§2. Ergodic Theorems

Our setup here will be:

- X a metric space (can be relaxed to just a measure space...)
- $T: X \to X$ (Borel) measurable map
- μ a T-invariant probability measure on X

For $f: X \to \mathbb{R}$ measurable,
we study the ergodic averages

$$\langle f \rangle_n (x) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)), \quad x \in X$$

An ergodic theorem would state that under certain assumptions on T, μ,

$$\langle f \rangle_n \longrightarrow \int_X f \, d\mu$$

in some sense for some class of functions f

function of $x \in X$ constant

i.e. time average (over a trajectory of T)

$\big\downarrow$ time $\to \infty$

Space average (w.r.t. the measure μ)

§2.1. The von Neumann ergodic theorem

We start with an easier to prove version of ergodic theorem, in the space $L^2(X,\mu)$.

Define the subspace of $L^2(X,\mu)$

$$\mathrm{Inv} := \{ f \in L^2(X,\mu) \mid f \circ T = f \ \ \mu\text{-almost everywhere} \}$$

Note that Inv is a closed subspace of $L^2(X,\mu)$ as will be seen in a moment.

Let $P : L^2(X,\mu) \to \mathrm{Inv}$ be the orthogonal projector onto Inv.

Thm [L^2 Ergodic Theorem]

For each $f \in L^2(X,\mu)$ we have

$$\langle f \rangle_n \longrightarrow Pf \ \text{ in } L^2(X,\mu).$$

Proof Define the linear operator

$$U : L^2(X,\mu) \circlearrowleft, \quad Uf = f \circ T.$$

Then $\|Uf\|_{L^2(X,\mu)} = \|f\|_{L^2(X,\mu)}$ $\forall f \in L^2(X,\mu)$

Since μ is T-invariant:

$$\int_X |f \circ T|^2 \, d\mu = \int_X |f|^2 \, d\mu.$$

Define next the operator

$$B_n : L^2(X,\mu) \circlearrowleft, \quad B_n := \frac{1}{n}\sum_{j=0}^{n-1} U^j,$$

so that $\langle f \rangle_n = B_n f$.

Note that $\|B_n\|_{L^2(X,\mu)\circlearrowleft} \le 1$.

We have $\boxed{\mathrm{Inv} = \ker(\mathrm{Id} - U)}$

which is closed since U is a bounded operator on $L^2(X,\mu)$.

We have (by the Orthogonal Projection Thm from functional analysis)

$$L^2(X,\mu) = \mathrm{Inv} \oplus \mathrm{Inv}^\perp$$

where $\mathrm{Inv}^\perp = \{ f \in L^2(X,\mu) \mid \langle f, g \rangle_{L^2(X,\mu)} = 0 \ \forall g \in \mathrm{Inv} \}$.

For $f \in \mathrm{Inv}$, we have $Uf = f$ and thus $B_n f = f \Rightarrow B_n f \to f$ in $L^2(X,\mu)$

So it suffices to show that

(*) $B_n f \to 0$ in $L^2(X,\mu)$ for all $f \in \mathrm{Inv}^\perp$.

Define the range of $\mathrm{Id} - U$:

$$\mathrm{Ran}(\mathrm{Id} - U) = \{ f - Uf \mid f \in L^2(X,\mu) \}$$

We show $(*)$ in 2 steps:

(1) $B_n f \to 0$ in $L^2(X,\mu)$ for all $f \in \mathrm{Ran}(\mathrm{Id}-U)$

(2) $\mathrm{Ran}(\mathrm{Id}-U)$ is <u>dense</u> in $\mathrm{Inv}^\perp$.

Together (1)+(2) imply $(*)$ by the Lemma in §1.1.

(1) Assume $f \in \mathrm{Ran}(\mathrm{Id}-U)$, that is $f = g - Ug$ for some $g \in L^2(X,\mu)$.

