Ergodic Theory - A Summary

Ergodic Theory is the study of dynamical systems by means of invariant measures.

Definition: A measure is T-invariant if $\mu(T^{-1}(A)) = \mu(A)$ for every set A in the algebra of μ-measurable sets.

Equivalently: for every measurable function $f : X \to \mathbb{R}$,

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

A measure is called non-atomic if $\mu(\{x\}) = 0$ for every $x \in X$.

A measure is called absolutely continuous (w.r.t. Lebesgue m) if $m(A) = 0$ implies $\mu(A) = 0$.

Existence of Invariant Measures

The existence of invariant measures is (usually) guaranteed by the:

Theorem of Krylov-Bogol'ubov: If $T : X \to X$ is a continuous map on a nonempty compact metric space X, then the set of invariant probability measures $\mathcal{M}(X, T) \neq \varnothing$.

Examples:

- If $T^p(x) = x$, then the equidistribution $\frac{1}{p}\sum_{j=0}^{p-1}\delta_{T^j(x)}$ is an invariant measure.
- If $P_T f(x) = \sum_{Ty=x} \frac{f(y)}{|DT(y)|}$ is the transfer operator w.r.t. Lebesgue measure, then

$$d\mu = f\,dx \text{ is invariant if } P_T f = f.$$

Ergodicity

Definition: A measure μ for a dynamical system (X, T) is ergodic if

$$\mu(A) = 0 \text{ or } \mu(A^c) = 0$$

for every measurable set $A \subset X$ such that $T^{-1}(A) = A \bmod \mu$.

This says that (modμ), the space doesn't decompose into parts that don't communicate with each other.

Equivalent expression for ergodicity are:

- The only T-invariant functions $\psi \in L^1(\mu)$ *i.e.,* $\psi \circ T = \psi$ μ-a.e., are constant μ-a.e.
- $\frac{1}{n}\sum_{j=0}^{n} \mu(A \cap T^j B) - \mu(A)\mu(B) \to 0$ for all measurable sets $A, B \subset X$.

Ergodicity

Examples:

- The doubling map $T : \mathbb{S}^1 \to \mathbb{S}^1$, $x \mapsto 2x \bmod 1$ preserves Lebesgue measure m, and it is ergodic. However, $\frac{1}{2}m + \frac{1}{2}\delta_0$ is invariant but not ergodic.
- The Gauß map $G : [0,1) \to (0,1]$, $x \mapsto \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ preserves the measure $d\mu = \frac{1}{\log 2} \frac{dx}{1+x}$, and it is ergodic (Folklore Theorem).
- Circle rotations

$$R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1, \qquad R_\alpha(x) = x + \alpha \bmod 1.$$

  preserve Lebesgue measure.
  - If $\alpha \in \mathbb{Q}$, then every orbit is periodic. Lebesgue measure is not ergodic.
  - If $\alpha \notin \mathbb{Q}$, then every orbit is dense in $\mathbb{S}^1$. Lebesgue measure is ergodic; in fact it is the only R_α-invariant probability measure.

Unique Ergodicity

Definition: A system (X, T) is called uniquely ergodic if there is exactly one T-invariant probability measure.

This measure is automatically ergodic.

Oxtoby's Theorem: Let X be a compact space and $T : X \to X$ continuous. A transformation (X, T) is uniquely ergodic if and only if, for every continuous function ψ and every point $x \in X$, the Birkhoff averages

$$\frac{1}{n}\sum_{i=0}^{n-1} \psi \circ T^i(x)$$

converge uniformly to a constant.

Birkhoff's Ergodic Theorem

Birkhoff's Ergodic Theorem formalizes a frequent observation in physics:

Space Average = Time Average (for typical points).

This is expressed in:

Birkhoff's Ergodic Theorem: Let μ be a probability measure and $\psi \in L^1(\mu)$. Then the ergodic average

$$\psi^*(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ T^i(x)$$

exists μ-a.e., and ψ^* is T-invariant, i.e., $\psi^* \circ T = \psi^*$ μ-a.e.
If in addition μ is ergodic then

$$\psi^* = \int_X \psi \, d\mu \qquad \mu\text{-a.e.}$$

Absolutely Continuous Measures

Definition: A measure μ is called absolutely continuous w.r.t. the measure ν (notation: $\mu \ll \nu$) if $\nu(A) = 0$ implies $\mu(A) = 0$. If both $\mu \ll \nu$ and $\nu \ll \mu$, then μ and ν are called equivalent.

