random* behavior in *deterministic* dynamics

Example 1: Coin toss

Example 2: A box filled with 1 liter of gas

Any measured quantity appears to have some random behavior, sort of fluctuations, yet the evolution of the system is completely deterministic by Hamilton's equations.

1.1 Deterministic Setup:

A continuous dynamical system ("semiflow") is a 1-parameter family of maps $T_t : X \rightarrow X$, ($t \geq 0$) st $T_t \circ T_s = T_{t+s}$.

Interpretation

- X ("space" or "phase space" for physicists) = { all possible states of the system } where a state is a complete description of the system (enough information to deterministically evolve it in time).
- T_t ("law of motion"): $T_t \begin{bmatrix} \text{state at} \\ \text{time 0} \end{bmatrix} = \begin{bmatrix} \text{state at} \\ \text{time } t \end{bmatrix}$

Determinism gives us that T_t is a function, while $T_t \circ T_s = T_{t+s}$ expresses consistency, evolving a time t then a time s is the same to evolving a time $t + s$.

An orbit corresponds to the set of points taken by a particle as it is evolved in time, $\{T_t(x)\}_{t=0}^{\infty}$, where x is the initial state.

1.2 Measurement

A measurement, $f : X \rightarrow \mathbb{R}$ corresponds to the mapping of the state of our system to a number (or other mathematical object like matrix, etc), which is the outcome of our measurement.

Example: $f(\vec{q}_1, \dots, \vec{q}_N, \vec{p}_1, \dots, \vec{p}_N) = \sum_{\text{particle } i} \frac{1}{2m_i} |\vec{p}_i|^2$, measures the kinetic energy of the system.

Although we can't tract the orbit of a system, we can characterize the measurements, by having the time series of measurements $f(T_t(x))$, parametrized by a parameter t .

This time series will be the one with properties of a random process.

1.3 Discrete time dynamical system

Instead of a family of maps, there exists only one map $T : X \rightarrow X$, st:

$$T \begin{bmatrix} \text{state at} \\ \text{time 0} \end{bmatrix} \rightarrow \begin{bmatrix} \text{state at} \\ \text{time 1} \end{bmatrix}$$

which gives that:

$$T^n \begin{bmatrix} \text{state at} \\ \text{time 0} \end{bmatrix} \rightarrow \begin{bmatrix} \text{state at} \\ \text{time n} \end{bmatrix}$$

Dynamical systems is a theory of *iterative functions*.

1.4 Modeling Randomness

What is the mathematical setup for studying randomness?

Example: I choose $x \in [0, 1]$ randomly and keep it secret. I will answer a countable collection of “reasonable” questions of the form yes or no, about x .

Def: A *probability space* is $(\Omega, \mathcal{F}, \mu)$.

- Ω is a set corresponding to the “sample space”, the possible values of the system.
- \mathcal{F} , a σ -algebra, is a collection of subsets of Ω , called *measurable sets*, such that:
 - \mathcal{F} contains \emptyset .
 - \mathcal{F} is closed under forming complements: If $E \in \mathcal{F}$ then $E^c = \Omega \setminus E \in \mathcal{F}$.
 - If $E_i \in \mathcal{F}$ for all $i \in \mathbb{N}$ (in short a countable number of sets), then $\bigcup_{i=1}^{\infty} E_i$ and $\bigcap_{i=1}^{\infty} E_i$ both belong to \mathcal{F}

(Note: A reasonable question then becomes a question of the form, “does x belong to E for some E in \mathcal{F} ?”)

- μ , the probability measure, is a function, $\mu : \mathcal{F} \rightarrow [0, 1]$, st:
 - $\mu(\emptyset) = 0, \mu(\Omega) = 1$.
 - if E_i are countable pairwise disjoint sets in \mathcal{F} , then: $\mu(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n \mu(E_i)$.

1.5 Important theorems in Measure Theory

Theorem (Lebesgue)

There exists a probability space $([0, 1], \mathcal{B}, \mu)$, where:

- \mathcal{B} is a σ -algebra which contains all subintervals (and many, many more)
- $\mu([a, b]) = b - a$
- $\mu(a + E) = \mu(E), \forall E \in \mathcal{F}$ st $E + a \in \mathcal{F}$ where $x \in a + E$ if $x - a \in E$

Theorem (Vitaly) Lebesgue theorem is false for $\mathcal{F} = \{ \text{all subsets of } [0, 1] \}$.

Def: A *measurable function* is a function $f : \Omega \rightarrow \mathbb{R}$ st $\forall t \in \mathbb{R} [f > t] \in \mathcal{F}$ where $[f > t] := \{ \omega \in \Omega | f(\omega) > t \}$.