Then
$$B_n f = \frac{1}{n}\sum_{j=0}^{n-1} U^j(g - Ug)$$
$$= \frac{1}{n}\left[\sum_{j=0}^{n-1} U^j g - \sum_{j=0}^{n-1} U^{j+1} g\right]$$
$$= \frac{1}{n}(g - U^n g) \xrightarrow[n\to\infty]{} 0 \text{ in } L^2(X,\mu),$$
since $\|U^n g\|_{L^2} = \|g\|_{L^2}$ $\forall n$.

(2) It suffices to show that $\mathrm{Ran}(\mathrm{Id}-U)^\perp \subset \mathrm{Inv}$ where
$$\mathrm{Ran}(\mathrm{Id}-U)^\perp = \{f \in L^2(X,\mu) \mid \langle f, g - Ug\rangle_{L^2(X,\mu)} = 0 \ \forall g \in L^2(X,\mu)\}$$

Indeed, if $h \in L^2(X,\mu)$ does not lie in the closure $\overline{\mathrm{Ran}(\mathrm{Id}-U)}$ then (by the Orthogonal Complement Thm)

we can find $f \in L^2(X,\mu)$ which is orthogonal to $\mathrm{Ran}(\mathrm{Id}-U)$ and $\langle f, h\rangle_{L^2(X,\mu)} = 1$.

If $\mathrm{Ran}(\mathrm{Id}-U)^\perp \subset \mathrm{Inv}$ then $f \in \mathrm{Inv} \Rightarrow h$ cannot lie in $\mathrm{Inv}^\perp$.

Now, to show that $\mathrm{Ran}(\mathrm{Id}-U)^\perp \subset \mathrm{Inv}$, we take $f \in \mathrm{Ran}(\mathrm{Id}-U)^\perp$,

So that $\langle f, g-Ug\rangle_{L^2(X,\mu)} = 0 \quad \forall g \in L^2(X,\mu)$.

Taking $g := f$, we get $\langle f, f-Uf\rangle_{L^2(X,\mu)} = 0$,

So $\|f\|^2_{L^2(X,\mu)} = \langle f, Uf\rangle_{L^2(X,\mu)}$.

Now $\|f-Uf\|^2_{L^2(X,\mu)} = \|f\|^2_{L^2} + \|Uf\|^2_{L^2} - 2\langle f, Uf\rangle_{L^2}$

$= 2(\|f\|^2_{L^2} - \langle f, Uf\rangle_{L^2}) = 0$, so $Uf = f$ and $f \in \mathrm{Inv}$ as needed. $\square$

since $\|Uf\|_{L^2} = \|f\|_{L^2}$

Remarks.

(1.) Most of the time, the range $\mathrm{Ran}(I-U)$ is <u>not</u> closed in L^2.

E.g. the irrational shift of §1.1:
if $e_k(x) = e^{2\pi i k x}$, $k \in \mathbb{Z}$, then
$U e_k = e^{2\pi i k r} e_k$ (here $T(x) = (x+r) \bmod \mathbb{Z}$)

If $f = \sum_{k \in \mathbb{Z}\setminus\{0\}} c_k e_k \in \mathrm{Inv}^\perp = \overline{\mathrm{Ran}(I-U)}$ (here $\mathrm{Inv} = \{\text{constants}\}$)

then $f \in \mathrm{Ran}(I-U)$ iff $\sum_{k \in \mathbb{Z}\setminus\{0\}} \left|\frac{c_k}{e^{2\pi i k r}-1}\right|^2 < \infty$

and this is a stronger condition than $\sum_{k \in \mathbb{Z}\setminus\{0\}} |c_k|^2 < \infty$
since $e^{2\pi i k r} - 1$ can get arbitrarily small as $k \to \infty$.

