§2. Ergodic Theorems

Our setup here will be:
- $X$ a metric space (can be relaxed to just a measure space...)
- $T: X \rightarrow X$ (Borel) measurable map
- $\mu$ a $T$-invariant probability measure on $X$

For $f: X \rightarrow \mathbb{R}$ 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 $T, \mu$,

$$\langle f \rangle_n \xrightarrow{n \to \infty} \int_X f \, d\mu$$

in some sense for some class of functions $f$.

i.e. time average (over a trajectory of $T$) $\downarrow$ time $\rightarrow \infty$

constant of functions $f$

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

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

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

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

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

**Thm [L^2 Ergodic Theorem]**

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

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

**Proof** Define the linear operator

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

Then

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

Since $\mu$ is $T$-invariant,

$$\int_X |f \circ T|^2 \, d\mu = \int_X |f|^2 \, d\mu.$$
Define next the operator
\[ B_n : L^2(X, \mu) \mapsto B_n := \frac{1}{n} \sum_{j=0}^{n-1} U^j, \]
so that \( \langle f \rangle_n = B_n f. \)

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

We have \( \text{Inv} = \ker (\text{Id} - U) \)

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

We have (by the Orthogonal Projection Thm from functional analysis)
\[ L^2(X, \mu) = \text{Inv} \oplus \text{Inv}^+ \]
where \( \text{Inv}^+ = \{ f \in L^2(X, \mu) | \langle f, g \rangle_{L^2(X, \mu)} = 0 \forall g \in \text{Inv} \}. \)

For \( f \in \text{Inv}, \) we have \( U f = f \)

and thus \( B_n f = f \Rightarrow B_n f \rightarrow f \) in \( L^2(X, \mu) \)

So it suffices to show that
(1) \( B_n f \rightarrow 0 \) in \( L^2(X, \mu) \) for all \( f \in \text{Inv}^+. \)

Define the range of \( \text{Id} - U: \)
\[ \text{Ran}(\text{Id}-U) = \{ f - U f \mid f \in L^2(X, \mu) \}? \]
We show (*) in 2 steps:

1. \( B_n f \rightarrow 0 \) in \( L^2(X, \mu) \) for all \( f \in \text{Ran}(\text{Id}-U) \)

2. \( \text{Ran}(\text{Id}-U) \) is dense in \( \text{Inv}^+ \).

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

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

Then

\[
B_n f = \frac{1}{n} \sum_{j=0}^{n-1} Ud (g-Ug)
\]

\[
= \frac{1}{n} \left[ \sum_{j=0}^{n-1} U^j g - \sum_{j=0}^{n-1} U^{j+1} g \right]
\]

\[
= \frac{1}{n} (g - U^ng) \xrightarrow{n \to \infty} 0 \text{ in } L^2(X, \mu),
\]

since \( \|U^ng\|_{L^2} = \|g\|_{L^2} \) for all \( n \).

2. It suffices to show that \( \text{Ran}(\text{Id}-U)^\perp \subset \text{Inv} \), where

\[
\text{Ran}(\text{Id}-U)^\perp = \{ f \in L^2(X, \mu) | \langle f, g-Ug \rangle_{L^2(X, \mu)} = 0 \text{ for all } g \in L^2(X, \mu) \}.\]
Indeed, if $h \notin L^2(X,\mu)$ does not lie in the closure \(\text{Ran} (\text{Id}-U)\) then (by the orthogonal complement theorem) we can find $f \in L^2(X,\mu)$ which is orthogonal to $\text{Ran} (\text{Id}-U)$ and $\langle f, h \rangle_{L^2(X,\mu)} = 1$.

If $\text{Ran} (\text{Id}-U)^+ \subseteq \text{Inv}^+$ then $f \in \text{Inv} \Rightarrow h$ cannot lie in $\text{Inv}^+$.

Now, to show that $\text{Ran} (\text{Id}-U)^+$ we take $f \in \text{Ran} (\text{Id}-U)^+$, so that $\langle f, g-Ug \rangle_{L^2(X,\mu)} = 0 \; \forall g \in L^2(X,\mu)$.