Exercise: f is measurable iff there is a countable list of reasonable questions whose answer determines the value of f .

Def: A stochastic process is a sequence of measurable functions $f_n : \Omega \rightarrow \mathbb{R}$, on the same probability space.

Exercise: Show that $\mathbb{P}(a_k < f_{i_k} < b_k, k = \{1, 2, \dots\}) := \mu(\omega \in \Omega : f_{i_k} > a_k \wedge f_{i_k} < b_k)$ is a stochastic process.

Def: A probability preserving transformation is $(\Omega, \mathcal{F}, T, \mu)$ where:

- $(\Omega, \mathcal{F}, \mu)$ is a probability space.
- $T : \Omega \rightarrow \Omega$ is a *measurable map*, a map such that $\forall E \in \mathcal{F}, T^{-1}(E) = \{ \omega \in \Omega | T(\omega) \in E \} \in \mathcal{F}$
- $\mu(T^{-1}E) = \mu(E), \forall E \in \mathcal{F}$.

Every measurement gives rise to a stochastic process:

$$f, f \circ T, f \circ T^2, \dots$$

If we show that a stochastic process has the characteristic of a random process, we can make the relationship between a random process and a dynamical system precise.

2 Lecture 2

Probability-preserving $\Rightarrow \mu(T^{-1}E) = \mu(E)$, in $(\Omega, \mathcal{F}, \mu, T)$ with $(\Omega, \mathcal{F}, \mu)$ a probability space and $T : X \rightarrow X$.

Remember a measurement is a function that associates to each state a certain value, $f : \Omega \rightarrow \mathbb{R}$
Our goal is to show that these functions show stochastic behavior.

2.1 Strong Law of Large Numbers

Let x_i independent, identically distributed random variables.

Then we have: $\boxed{\frac{\sum_i^N x_i}{N} \xrightarrow{N \rightarrow +\infty} \mathbb{E}(x_0)}$

Similarly we will arrive at:

$$\frac{1}{N} \sum_{i=0}^N f(T^i x) \xrightarrow{N \rightarrow \infty} \int_{\Omega} f d\mu, \text{ almost everywhere}$$

2.2 Measure Theory - Integrable Functions and their Integrals

Recall: $f : \Omega \rightarrow \mathbb{R}$ is *measurable* (with respect to a σ -algebra) if:

$$\forall a \in \mathbb{R}, [f > a] := \{\omega \in \Omega : f(\omega) > a\} \text{ is in } \mathcal{F}$$

Example 1: Indicators of measurable sets:

$$1_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases} \text{ for } E \in \mathcal{F}$$

Since:

$$[1_E > a] = \begin{cases} \Omega & a < 0 \\ E & 0 \leq a < 1 \\ \emptyset & a \geq 1 \end{cases}$$

Example 2: Simple functions / discrete random variable

$$f = \sum_{i=1}^N \alpha_i 1_{E_i} \quad (E_1, \dots, E_N \in \mathcal{F})$$

Example 3: Pointwise limits of simple functions.

This follow from **Lemma:**

If $f_n : \Omega \rightarrow \mathbb{R}$ are measurable and $|f_n| \leq M$, then $\limsup_{n \rightarrow \infty} f_n(\omega)$ and $\liminf_{n \rightarrow \infty} f_n(\omega)$ are measurable.

Proof:

If $(a_n)_{n \geq 0}$, is bounded then

$$\begin{aligned} \limsup_{n \rightarrow \infty} a_n &= \max\{L \mid \exists \text{ subsequence } a_{n_k} \rightarrow L\} \\ \liminf_{n \rightarrow \infty} a_n &= \min\{L \mid \exists \text{ subsequence } a_{n_k} \rightarrow L\} \end{aligned}$$

Then:

$$\begin{aligned} \limsup_{n \rightarrow \infty} f_n(\omega) > a &\Leftrightarrow \exists_n \left(\limsup_{n \rightarrow \infty} f_n(\omega) \geq a + \frac{1}{n} \right) \\ &\Leftrightarrow \exists_n [\forall_m \forall_N \exists_{k > N} (f_k(\omega) < a + 1/n - 1/m)] \end{aligned}$$

Transforming from \exists and \forall into unions and intersections:

$$\{\omega \in \Omega \mid \limsup_{n \rightarrow \infty} f_n(\omega) > a\} = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \left[f_k > a + \frac{1}{n} - \frac{1}{m} \right]$$

Since it is a countable union and intersection of sets in \mathcal{F} , then it also belongs to \mathcal{F} , hence it is a measurable function.