(2) If T is not invertible then U need not be unitary (i.e. might not have $U^{-1} = U^*$)

<u>Basic example:</u> $X = S^1 = \mathbb{R}/\mathbb{Z}$, $T(x) = (2x) \bmod \mathbb{Z}$, $\mu =$ Lebesgue measure. Then

$Uf(x) = f(2x \bmod \mathbb{Z})$ but

$U^* f(x) = \frac{1}{2}\left(f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right)\right)$ (transfer operator)

See Pset 1.

An important case is when

$$\mathrm{Inv} = \{\text{constant functions}\} = \mathrm{Span}(\mathbb{1}).$$

Then T is called <u>ergodic</u> w.r.t. μ

The L^2 ergodic thm implies

<u>Corollary</u> If μ is ergodic w.r.t. T then for each $f \in L^2(X,\mu)$ we have

$$\langle f \rangle_n \to \int_X f \, d\mu \quad \text{in } L^2(X,\mu).$$

<u>Proof</u> If $\mathrm{Inv} = \mathrm{Span}(\mathbb{1})$ then the orthogonal projector onto Inv is

$$Pf(x) = \langle f, \mathbb{1} \rangle_{L^2(X,\mu)} \mathbb{1} = \int_X f \, d\mu. \quad \square$$

<u>Exercise (no credit)</u>: show that corollary holds with $L^2(X,\mu)$ replaced by $L^p(X,\mu)$ for <u>any</u> p, $1 \le p < \infty$.

What about $p = \infty$?

We now discuss
<u>equivalent definitions of ergodicity</u>:

<u>Prop.</u> Let $T: X \to X$ be measurable, μ a T-invariant prob. measure on X.
TFAE:

① $\{f \in L^2(X,\mu) \mid f = f \circ T \ \mu\text{-almost everywhere}\} = \{\text{constant functions a.e.}\}$

② If $A \subset X$ is a Borel set and $A = T^{-1}(A)$ then $\mu(A) = 0$ or $\mu(A) = 1$.
(Cannot split X μ-nontrivially in a T-invariant way.)

③ If $A \subset X$ is a Borel set and $\mu(A \,\Delta\, T^{-1}(A)) = 0$ where $A \Delta B = (A \setminus B) \cup (B \setminus A)$
then $\mu(A) = 0$ or $\mu(A) = 1$.

If ①-③ hold then we say that μ is an <u>ergodic</u> measure for T, or T is <u>ergodic</u> with respect to μ.

Proof 1 $\Rightarrow$ 2:

Assume $A \subset X$ is a Borel set and $A = T^{-1}(A)$. Then take
$f = 1_A$ (the indicator fn. of A).
We see that $f \in L^2(X, \mu)$ and
$f \circ T = f$, so $\exists c_1 \in \mathbb{R}$:
$f = c_1$ μ-almost everywhere.
Thus $\mu(A) = 0$ or $\mu(A) = 1$.

2 $\Rightarrow$ 3: Assume A is a Borel set and $\mu(A \triangle T^{-1}(A)) = 0$.
Define the Borel set $B \subset X$ as follows:
$x \in X$ lies in $B \iff T^n(x) \in A$ for all sufficiently large n.

Then $x \in B \iff T(x) \in B$,
so $B = T^{-1}(B)$. Thus $\mu(B) = 0$ or $\mu(B) = 1$.

On the other hand, writing

$$B = \bigcup_{m \geq 0} \bigcap_{n \geq m} T^{-n}(A)$$

and using that $\mu(A \Delta T^{-1}(A)) = 0$

and thus $\mu(A \Delta T^{-n}(A)) = 0 \quad \forall n$

(since $\mu(T^{-n}(A) \Delta T^{-n-1}(A)) =$

$= \mu(T^{-n}(A \Delta T^{-1}(A)))$

$= \mu(A \Delta T^{-1}(A)) = 0 \quad \forall n$)

we see that $\mu(A \Delta B) = 0$,

thus $\mu(B) = 0$ or $\mu(B) = 1$.

3 => 1: Assume that $f \in L^2(X, \mu)$

and $f \circ T = f$ μ-almost everywhere.