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

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

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

$= 2(\|f\|_{L^2(X,\mu)}^2 - \langle f, Uf \rangle_{L^2(X,\mu)}) = 0$, so $Uf=f$

since $\|Uf\|_{L^2(X,\mu)} = \|f\|_{L^2(X,\mu)}$ and $f \in \text{Inv}$ as needed.$\blacksquare$
Remarks.

1. Most of the time, the range $\text{Ran}(I-U)$ is not closed in $L^2$.
   
   E.g. the irrational shift of $\Sigma_1$:
   
   If $e_k(x) = e^{2\pi ikx}$, $k \in \mathbb{Z}$, then
   
   $Ue_k = e^{2\pi ikx}e_k$ (here $T(x) = (x+r) \mod \mathbb{Z}$)
   
   If $f = \sum C_k e_k \in \text{Inv}^+ = \overline{\text{Ran}(I-U)}$
   
   where $\text{Inv} = \{\text{constants}\}$
   
   Then $f \in \text{Ran}(I-U)$ iff $\sum_{k \in \mathbb{Z} \setminus \{0\}} \left| \frac{C_k}{e^{-2\pi ikr}} \right|^2 < \infty$
   
   and this is a stronger condition than $\sum_{k \in \mathbb{Z} \setminus \{0\}} |C_k|^2 < \infty$
   
   Since $e^{2\pi ikx}e_k$ 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 $U^* = U^t$).

   Basic example: $X = S^1 = \mathbb{R}/\mathbb{Z}$, $T(x) = (2x) \mod \mathbb{Z}$, $\mu$ = Lebesgue measure. Then
   
   $Uf(x) = f(2x \mod \mathbb{Z})$ but
   
   $U^*f(x) = \frac{1}{2}(f(\frac{x}{2}) + f(\frac{x+1}{2}))$ (transfer operator)

   See Pset 1.
An important case is when

\[ \text{Inv} = \{ \text{constant functions} \} = \text{Span}(1). \]

Then \( T \) is called \underline{ergodic} w.r.t. \( \mu \)

The \( L^2 \) ergodic thm implies

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

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

Proof: if \( \text{Inv} = \text{Span}(1) \), then the orthogonal projector onto \( \text{Inv} \) is

\[ P f(x) = \langle f, 1 \rangle_{L^2(X,\mu)} \times 1 = \int_X f \, d\mu. \]

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

What about \( p = \infty \)?
We now discuss equivalent definitions of ergodicity.

Prop. Let $T: X \to X$ be measurable, $\mu$ a $T$-invariant prob. measure on $X$.

TFAE:

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

2. If $A \subset X$ is a Borel set and $A = T^{-1}(A)$ then $\mu(A) = 0$ or $\mu(A) = 1$.
   (Cannot split $X$ $\mu$-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 $\mu$ is an ergodic measure for $T$, or $T$ is ergodic with respect to $\mu$. 
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 \( \exists c \in \mathbb{R} : f = c \) \( \mu \)-almost everywhere.
Thus \( \mu(A) = 0 \) or \( \mu(A) = 1 \).

\( 2 \Rightarrow 3 \): Assume \( A \) is a Borel set and \( \mu(A \Delta T^{-1}(A)) = 0 \).
Define the Borel set \( B \subset X \) as follows:
\[ x \in X \text{ lies in } B \iff T^n(x) \in A \text{ for all sufficiently large } n. \]
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
\[
B = \bigcup_{m \geq 0} T^{-m}(A)
\]
and using that \(\mu(A \Delta T^{-1}(A)) = 0\)
and thus \(\mu(A \Delta T^{-m}(A)) = 0\) \(\forall n\)
(since \(\mu(T^{-m}(A) \Delta T^{-m-1}(A)) = \mu(T^{-m}(A) \Delta T^{-m}(A)) = \mu(A \Delta T^{-1}(A)) = 0\) \(\forall n\))

we see that \(\mu(A \Delta B) = 0\),
thus \(\mu(B) = 0\) or \(\mu(B) = 1\).