Proposition: Every bounded measurable function is the uniform limit of simple functions.

Proof: Suppose $|f| \leq M$, f is measurable, let:

$$f_n = \sum_{k=-Ma}^{Ma} \frac{k}{n} 1_{[k/n \leq f \leq (k+1)/n]}$$

For every $\omega \in \Omega$, there's a unique k , st $k/n \leq f_n(\omega) \leq (k+1)/n$. For this k :

$f(\omega) \in (k/n, (k+1)/n]$, $f_n(\omega) = k/n$

So $|f_n(\omega) - f(\omega)| < \frac{1}{n} \Rightarrow \sup |f_n - f| \leq \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$.

As we wanted.

2.3 Definition of the integral

Case 1: Indicators of measurable sets, $f = 1_E, E \in \Omega$.

The integral should be equal to the measure of the interval, $\int_{\Omega} 1_E d\mu = \mu(E)$.

Case 2: Simple functions $f = \sum_{i=1}^N \alpha_i 1_{E_i}$:

We expect the integral to be linear, $\int f d\mu = \sum_{i=1}^N \alpha_i \mu(E_i)$.

Suppose that E_1, \dots, E_N are disjoint sets, then $\mathbb{P}(f = \alpha_i) = \mu(E_i)$. So $\int f d\mu = \sum_{i=1}^N \alpha_i \mathbb{P}(f = \alpha_i) = \mathbb{E}(f)$, the expectation of the function.

Case 3: General bounded measurable functions

Every function f like this is the uniform limit of simple functions.

$f = \lim_{n \rightarrow \infty} f_n$, f_n simples.

So we define:

$$\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

Proposition: The limit exists, is independent of the choice of f_n , and has the properties:

- Linearity: $\int_{\Omega} (\alpha f + \beta g) d\mu = \alpha \int_{\Omega} f d\mu + \beta \int_{\Omega} g d\mu$
- Positivity: $f \geq 0 \Rightarrow \int_{\Omega} f d\mu \geq 0$
- Monotocity: $f \geq g \Rightarrow \int_{\Omega} f d\mu \geq \int_{\Omega} g d\mu$
- Δ -ineq: $|\int_{\Omega} f d\mu| \leq \int_{\Omega} |f| d\mu$

Proof: It is true for simple functions (exercise). Suppose f is bounded and measurable, and let f_n , be simple functions st $\sup |f - f_n| \rightarrow 0$.

Claim: $\lim \int f_n d\mu$ exists.

Proof: We check the Cauchy criteria. Fix $\epsilon \geq 0$. Choose N such that, $n > M \Rightarrow |f_n(\omega) - f(\omega)| < \epsilon$ for all ω . Then, for all $m, n > M$, $|f_n(\omega) - f_m(\omega)| \leq |f_n(\omega) - f(\omega) + f(\omega) - f_m(\omega)| < 2\epsilon$.

By the Δ -ineq and linearity:

$$\left| \int f_n d\mu - \int f_m d\mu \right| = \left| \int (f_n - f_m) d\mu \right| \leq \int |f_n - f_m| d\mu \leq \int 2\epsilon d\mu = 2\epsilon$$

So, $a_n = \int f_n d\mu$ is a Cauchy sequence.

Claim: If $\{f_n\}, \{g_n\}$ are two sequences of simple functions which converge uniformly to f , then:
 $\lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu$

Proof: Fix $\epsilon > 0$. Choose N st if $n > N$, then $|f_n(\omega) - f(\omega)| < \epsilon$ for all ω as well for g_n .

Then, necessarily, $n > N \Rightarrow |f_n(\omega) - g_n(\omega)| < 2\epsilon$, hence $|\int_{\Omega} f_n d\mu - \int_{\Omega} g_n d\mu| = |\int_{\Omega} (f_n - g_n) d\mu| \leq \int |f_n - g_n| d\mu \leq \int 2\epsilon d\mu = 2\epsilon$.

Therefore $\boxed{\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \lim_{m \rightarrow \infty} \int_{\Omega} g_m d\mu}$.

Example

$$D(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases} \Rightarrow \int_{[0,1]} D(x) d\mu = 1\mu(\mathbb{Q}) + 0\mu(\mathbb{Q}^c) = 0$$

2.4 Bounded Convergence Theorem

Suppose f_n are measurable functions on a probability space st:

$f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ exists for all ω .