Then for each $c \in \mathbb{R}$, the set

$$A_c := \{x \in X \mid f(x) \leq c\}$$

satisfies $\mu(A_c \Delta T^{-1}(A_c)) = 0$.

Thus $\forall c$, $\mu(A_c) = 0$ or $\mu(A_c) = 1$.

This implies that $f =$ constant

μ-almost everywhere. $\square$

§2.2. Example: an expanding map

Let us again take $X = S^1 = \mathbb{R}/\mathbb{Z}$ and consider the map $T: X \to X$ given by

$$T([x]) = 2x \mod \mathbb{Z}, \quad [x] \in \mathbb{R}/\mathbb{Z}.$$

<u>Thm</u>. The map T is ergodic w.r.t. the Lebesgue measure μ on $[0,1]$.

<u>Proof</u> μ is T-invariant: see Pset 1.

Enough to show that $\forall f \in L^2(X,\mu)$,

$$\langle f \rangle_n \to \int_X f \, d\mu \quad \text{in } L^2(X,\mu)$$

(indeed, can apply this to any $f \in L^2(X,\mu)$ such that $f = f \circ T$ μ-almost everywhere and get $f = \text{const}$ μ-almost everywhere)

As in §1.1, enough to consider the case when $f(x) = e_\ell(x) = e^{2\pi i \ell x}$, $\ell \in \mathbb{Z}$.

For $\ell = 0$, get $f \equiv 1 \Rightarrow \langle f \rangle_n \equiv 1$.

For $\ell \neq 0$, get

$$\langle e_\ell \rangle_n = \frac{1}{n} \sum_{j=0}^{n-1} e_\ell(2^j x)$$

$$= \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell \cdot 2^j x} = \frac{1}{n} \sum_{j=0}^{n-1} e_{2^j \cdot \ell}.$$

Now $\|\langle e_\ell \rangle_n\|^2_{L^2(X,\mu)} = \frac{1}{n^2} \sum_{j=0}^{n-1} 1 = \frac{1}{n}$

as $e_{2^j \cdot \ell}$ form an orthonormal system in $L^2(X,\mu)$.

So $\langle e_\ell \rangle_n \xrightarrow{n\to\infty} 0$ in $L^2(X,\mu)$ as needed. $\square$

However, the convergence above is <u>not pointwise</u> (even for $f = e_\ell$) since T is not uniquely ergodic.

To see this, we note that T has a <u>fixed point</u> at 0: $T(0)=0$.

Take $\mu_0 = \delta_0 \leftarrow$ delta measure at 0

Then μ_0 is T-invariant: $\forall$ Borel $A \subset X$

$$\mu_0(T^{-1}(A)) = \begin{cases} 1, & 0 \in T^{-1}(A) \\ 0, & \text{else} \end{cases}$$

$$\mu_0(A) = \begin{cases} 1, & 0 \in A \\ 0, & \text{else} \end{cases}$$

In fact, we can easily see that

$$\forall f \in C^0(X), \quad \langle f \rangle_n(0) \xrightarrow[n\to\infty]{} f(0)$$

but (as will follow from the Birkhoff ergodic theorem below)

$$\langle f \rangle_n(x) \xrightarrow{n\to\infty} \int_0^1 f(x)\,dx$$

for Lebesgue almost every x.

More generally, if we have a <u>periodic orbit</u> $\gamma = \{x_0, x_1, \ldots, x_{m-1}\}$ of T, i.e. $x_0 \xmapsto{T} x_1 \xmapsto{T} \cdots \xmapsto{T} x_{m-1} \xmapsto{T} x_0$

then the measure $\delta_\gamma = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{x_k}$

is <u>T-invariant</u> and <u>ergodic</u> (exercise, no credit)

18.118
2-14

What are the periodic orbits of T?

Need to find X, m such that

$\boxed{T^m(x) = x}$, i.e.