Theorem of Radon-Nikodym: If μ is a probability measure and $\mu \ll \nu$ then there is a function $h \in L^1(\nu)$ (called Radon-Nikodym derivative or density) such that $\mu(A) = \int_A h(x)\, d\nu(x)$ for every measurable set A.

Notation: $h(x) = \frac{d\mu(x)}{d\nu(x)}$.

Suppose that $\mu \ll \nu$ are both T-invariant probability measures, with a common σ-algebra $\mathcal{B}$ of measurable sets. If ν is ergodic, then $\mu = \nu$.

Poincaré Recurrence

The Poincaré Recurrence Theorem: If (X, T, μ) is a measure preserving system with $\mu(X) = 1$, then for every measurable set $Y \subset X$ of positive measure, μ-a.e. $y \in Y$ returns to Y, i.e., the first return time to Y: $\tau_Y(y) < \infty$ μ-a.e.

Definition: A system $(X, T, \mathcal{B}, \mu)$ is called conservative if for every set $A \in \mathcal{B}$ with $\mu(A) > 0$, there is $n \geq 1$ such that $\mu(T^n(A) \cap A) > 0$. The Poincaré Recurrence Theorem thus implies that probability measure preserving systems are conservative.

If not conservative, then the system is called dissipative. It is called totally dissipative if for every set $A \in \mathcal{B}$,

$$\mu(\{x \in A : T^n(x) \in A \text{ infinitely often}\}) = 0.$$

Kac' Lemma

Kac's Lemma quantifies the expected value of the first return time τ_Y to $Y \subset X$.

Kac' Lemma: Let (X, T) preserve an ergodic measure μ. Take $Y \subset X$ measurable such that $\mu(Y) > 0$, and let $\tau : Y \to \mathbb{N}$ be the first return time to Y. Take $Y \subset X$ measurable such that $\mu(Y) > 0$. Then

$$\mathbb{E}_\mu(\tau_Y) = \int_Y \tau_Y \, d\mu = \sum_{n \geq 1} n\mu(Y_n) = \mu(X)$$

for $Y_n := \{y \in Y : \tau(y) = n\}$.

Induced Systems

Proposition: Let $(X, \mathcal{B}, T, \mu)$ be an ergodic dynamical system and $Y \in \mathcal{B}$ a set with $\mu(Y) > 0$. Let $T_Y = T^{\tau_Y}$ be the first return map to Y.

If μ is T-invariant, then $\nu(A) := \frac{1}{\mu(Y)}\mu(A \cap Y)$ is T_Y-invariant.

Conversely, if ν is T_Y-invariant, and

$$\Lambda := \int_Y \tau(y) d\nu < \infty,$$

then

$$\mu(A) = \frac{1}{\Lambda} \sum_{j=1}^{\infty} \nu(T^{-j}(A) \cap \{y \in Y : \tau(y) \geq j\})$$

is a T-invariant probability measure. Moreover μ is ergodic for T if and only if ν is ergodic for T_Y.

Isomorphic Systems

Definition: Two measure preserving dynamical systems $(X, \mathcal{B}, T, \mu)$ and $(Y, \mathcal{C}, S, \nu)$ are called isomorphic if there are $X' \in \mathcal{B}$, $Y' \in \mathcal{C}$ and $\phi : Y' \to X'$ such that

- $\mu(X') = 1$, $\nu(Y') = 1$;
- $\phi : Y' \to X'$ is a bi-measurable bijection;
- ϕ is measure preserving: $\nu(\phi^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$.
- $\phi \circ S = T \circ \phi$.

That is, the below diagram commutes, and $\phi : Y \to X$ is one-to-one almost everywhere.

$$\begin{array}{ccc} (Y, \mathcal{C}, \nu) & \xrightarrow{S} & (Y, \mathcal{C}, \nu) \\ \phi \downarrow & & \downarrow \phi \\ (X, \mathcal{B}, \mu) & \xrightarrow{T} & (X, \mathcal{B}, \mu) \end{array}$$

The Bernoulli Property

Definition: Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving dynamical system.

1. If T is invertible, then the system is called Bernoulli if it is isomorphic to a 2-sided Bernoulli shift.
2. If T is non-invertible, then the system is called one-sided Bernoulli if it is isomorphic to a 1-sided Bernoulli shift.
3. If T is non-invertible, then the system is called Bernoulli if its natural extension is isomorphic to a 2-sided Bernoulli shift.