If $\exists M > 0$ st $|f_n| \leq M$ for all n , then:

$$\boxed{\int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu}$$

3 Lecture 3

3.1 Ergodic Theorem / Strong Law of Large Number

Our goal is to get a Law of Large number for an ergodic process.

$$\frac{1}{N} \sum_{i=0}^{N-1} f(T^i \omega) \xrightarrow{N \rightarrow \infty} \int_{\Omega} f d\mu \text{ ae}$$

Definition: If property P occurs almost everywhere (ae) then:

$$\mu(\omega | \text{Property } P \text{ does not happen in } \omega) = 0$$

Obstruction: Conserved quantities do not change over time.

Def: A measurable function $f : \Omega \rightarrow \mathbb{R}$ is T -invariant if $f \circ T = f$. ae.

Def: A probability preserving system $(\Omega, \mathcal{F}, \mu, T)$ is called *ergodic* if every T -invariant function f is equal to a constant function ae.

In order words: \exists_c st $\mu[f \neq c] = 0$

Birkhoff's Pointwise Ergodic Theorem

Suppose $(\Omega, \mathcal{F}, \mu, T)$ is a probability preserving system, and let f be a bounded measurable function.

Then $\bar{f}(\omega) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i \omega)$ exists ae.

If the system is ergodic, then $\bar{f}(\omega) = \int_{\Omega} f d\mu$ ae.

Proof

Some reductions, first:

- $|f| \leq 1$ by rescaling
- $0 \leq f \leq 1$ otherwise $f = f1_{[f \geq 0]} - |f|1_{[f < 0]}$. Prove for each term and then the limit of the difference is the difference of the limit.

Idea: Show that $\limsup = \liminf$.

Let:

$$\bar{A} = \lim_{N \rightarrow \infty} \sup \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega)$$

$$\underline{A} = \lim_{N \rightarrow \infty} \inf \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega)$$

If $\bar{A} = \underline{A} \Rightarrow$ limit exists.

Notice that $\bar{A} \geq \underline{A}$, then if the limit exists $\bar{A} = \underline{A}$ ae $\Rightarrow \mu(\bar{A} > \underline{A}) = 0$

Exercise: Show that $\mu(\bar{A} > \underline{A}) = 0 \Leftrightarrow \forall \mu(\bar{A} - \underline{A} > 1/n)$.

We'll use show $\int (\bar{A} - \underline{A}) d\mu = 0 \Rightarrow \mu(\bar{A} - \underline{A} > 1/n) = \int 1_{\bar{A} - \underline{A} > 1/n} d\mu \leq n \int (\bar{A} - \underline{A}) d\mu = 0$

Therefore $\bar{A} = \underline{A}$ ae.

Note that: $\int 1_{\bar{A} - \underline{A} > 1/n} d\mu \leq \bar{A} - \underline{A}$

Exercise: Check first equality in previous expression.

Start by proving $\int (\bar{A} - \underline{A}) d\mu = 0$.

Fix $C > 0$ very big and, $\epsilon > 0$ very small and define:

$$\tau(\omega) = \min \left\{ n > 0 \mid \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) > \bar{A}(\omega) - \epsilon \right\}$$

Well defined value because of the \limsup .

We now color the interval $0, 1, \dots, N$ as follows.

- $k=0$: If $\tau(\omega) > C$, color k "red". If $\tau(\omega) \leq C$, color $0, 1, \dots, \tau(\omega) - 1$ "blue". Move to next colorless k
- If $\tau(T^k \omega) > C$, color k "red", otherwise color the next $\tau(T^k \omega)$ k 's "blue". Move to next colorless k .
- Stop when all k 's are colored or you want to color a segment blue but there are not enough k 's left.

We estimate, from below:

$$\sum_{k=0}^{N-1} (f(T^k(\omega)) + 1_{[\tau \geq C]}(T^k \omega))$$

Contribution of red k 's, if k is red, the $T^k\omega \in [\tau > C]$ so:

$$\sum_{k \text{ red}} \underbrace{f(T^k\omega)}_{\geq 0} + \underbrace{1_{[\tau > C]}(T^k\omega)}_{\geq 1} \geq \# \text{ reds} \geq (\# \text{ reds})(\bar{A}(\omega) - \epsilon)$$

Contribution of blue k 's

$$\begin{aligned} & \sum_{\text{blue segments}} \left(\sum_{k \in \text{segment}} (f(T^k\omega) + 1_{[\tau > C]}) \right) \geq \\ & \geq \sum_{\text{blue segments}} \left(\sum_{k \in \text{segment}} f(T^k\omega) \right) \geq \sum_{\text{blue segments}} (\text{length})(\bar{A}(T^{\text{start of segment}}\omega) - \epsilon) \end{aligned}$$

But \bar{A} is T-invariant so we have:

$$= \sum_{\text{blue segments}} (\text{length})(\bar{A}(\omega) - \epsilon) = (\# \text{ blues})(\bar{\omega} - \epsilon)$$