$2^m x - x \in \mathbb{Z}$, that is

$x \in \dfrac{\mathbb{Z}}{2^m - 1}$. For each $m \geq 1$

there are exactly $2^m - 1$ such points (though some of these lie on shorter periodic orbits)

and we see in particular that for the specific $T(x) = [2x] \bmod \mathbb{Z}$, the set of periodic points is <u>dense</u>.

§2.3. The almost everywhere ergodic theorem

Here we prove

Thm [Birkhoff] Let X be a metric space, $T: X \to X$ a Borel measurable map, and μ an ergodic T-invariant probability measure on X.

Then for each $f \in L^1(X, \mu)$ we have

$$\langle f \rangle_n (x) \overset{\text{def}}{=} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \xrightarrow[n \to \infty]{} \int_X f \, d\mu$$

for μ-almost every $x \in X$.

Remarks ① A common interpretation is that a μ-typical trajectory equidistributes according to μ.

② The general version of the Thm does not need μ to be ergodic, and $\int_X f d\mu$ is replaced by a certain projector applied to f.

The proof of Birkhoff's Ergodic Thm in these notes is short but hard to absorb. For a longer but easier to read proof, see e.g. Coudène, "Ergodic Theory & Dynamical Systems"

Our proof relies on the following key

Lemma Let X, T, μ be as in the Thm and $f \in L^1(X,\mu)$ satisfy $\boxed{\int_X f \, d_\mu < 0}$.

Then for μ-almost every x, we have

$$\boxed{\limsup_{n\to\infty} \langle f \rangle_n (x) \le 0.}$$

Proof that Lemma $\Rightarrow$ Thm

Let $f \in L^1(X,\mu)$ and put $I := \int_X f d_\mu \in \mathbb{R}$.

Take arbitrary $N \in \mathbb{N}$, then

$$\int_X \left(f - I - \frac{1}{N}\right) d\mu = -\frac{1}{N} < 0.$$

Applying the Lemma to $f - I - \frac{1}{N}$ we see that there is a set $A_N \subset X$, $\mu(A_N) = 0$, such that

$$\forall x \in X \setminus A_N, \quad \limsup_{n\to\infty} \langle f\rangle_n(x) \le I + \frac{1}{N}.$$

Take $A_\infty := \bigcup_N A_N$, then $\mu(A_\infty) = 0$ and $\forall x \in X \setminus A_\infty \ \forall N \ \limsup_{n\to\infty} \langle f\rangle_n(x) \le I + \frac{1}{N}$.

That is, for μ-almost every x we have

$$\boxed{\limsup_{n\to\infty} \langle f\rangle_n(x) \le I.}$$

A similar argument (replacing f with $-f$) gives

$$\boxed{\liminf_{n\to\infty} \langle f\rangle_n(x) \ge I}$$

for μ-almost every x.

So for μ-almost every x we have

$$\lim_{n\to\infty} \langle f\rangle_n(x) = I. \qquad \square$$

Proof of Lemma

Assume $f \in L^1(X, \mu)$ and $\int_X f \, d\mu < 0$.

Define the sums

$$\boxed{S_n(x) = \sum_{j=0}^{n-1} f(T^j(x)),}$$

so that $\langle f \rangle_n (x) = \frac{S_n(x)}{n}$.

We have the following identity for all $n \geq 0$:

$$\boxed{S_{n+1}(x) = f(x) + S_n(T(x)).} \quad (*)$$

Define the Borel set $A \subset X$ by

$$\boxed{A = \{x \in X \mid \sup_{n \geq 1} S_n(x) = \infty\}}$$

i.e. $x \in A \iff S_n(x)$ is not bounded from above.