Mixing

Definition: A probability measure preserving dynamical system $(X, \mathcal{B}, \mu, T)$ is mixing (or strong mixing) if

$$\mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B) \text{ as } n \to \infty$$

for every $A, B \in \mathcal{B}$.

This says that the "events" A and B are asymptotically independent.

Equivalently, a probability preserving dynamical system $(X, \mathcal{B}, T, \mu)$ is mixing if and only if

$$\int_X f \circ T^n(x) \cdot \overline{g(x)}\, d\mu \to \int_X f(x)\, d\mu \cdot \int_X \overline{g(x)}\, d\mu \text{ as } n \to \infty$$

for all $f, g \in L^2(\mu)$

Measure-Theoretic Entropy

Given a finite partition $\mathcal{P}$ of a probability space (X, μ), let

$$H_\mu(\mathcal{P}) = -\sum_{P \in \mathcal{P}} \mu(P) \log(\mu(P)). \tag{1}$$

For a T-invariant probability measure μ on $(X, \mathcal{B}, T)$, and a partition $\mathcal{P}$, define the entropy of μ w.r.t. $\mathcal{P}$ as

$$h_\mu(T, \mathcal{P}) = \lim_{n \to \infty} \frac{1}{n} H_\mu\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}\right). \tag{2}$$

Finally, the measure theoretic entropy of μ is

$$h_\mu(T) = \sup\{h_\mu(T, \mathcal{P}) \ : \ \mathcal{P} \text{ is a finite partition of } X\}. \tag{3}$$

Measure-Theoretic Entropy

Remarks concerning measure-theoretic entropy

- The existence of the limit in (2) depends on:

  Fekete's Lemma: If $(a_n)_{n\geq 1}$ is subadditive, then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{q\geq 1} \frac{a_q}{q}.$$

- By Sinaĭ's Theorem, instead of taking the supremum over all partitions, it suffices to take a generating partition.
- Entropy is preserved under isomorphism (and is non-increasing under taking measure-theoretical factors).
- The entropy of the $(p_1, \dots, p_n)$-Bernoulli shift (both one-sided and two-sided) is $h(\mu_p) = -\sum_i p_i \log p_i$.

The Shannon-Breiman-McMillan Theorem

The Shannon-Breiman-McMillan Theorem uses entropy to measure how large sets in the n-th joint $\mathcal{P}_n$ are. Typically, they decrease exponentially and the exponential rate is exactly the measure-theoretical entropy.

Shannon-McMillan-Breiman Theorem: Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving transformation and $\mathcal{P}$ a (countable or finite) partition with $H(\mathcal{P}) < \infty$ Let $\mathcal{P}_n = \bigvee_{k=0}^{n-1} T^{-k}(\mathcal{P})$ and $\mathcal{P}_n(x)$ the element of $\mathcal{P}_n$ containing x. Then

$$-\lim_{n\to\infty} \frac{1}{n} \log \mu(\mathcal{P}_n(x)) = h(\mathcal{P}, T) \quad \mu\text{-a.e.}$$

Topological Entropy

Definition: Topological entropy can be defined as the xponential growth rate of the

- (Adler-Konheim-McAndrew) minimal cardinality of subcovers of joints:

$$h_{top}(T) = \lim_{\varepsilon\to 0} \sup_{\mathcal{U}} \lim_{n\to\infty} \frac{1}{n} \log \mathcal{N}(\mathcal{U}^n), \tag{4}$$

  where the supremum is taken over all open ε-covers $\mathcal{U}$.

- (Bowen) maximal cardinalty of n, ε-separated sets:

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon). \tag{5}$$

- (Bowen) minimal cardinalty of n, ε-spanning sets:

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon). \tag{6}$$

Topological Entropy

For maps T on the interval we have:

Theorem of Szlenk & Misiurewicz Let $T : [0,1] \to [0,1]$ has finitely many laps. Then

$$\begin{aligned} h_{top}(T) &= \lim_{n\to\infty} \frac{1}{n} \log \ell(T^n) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \#\{\text{clusters of } n\text{-periodic points}\} \\ &= \max\{0, \lim_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n)\}. \end{aligned}$$

where two n-periodic points are in the same cluster if they belong to the same lap of T^n.

The Variational Principle

Topological and measure-theoretical entropy are related by the Variational Principle which say that (for continuous map on compact metric space)

$$h_{top}(T) = \sup\{h_\mu(T) : \mu \text{ is } T\text{-invariant probability measure}\}$$

If μ is such that $h_\mu(T) = h_{top}(T)$, then μ is called a measure of maximal entropy.