Then, from below, the sum is bounded by:

$$(\# \text{ red} + \# \text{ blue})(\bar{A}(\omega) - \epsilon) \geq (N - C)(\bar{A} - \epsilon)$$

Then we have:

$$\frac{1}{N} \sum_{k=0}^{N-1} (f(T^k\omega) + 1_{[\tau > C]}(T^k\omega)) \geq (1 - C/N)(\bar{A}(\omega) - \epsilon)$$

By monotonicity and linearity:

$$\frac{1}{N} \sum_{k=0}^{N-1} \left(\int_{\Omega} f \circ T^k d\mu + \int_{\Omega} 1_{[\tau > C]} \circ T^k d\mu \right) \geq (1 - C/N) \left(\int_{\Omega} (\bar{A} - \epsilon) d\mu \right)$$

Exercise: Show that $\int_{\Omega} f \circ T^k d\mu = \int_{\Omega} f d\mu$, $\forall f$ bounded measurable and assuming that μ is T-invariant.

Then we get:

$$\int_{\Omega} f d\mu + \mu(\tau > C) \geq (1 - C/N) \left(\int_{\Omega} \bar{A} d\mu - \epsilon \right)$$

Taking the limit as $N \rightarrow \infty$ and then as $C \rightarrow \infty$ (note that $\mu(\tau > C) \rightarrow 0$), gives us:

$$\int_{\Omega} f d\mu \geq \left(\int_{\Omega} \bar{A} d\mu - \epsilon \right)$$

The limit $\epsilon \rightarrow 0$ gives us the upper bound we wanted.

Exercise: Redefine τ to obtain the argument for \liminf , that $\int_{\Omega} f d\mu \leq \int_{\Omega} \underline{A} d\mu$

In summary, $\int (\bar{A} - \underline{A}) d\mu \leq \int_{\Omega} f d\mu - \int_{\Omega} f d\mu = 0$.

However $\bar{A} - \underline{A} \geq 0$, meaning that $\int (\bar{A} - \underline{A}) d\mu = 0$, hence the limit exists.

We showed for almost every point, that $\bar{f} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega)$ exists.

It is easy to see that \bar{f} is T-invariant, $\bar{f} \circ T = \bar{f}$.

So in the ergodic case, every invariant function is constant ae.

Let's calculate C .

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega)$$

$$C = \int_{\Omega} c d\mu = \int_{\Omega} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega) d\mu$$

By the bounded convergence theorem (from end of last lecture), we can switch the limit and the integration.

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \int_{\Omega} f(T^k \omega) d\mu$$

$$C = \lim_{N \rightarrow \infty} \int_{\Omega} f(\omega) d\mu = \int_{\Omega} f(\omega) d\mu$$

As we wanted to show, if we have an **ergodic** dynamical system.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k \omega) = \int_{\Omega} f(\omega) d\mu$$

As we wanted to show.

4 Lecture 4

4.1 Independence and Mixing

Probability Theory: Suppose $(\Omega, \mathcal{F}, \mu)$ is a probability space.

Two events $E, F \in \mathcal{F}$ are *independent* if:

$$\mu(E \cap F) = \mu(E)\mu(F)$$

In this case, the conditional probability reduces to: $\mathcal{P}(E|F) = \frac{\mu(E \cap F)}{\mu(F)} = \mu(E)$

The occurrence of F yields no further information on the event E .

Approximate Independence When $\mu(E \cap F) \approx \mu(E)\mu(F)$.

Example: Coin-tossing

For a coin, the face coming out depends on the time the coin is in the air $t = v_0/g$, and the angular frequency ω (in rad/s).

The number of half turns is then $\lfloor \frac{\omega v_0}{g\pi} \rfloor$. The product of ωv_0 is then the determining factor and in the space $(\omega-v_0)$, the different outcomes are separated by hyperboles.

Imagining that there is a spread of the initial velocities and angular momentum, the intersection of the uncertainty ellipse with the hyperbolas gives us the probability of each outcome.

If the ellipse is large or the hyperbole very thin, the result is that the ellipse intersects about the same area of head and tails phase space.