\(\circ \Rightarrow 1\): Assume that \(f \in L^2(X, \mu)\)
and \(f \circ T = f\) \(\mu\)-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 \(f = \text{constant}\)
\(\mu\)-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

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

Thm. The map $T$ is ergodic w.r.t. the Lebesgue measure $\mu$ on $[0,1]$.

Proof. $\mu$ is $T$-invariant: see Part 1.

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

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

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

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

For $l = 0$, get $f \equiv 1 \Rightarrow \langle f \rangle_n \equiv 1$. 
For \( \ell \neq 0 \), get

\[
\langle e_\ell \rangle_n = \frac{1}{n} \sum_{j=0}^{n-1} e_\ell (2j \delta x) = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i e.2j \delta x} = \frac{1}{n} \sum_{j=0}^{n-1} e_2j \delta e.
\]

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

as \( e_2j \delta e \) 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. \( \blacksquare \)

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

To see this, we note that \( T \) has a fixed point at \( 0 : T(0) = 0 \).

Take \( \mu_0 = \delta_0 \leftarrow \) delta measure at \( 0 \).
Then $\mu_0$ is $T$-invariant if $\forall \delta \in \mathcal{B}$ 

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

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

In fact, we can easily see that

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

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

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

for Lebesgue almost every $x$.

More generally, if we have a periodic orbit $\delta = \{x_0, x_1, \ldots, x_{m-1}\}$

of $T$, i.e. $x_0 \xrightarrow{\delta} x_1 \xrightarrow{\delta} \cdots \xrightarrow{\delta} x_{m-1} \xrightarrow{\delta} x_0$

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

is $T$-invariant and ergodic (exercise, no credit).
What are the periodic orbits of $T$?

Need to find $x, m$ such that

\[ T^m(x) = x \], i.e.

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

\[ x \in \mathbb{Z} \cdot \frac{2^m - 1}{2^m - 1} \].

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 $T(x) = \lfloor 2x \rfloor \mod 12$, the set of periodic points is dense.
§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 $\mu$ an ergodic $T$-invariant probability measure on $X$. Then for each $f \in L^1(X, \mu)$ we have

$$\langle f \rangle_n(x) \overset{\text{def}}{=} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \underset{n \to \infty}{\longrightarrow} \int_X f \, d\mu \quad \text{for } \mu\text{-almost every } x \in X.$$

Remarks

1. A common interpretation is that a $\mu$-typical trajectory equidistributes according to $\mu$.

2. The general version of the Thm does not need $\mu$ to be ergodic, and $\langle f \rangle_n(x)$ is replaced by a certain projector applied to $f$. 
The proof of Birkhoff’s Ergodic Theorem 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 $X, T, \mu$ be as in the Thm and $f \in L^1(X, \mu)$ satisfy $\int_X f \, d\mu < 0$. Then for $\mu$-almost every $x$, we have

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

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

$$\int_X (f - I - \frac{1}{N}) \, d\mu = -\frac{1}{N} < 0.$$
Applying the lemma to \( f-\frac{1}{n} \), we see that there is a set
\[ A_N \in \mathcal{X}, \quad \mu(A_N) = 0, \quad \text{such that} \]
\[ \forall x \in \mathcal{X} \setminus A_N, \quad \limsup_{n \to \infty} \langle f \rangle_n(x) \leq I + \frac{1}{n}. \]
Take \( A_\infty := \bigcup_{N=1}^{\infty} A_N \), then \( \mu(A_\infty) = 0 \) and \( \forall x \in \mathcal{X} \setminus A_\infty \quad \forall N \quad \limsup_{n \to \infty} \langle f \rangle_n(x) \leq I + \frac{1}{n}. \)

That is, for \( \mu \)-almost every \( x \) we have
\[ \limsup_{n \to \infty} \langle f \rangle_n(x) \leq I. \]

A similar argument (replacing \( f \) with \( -f \)) gives
\[ \liminf_{n \to \infty} \langle f \rangle_n(x) \geq I \]
for \( \mu \)-almost every \( x \).