If there is a unique measure of maximal entropy $\mu_{\max}$, then (X, T) is called intrinsically ergodic. In this case, $\mu_{\max}$ is ergodic.

The Variational Principle

- The full shift on N symbols (one-sided or two-sided) has entropy $\log N$, and the measure of maximal entropy is then$(\frac{1}{N}, \dots, \frac{1}{N})$-Bernoulli measure.
- A subshift of finite type with transition matrix A has the logarithm of the leading eigenvalue as entropy. The Parry measure is the measure of maximal entropy.
- Interval maps with constant slope $\pm s$ for $s > 1$ have an absolutely continuou smeasure, which is the measure of maximal entropy.

Mixing

Review of ergodic properties:

- Bernoulli if it is isomorphic to a two-sided Bernoulli shift.
- strong mixing if for all $A, B \in \mathcal{B}$

$$\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B) \to 0.$$

- weak mixing if for all $A, B \in \mathcal{B}$ the average

$$\frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)| \to 0.$$

- ergodic if $T^{-1}(A) = A \bmod \mu$ implies $\mu(A) = 0$ or $\mu(A^c) = 0$.
- conservative if for all $A \in \mathcal{B}$ with $\mu(A) > 0$ there is $n \geq 1$ such that $\mu(T^n(A) \cap A) > 0$.

Bernoulli - Mixing - Ergodic - Conservative

Theorem We have the implications:

Bernoulli ⇒ mixing ⇒ weak mixing ⇒ ergodic ⇒ conservative.

None of the reverse implications holds.

Classifying Systems up to Isomorphism

Definition: Two dynamical systems (X, T) and (Y, S) are topologically conjugate if there is a homeomorphism $\phi : Y \to X$ such that $\phi \circ S = T \circ \phi$.

We can classify dynamical systems up to conjugacy, and measure preserving systems up to isomorphism.
The one does not imply the other or vice versa:

- The doubling map (with Lebesgue measure) and the one-sided $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift are isomorphic. They are not conjugate (note: they are defined on non-homeomorphic spaces).
- The doubling map with Lebesgue measure is conjugate to the doubling map with δ_0, but they are not isomorphic. (Conjugacies are topological objects; they don't care about measures).

Classifying Systems up to Isomorphism

The following properties and quantities are preserved under isomorphisms.

- ergodicity, weak and strong mixing, the Bernoulli property
- measure-theoretic entropy. Moreover,

  Ornstein's Theorem: Two two-sided Bernoulli shifts (X, μ_p, σ) and $(X', \mu_{p'}, \sigma)$ are isomorphic if and only if $h(\mu_p) = h(\mu_{p'})$.

  This fails for one-sided Bernoulli shifts.
- eigenvalues of the Koopman operator $U_T f = f \circ T$. Moreover,

  Theorem (Halmos & Von Neumann) Two measure-preserving dynamical systems with pure point spectra are isomorphic if and only if their eigenvalues are the same.

  This fails without the assumption of pure point spectrum.

Toral Automorphims

Definition: A toral automorphism $T : \mathbb{T}^d \to \mathbb{T}^d$ is an invertible linear map on the (d-dimensional) torus $\mathbb{T}^d$. Each such T is of the form $T_A(x) = Ax \pmod 1$, where the matrix A satisfies:

- A is an integer matrix with $\det(A) = \pm 1$;
- To avoid degenerate examples including $A = Id$, we assume that A is primitive, i.e., A^n is strictly positive for some $n \geq 1$.
- If the eigenvalues of A are not on the unit circle, then the toral automorphism is called hyperbolic.

Toral Automorphims

Properties of the toral automorphisms T_A:

- A preserves the integer lattice $\mathbb{Z}^d$, so T_A is well-defined and continuous.
- $\det(A) = \pm 1$, so Lebesgue measure m is preserved (both by A and T_A). Also A and T_A are invertible, and A^{-1} is still an integer matrix (so T_A^{-1} is well-defined and continuous too).
- One can show that Lebesgue measure is ergodic if and only if A has no eigenvalues that are roots of unity.
- Hyperbolic toral automorphisms have a Markov partition w.r.t. which the symbolic dynamics is a subshift of finite type, and Lebesgue measure is the measure of maximal entropy.
- Hyperbolic toral automorphisms are mixing w.r.t. Lebesgue.