4.2 Dynamical System

Def: A probability preserving system $(\Omega, \mathcal{F}, \mu, T)$ is called *mixing* if for all $E, F \in \mathcal{F}$
 $(E \cap T^{-n}F) \xrightarrow{n \rightarrow \infty} \mu(E)\mu(F)$

Idea: $\mathcal{P}(T^n(\omega) \in F | \omega \in E) \rightarrow \mathcal{P}(\omega \in F) \rightarrow \mathcal{P}(T^n\omega \in F)$

Proposition: A probability preserving system $(\Omega, \mathcal{F}, \mu, T)$ is ergodic iff $\forall E, F \in \mathcal{F}$

$$\frac{1}{N} \sum_{k=0}^{N-1} \mu(E \cap T^{-k}F) \xrightarrow{N \rightarrow \infty} \mu(E)\mu(F)$$

Proof (\Rightarrow):

Suppose T is ergodic, and $E, F \in \mathcal{F}$.

By the ergodic theorem:

$$\frac{1}{N} \sum_{k=0}^{N-1} f \circ T^{-k}\omega \xrightarrow{N \rightarrow \infty} \int_{\Omega} f d\mu$$

Choosing $f = 1_F$ we have:

$$\frac{1}{N} \sum_{k=0}^{N-1} 1_F \circ T^{-k}\omega \xrightarrow{N \rightarrow \infty} \mu(F)$$

Multiplying by $1_E(\omega)$ and integrating over all phase space we have:

$$\int_{\Omega} \left[\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} (1_E(\omega))(1_F \circ T^{-k})\omega \right] d\mu = \int_{\Omega} 1_E \mu(F) d\mu$$

By the bounded convergence theorem, linearity of integral:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \int_{\Omega} (1_E(\omega))(1_F \circ T^{-k})\omega d\mu = \mu(E)\mu(F)$$

Note: $1_E 1_F \circ T^{-k}$ is one if the state is in the intersection of E and $T^{-k}F$ and zero otherwise, so the integral over all space yields only the measure of $E \cap T^{-k}F$.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(E \cap T^{-k}F) = \mu(E)\mu(F)$$

As we wanted to show.

Proof (\Leftarrow)

To prove the inverse implication, consider $E = F = [f > c]$, for some T-invariant function. If this T-invariant function is constant in all Ω , then T is ergodic. Since f is T-invariant, then $T^{-k}F = F$

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(F \cap F) &= \mu(F)\mu(F) \\ \mu(F) &= (\mu(F))^2 \end{aligned}$$

This means that $\mu(F)$ is either 0 or 1.

Exercise: Use this to show $\mu(f < c) = 0$ and $\mu(f > c) = 0$ for some c .

This means that there is at most a set of measure 0, where the function is not a constant, which is the definition of ergodicity.

4.3 Angle Doubling Map

Given the probability preserving system $(\Omega, \mathcal{F}, \mu, T)$

$$\Omega = [0, 1]$$

\mathcal{F} := Smallest σ -algebra containing the intervals. $\bigcap \{ \mathcal{G} \subset 2^\Omega \mid \mathcal{G} \text{ is a } \sigma\text{-algebra and } \mathcal{G} \supset \{\text{intervals}\} \}$

μ = Lebesgue's measure $\mu([a, b]) = b - a$.

$$T(x) = 2x \text{ mod } 1$$

Exercise: Show that the system is probability-preserving.

In binary representation $T^k(0.x_1x_2\dots) = 0.x_kx_{k+1}\dots$

Dyadic intervals:

$$[a_0, \dots, a_{n-1}] = [0, a_1, \dots, a_{n-1}, *, *, \dots] = \left[\sum_{k=0}^{n-1} \frac{a_k}{2^{k+1}}, \sum_{k=0}^{n-1} \frac{a_k}{2^{k+1}} + \frac{1}{2^n} \right]$$

For any two dyadic intervals $[a_0, \dots, a_{\alpha-1}]$, $[b_0, \dots, b_{\beta-1}]$, if $n > \alpha$,

$$\begin{aligned} \#[a] \cap T^{-n}[b] &= \#[a_0, \dots, a_{\alpha-1}, *, \dots, *, b_0, \dots, b_{\beta-1}] \\ &= \# \bigsqcup_{\xi=(\xi_1, \dots, \xi_n)} [a, \xi, b] = 2^{n-\alpha} \end{aligned}$$

We have $2^{n-\alpha}$ intervals with length $2^{-n-\beta}$ each, which yields a total size $2^{-\alpha}2^{-\beta}$ which is the product of the size of each of the two sets. Mixing occurs.

To show mixing for general Borel sets, use:

Approximation Lemma: For every Borel set \mathcal{B} , and $\epsilon > 0$ there are dyadic intervals I_1, \dots, I_n st:

$$\mu(E\Delta \bigsqcup_{i=1}^m I_i) < \epsilon$$

Where $A\Delta B := (A \setminus B) \cup (B \setminus A)$, the non-intersecting part of both sets.