So for \( \mu \)-almost every \( x \) we have
\[ \lim_{n \to \infty} \langle f \rangle_n(x) = I. \]
Proof of Lemma

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

Define the sums
\[
S_n(x) = \sum_{j=0}^{n-1} f(T^j(x)),
\]
so that \( \langle f \rangle_n(x) = \frac{S_n(x)}{n} \).

We have the following identity for all \( n \geq 0 \):
\[
S_{n+1}(x) = f(x) + S_n(T(x)). \quad (\ast)
\]

Define the Borel set \( ACX \) by
\[
A = \{ x \in X \mid \sup_{n \geq 1} S_n(x) = \infty \}
\]
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

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

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

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

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

Now, since $A$ is $T$-invariant and $T$ is ergodic w.r.t. $\mu$, we have $\mu(A) = 0$ or $\mu(A) = 1$.

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

$\lim\sup_{n \to \infty} \langle f \rangle_n(x) = \lim\sup_{n \to \infty} \frac{S_n(x)}{n} \leq 0$, which gives the statement of the Lemma.
So assume now that \( \mu(A) = 1 \), i.e. \( S_n(x) \) is not bounded above for \( \mu \)-almost every \( x \). We will reach a contradiction.

Define the function for \( m \geq 1 \),

\[
F_m(x) = \max_{1 \leq n \leq m} S_n(x) = \max (f(x), f(x) + f(T(x)), \ldots, f(x) + \cdots + f(T^{m-1}(x))).
\]

Note that \( F_m \leq F_{m+1} \) and \( \forall m, \ F_m \in L^1(X, \mu) \):

Indeed, \( f \in L^1(X, \mu) \) & \( \mu \) 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

\[
F_{m+1}(x) = \max (f(x), \max_{1 \leq n \leq m} S_{n+1}(x)) = \max (f(x), f(x) + \max_{1 \leq n \leq m} S_n(T(x))) = f(x) + \max (0, F_m(T(x))).
\]
Therefore
\[ G_m(x) := F_{m+1}(x) - F_m(T(x)) \]
\[ = f(x) - \min(0, F_m(T(x))) \]
We see that (since \( F_m \geq F_1 \))
\[ f \leq G_m \leq G_1 \]
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 \( T(x) \in A \Rightarrow \)
\[ \Rightarrow F_m(T(x)) \uparrow \infty \text{ as } m \to \infty \]
\[ \Rightarrow F_m(T(x)) > 0 \text{ for large enough } m \]
\[ \Rightarrow G_m(x) = f(x) \text{ for large enough } m. \]
In particular, \( G_m(x) \xrightarrow{m \to \infty} f(x) \)
for \( \mu \)-almost every \( x \) (i.e. for \( x \in A \))
So, by the Dominated Convergence Theorem
\[ \int X G_m \, d\mu \xrightarrow{m \to \infty} \int X f \, d\mu < 0. \]
But $S G_{n} \, d\mu = \lim_{n \to \infty} \int_{\mathbb{R}} S G_{n} \, d\mu$

$= \int_{\mathbb{R}} S F_{n+1} \, d\mu - \int_{\mathbb{R}} S F_{n} \, d\mu$

$= \int_{\mathbb{R}} S F_{n+1} \, d\mu - \int_{\mathbb{R}} S F_{n} \, d\mu$

$= \int_{\mathbb{R}} (F_{n+1} - F_{n}) \, d\mu \geq 0$ since $F_{n} \leq F_{n+1}$.

This gives a contradiction. \( \square \)

Remark. We have $\int S S_{n} \, d\mu = \frac{n}{\mu} \int S f \, d\mu \to -\infty$.

The hard part of the proof was to exclude the possibility that $\sup S_{n} = \infty \mu$-almost everywhere even though $\int S S_{n} \, d\mu \to -\infty$ (which could happen for a general sequence $S_{n}$).