We claim that A is T-invariant:

$$A = T^{-1}(A).$$

Indeed, for each x we have

$$x \in T^{-1}(A) \Leftrightarrow T(x) \in A$$

$$\Leftrightarrow \sup_{n\geq 1} S_n(T(x)) = \infty$$

$$\Leftrightarrow \sup_{n\geq 1} S_{n+1}(x) = \infty$$

$$\Leftrightarrow \sup_{n\geq 2} S_n(x) = \infty \Leftrightarrow x \in A.$$

Now, since A is T-invariant and T is <u>ergodic</u> w.r.t. μ, we have

$\underline{\mu(A)=0}$ or $\underline{\mu(A)=1}$.

If $\boxed{\mu(A)=0}$ then we use that $\forall x \in X \setminus A$, the sequence $S_n(x)$ is bounded above and thus

$$\limsup_{n\to\infty} \langle f\rangle_n(x) = \limsup_{n\to\infty} \frac{S_n(x)}{n} \leq 0,$$

which gives the statement of the Lemma.

So assume now that $\boxed{\mu(A)=1}$,
i.e. $S_n(x)$ is <u>not</u> bounded above for μ-almost every x. We will reach a <u>contradiction</u>.

Define the function for $m \geq 1$,

$$F_m(x) = \max_{1 \leq n \leq m} S_n(x) =$$
$$= \max\left(f(x), f(x)+f(T(x)), \dots, f(x)+\dots+f(T^{m-1}(x))\right).$$

Note that $\boxed{F_m \leq F_{m+1}}$ and

$\forall m$, $F_m \in L^1(X,\mu)$:
indeed, $f \in L^1(X,\mu)$ & μ is T-invariant
$\Rightarrow f \circ T^j \in L^1(X,\mu)$ $\forall j \Rightarrow$
$\Rightarrow S_n \in L^1(X,\mu) \Rightarrow F_m \in L^1(X,\mu)$.
On the other hand, by $(*)$ we have

$$F_{m+1}(x) = \max\left(f(x), \max_{1 \leq n \leq m} S_{n+1}(x)\right)$$
$$= \max\left(f(x), f(x) + \max_{1 \leq n \leq m} S_n(T(x))\right)$$
$$= f(x) + \max\left(0, F_m(T(x))\right).$$

Therefore

$$G_m(x) := F_{m+1}(x) - F_m(T(x))$$

$$= f(x) - \min(0, F_m(T(x))).$$

We see that (since $F_m \geq F_1$)

$$\boxed{f \leq G_m \leq G_1}$$

and $f \in L^1(X,\mu)$, $G_1 \in L^1(X,\mu)$

(since $G_1(x) = f(x) - \min(0, f(T(x))))$

But now, $\forall x \in A$, we have $T(x) \in A \Rightarrow$

$\Rightarrow F_m(T(x)) \nearrow \infty$ as $m \to \infty$

$\Rightarrow F_m(T(x)) > 0$ for large enough m

$\Rightarrow G_m(x) = f(x)$ for large enough m.

In particular, $\boxed{G_m(x) \xrightarrow[m\to\infty]{} f(x)}$

for μ-almost every x (i.e. for $x \in A$)

So by the Dominated Convergence Theorem

$$\boxed{\int_X G_m \, d\mu \xrightarrow[m\to\infty]{} \int_X f \, d\mu < 0.}$$

But $\int_X G_m \, d\mu =$

$= \int_X F_{m+1} - (F_m \circ T) \, d\mu =$ (as $F_m \in L^1$)

$= \int_X F_{m+1} \, d\mu - \int_X F_m \circ T \, d\mu =$ (as μ is T-invariant)

$= \int_X F_{m+1} \, d\mu - \int_X F_m \, d\mu$

$= \int_X (F_{m+1} - F_m) \, d\mu \geq 0$ since $F_m \leq F_{m+1}$.

This gives a contradiction. $\square$

<u>Remark</u>. We have $\int_X S_n \, d\mu = n \int_X f \, d\mu \xrightarrow[n \to \infty]{} -\infty$.

The hard part of the proof was to exclude the possibility that $\sup S_n = \infty$ μ-almost everywhere even though $\int_X S_n \, d\mu \to -\infty$ (which could happen for a general sequence S_n)