5 Lecture 5

Let $(\Omega, \mathcal{B}, \mu, T)$ a probability preserving system and $f : \Omega \rightarrow \mathbb{R}$ a measurement.

Then $\{f \circ T^k\}_{k=0}^{\infty}$ defines a stochastic process.

If we can show that the series has properties of a random process then we can make a precise comparison between the two.

In particular it is interesting to ask **“how random” can the system look.**

5.1 Entropy

Def: A *Bernoulli Process* is a stochastic process $(\{X_n\}_{n=-\infty}^{\infty})$ for which \exists finite set $\{a_1, \dots, a_N\}$ and a vector $\vec{p} = (p_1, \dots, p_N), 0 \leq p_i \leq 1, \sum_{i=0}^N p_i = 1$, such that:

- $\mathcal{P}(X_i = a_k) = p_k$ for all i
- X_i are independent $\Leftrightarrow \forall a_{i-m}, \dots, a_{i_m} \mathcal{P}(X_{-m} = a_{i-m}, \dots, X_m = a_{i_m}) = \prod_{k=-m}^m p_{i_k}$

In short, $\mathcal{P}(\bigcap_{k=1}^m [X_k = a_{i_k}]) = \prod_{k=1}^m \mathcal{P}(X_k = a_{i_k})$

Most random stochastic process!

Def: A probability preserving system $(\Omega, \mathcal{F}, \mu, T)$ *simulates a Bernoulli process* if \exists measurable function $f : \Omega \rightarrow \{a_1, \dots, a_N\}$ st the time series is a Bernoulli process, i.e., there exists a probability vector (p_1, \dots, p_N) st:

$$\begin{aligned} \mu\{\omega \in \Omega \mid f(T^k\omega) = a_{i_k} (k = 0, \dots, m-1)\} &= \prod_{k=0}^{m-1} p_{i_k} \\ \text{for all } m \in \mathbb{N} \text{ and } a_{i_1}, \dots, a_{i_m} &\in \{a_1, \dots, a_n\} \end{aligned}$$

Quasi-example (one sided definition of stochastic process):

$$Tx = 2x \bmod 1.$$

$$\text{And } f(x) = 1_{[1/2, 1)}$$

Suppose $x = 0.x_1x_2\dots$ in binary expansion. Then $T^i x = 0.x_{i+1}x_{i+2}\dots$ (in binary), hence $f(T^i x) = x_{i+1}$.

$$\begin{aligned} \mu\{x \in [0, 1] | f(x) = a_1, \dots, f(T^{n-1}x) = a_n\} &= \\ \mu\{0.a_1a_2\dots a_n \text{ **}\} &= \mu\left(\left[\sum_{k=1}^n \frac{a_k}{2^k}, \sum_{k=1}^n \frac{a_k}{2^k} + \frac{1}{2^n}\right]\right) \\ &= \frac{1}{2^n} \end{aligned}$$

Def: The *entropy* of a Bernoulli process with prob. vector (p_1, \dots, p_N) is the non-negative number.

$$-\sum_{i=1}^N p_i \log(p_i)$$

Theorem (Sinai): If a probability preserving system simulates a Bernoulli process (B-process) with entropy H , then it simulates any B-process with entropy $< H$.

Def: The *metric entropy* of a probability preserving system is

$$\sup\{H | \text{the system simulates a B-process with entropy } H\}$$

Examples

1: Geodesic flow on compact surfaces with negative curvature. Given a point and a direction it slides the point across the geodesic $g^t(\omega)$

If sphere there is zero entropy.

2: Playing billiard, the ball reflects, irrotational motion \rightarrow zero entropy. Same in a rectangle. Yet on rectangle with two side substituted by semi circles there is positive entropy.

3: (Arnold's Cat Map)

$$T : \mathcal{T}^2 \rightarrow \mathbb{T}^2 \text{ st } T(x, y) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \text{ mod } \mathbb{Z}^2.$$

Two directions, a stable and an unstable. Positive entropy.

5.2 What are the mechanisms that produce positive entropy?

Setup: Ω is a compact smooth manifold and $T : \Omega \rightarrow \Omega$ is a twice differentially invertible map.

Def: T has *exponential sensitivity to initial conditions* at $x \in \Omega$ if $\exists \lambda > 0, \exists \epsilon_0 > 0$ st that $\forall n > 0, \exists y \in \Omega$ st:

- $\text{dist}(x, y) < e^{-\lambda n}$
- $\text{dist}(T^n(y), T^n(x)) > \epsilon_0$

In linear time we have an exponential growth in error.

Example: $Tx = 2x \text{ mod } 1$

To know $T^n x$ well we need to know x up to $\frac{1}{2^{n+1}}$. The larger the n , the error in our initial knowledge of x becomes exponentially smaller.

Theorem (Ruelle, Margulis,...)

Under the previous assumptions, if T preserves an ergodic probability measure μ with positive entropy, then:

$x \in \Omega$ ae has exponential sensitivity to initial conditions

Undelying idea: you need exponential sensitity to have positive entropy.

5.3 Katok Horseshoe

From now on assume T a twice continuously differential map on a compact surface of dim 2.

Theorem (Katok, '80): If T has an ergodic inv. prob. measure with positive entropy then the following picture must appear somewhere in the manifold.

There is a point A with a stable direction and an unstable direction. These directions form lines such that points in each line are sent to the same line, W^S and W^u respectively.

In more precise notation $T(W^s) \subset W^s$ and contracts uniformly and $T(W^u) \subset W^u$ expand uniformly.

There will also be a point B , where the lines meet again perpendicularly.

Considering B and a neighbourhood in W^u , the point will evolve in W^s while the neighbourhood expands from the point, as it approaches A .

However all these points also belong to W^u , meaning that it forms a very twisted curve.

The same procedure can be formed by considering A and a neighbourhood in W^s and applying T^{-1} , forming an intricate path for W^s .

Note though, that now, we have a countable number of intersections, between the two lines, where the same process occurs.

We then have an uncountable number of intersections, a **fractal structure!**

This is known as the Katok Horseshoe.

6 Acknolegments

I would like to thank Professor Omri Sarig for the terrific series of lectures.

I would also like to thank Sagar Pratapsi for providing the lecture notes for comparison as well as proof-reading these notes for mistakes.

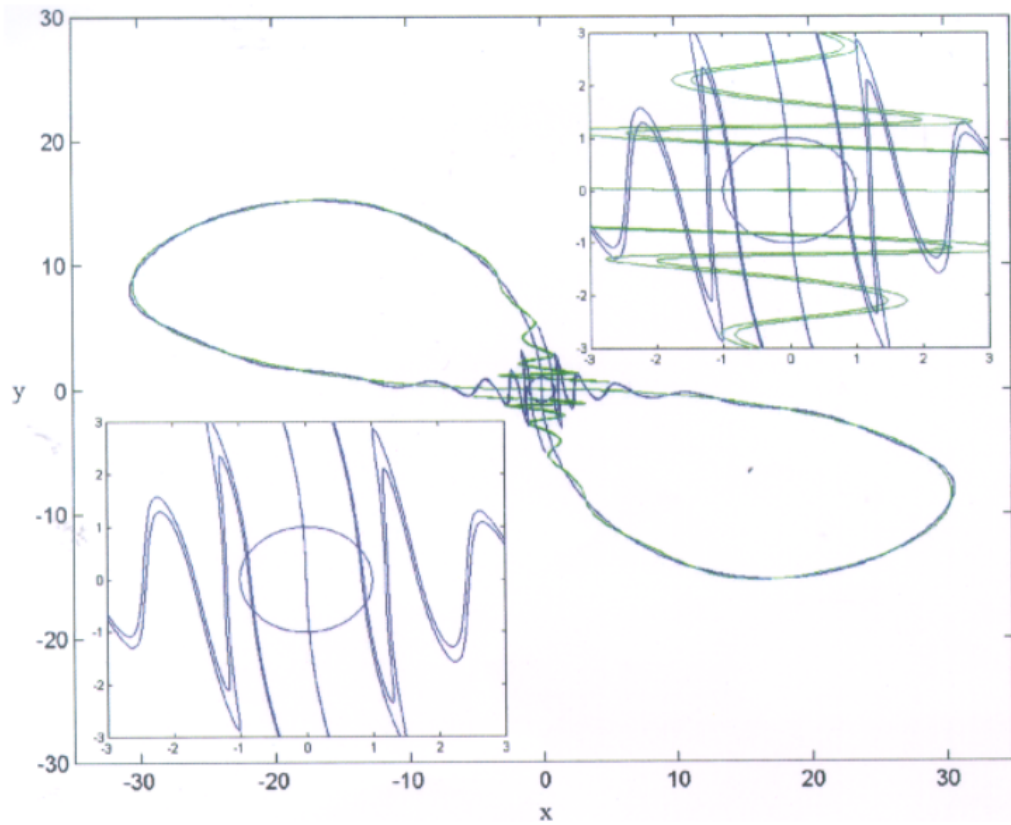


Figure 1: Figure with schematic the sort of curves arising in the Katov Horseshoe. Taken from International Journal of Bifurcation and Chaos, Vol. 9, No. 10 (1999)