

Ergodic Theory Notes

1 Basic Definitions

For this section we fix a probability space (X, \mathcal{B}, μ) and we have a transformation $T : X \rightarrow X$ which is measurable in our probability space.

We say T is a **measure preserving transformation (m.p.t.)** or μ is a **T -invariant measure** if

$$\mu(T^{-1}B) = \mu(B) \quad \forall B \in \mathcal{B}$$

The **push forward of μ by T** is defined to be

$$T_*\mu(B) = \mu(T^{-1}B) \quad \forall B \in \mathcal{B}$$

We say a measure μ is **regular** if $\forall B \in \mathcal{B}$ we have $\forall \epsilon > 0 \exists U \subseteq X$ open such that

$$B \subseteq U \quad \text{and} \quad \mu(U) < \mu(B) + \epsilon$$

An m.p.t T is said to be **ergodic** if

$$\forall B \in \mathcal{B}, T^{-1}B = B \implies \mu(B) = 0 \text{ or } 1$$

2 Facts on Fourier Series

Suppose $f \in L_1(\mathbb{T}^k)$ then we can define the **Fourier coefficients** by

$$\hat{f}(n) = \int_{\mathbb{T}^k} f(x) e^{-2\pi i n \cdot x} dx \quad \forall n \in \mathbb{Z}^k$$

Theorem 2.1 (Fejér's Theorem). *The average of the partial Fourier sums converges uniformly to f , i.e.*

$$\frac{1}{N} \sum_{k=0}^{N-1} S_k f \rightarrow f \quad \text{uniformly}$$

Theorem 2.2 (Riemann-Lebesgue Lemma). *For all $f \in L_1(\mathbb{T}^k)$,*

$$\lim_{|n| \rightarrow \infty} \hat{f}(n) = 0$$

Theorem 2.3 (Riesz-Fisher Theorem). *Define $S_N f(x) = \sum_{|n| \leq N} \hat{f}(n) e^{2\pi i(n \cdot x)}$ then $S_N f \rightarrow f$ in L^2 for all $f \in L^2(\mathbb{T}^k)$.*

Corollary 2.4. *If $f \in L^2(\mathbb{T}^k)$ and $\hat{f}(n) = 0 \forall n \in \mathbb{Z}^k \setminus \{0\}$, then f is constant.*

Theorem 2.5. *Given $f \in L^2$ which is T -invariant*

$$\hat{f}(n) = \lim_{N \rightarrow \infty} \int (S_N f)(Tx) e^{-2\pi i n \cdot x}$$

3 Criteria for measure preserving

Theorem 3.1. *Given $T : X \rightarrow X$ on a probability space (X, μ) , the following are equivalent:*

1. T is m.p.t
2. $\int f \circ T d\mu = \int f d\mu \quad \forall f \in L_1(X)$.

Recall the space $L_1(X) = \{f : x \rightarrow \mathbb{R} : \text{measurable} \quad \|f\|_1 := \int |f| d\mu < \infty\}$

Theorem 3.2. *Given $T : X \rightarrow X$ on a probability space (X, μ) , the following are equivalent:*

1. T is m.p.t
2. $\int f \circ T d\mu = \int f d\mu \quad \forall f \in C(X)$.

So we see that in fact it suffices to check that T does not affect the integral of any continuous function f . However, we can extend this further using the density of trigonometric polynomials in the space of continuous functions. First, we need to define a trigonometric polynomial in arbitrary dimension on the k -torus $X = \mathbb{T}^k$ with $\mu = \text{leb}$ and $\mathcal{B} = \text{Borel}$.

$P : \mathbb{T}^k \rightarrow \mathbb{T}^k$ is a **trigonometric polynomial** if for some $N \geq 1$ and $c_n \in \mathbb{C}$ we can write

$$P(x) = \sum_{|n| \leq N} c_n e^{2\pi i n \cdot x}$$

where $n = (n_1, \dots, n_k) \in \mathbb{Z}^k, x = (x_1, \dots, x_k), |n| = |n_1| + \dots + |n_k|$.

Note:

$$\int_{\mathbb{T}^k} e^{2\pi i n \cdot x} dx = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$$

and hence

$$\int_{\mathbb{T}^k} P = c_0$$

Theorem 3.3. *Given $T : \mathbb{T}^k \rightarrow \mathbb{T}^k$ continuous and denoting by μ the Lebesgue measure.*

1. T is m.p.t
2. $\int P \circ T d\mu = \int P d\mu \quad \forall \text{ trigonometric polynomials } P$.

4 Criteria for Ergodicity

First another few definitions.

Given $A, B \subseteq X$, their **symmetric difference** is

$$A \Delta B := (A \setminus B) \cup (B \setminus A)$$

A function f is **T -invariant** if $f \circ T = f$ a.e.

A function f is **constant** if $\exists c \in \mathbb{R}$ such that $f(x) = c$ almost everywhere.

Theorem 4.1. *Given a measure preserving transformation $T : X \rightarrow X$ and some $1 \leq p \leq \infty$. TFAE:*

1. T is ergodic.
2. For all f measurable f invariant $\iff f$ constant.
3. For all $f \in L^p(X)$, f invariant $\iff f$ constant.

Note: To check that T is ergodic it suffices to show that all invariant L^2 functions have zero Fourier coefficients away from zero.

To this end we present the following formula for computing the Fourier coefficients of invariant L^2 functions.

Theorem 4.2. *Given $f \in L^2$ which is invariant*

$$\hat{f}(n) = \lim_{N \rightarrow \infty} \int (S_N f)(Tx) e^{-2\pi i n \cdot x} dx$$

5 Theorems using Measure Preserving

Theorem 5.1 (Poincaré Recurrence Theorem). *Given a probability space (X, \mathcal{B}, μ) and $T : X \rightarrow X$ measure preserving. Then*

$$\mu\{x \in B : T^n x \in B \text{ infinitely often}\} = \mu(B) \quad \forall B \in \mathcal{B}$$

6 Theorems using Ergodicity

Theorem 6.1 (Pointwise Ergodic Theorem - Birkhoff 1931). *Given a measure space (X, \mathcal{B}, μ) and a measure preserving transformation $T : X \rightarrow X$ and $f \in L^1(X)$. Then $\exists f^* \in L^1(X)$ invariant such that*

$$\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \rightarrow f^* \text{ a.e.} \quad \text{and} \quad \int f^* = \int f$$

Note this does not actually need ergodicity. However, if we additionally assume ergodicity we can prove the following stronger result.

Corollary 6.2. *Given a probability space (X, \mathcal{B}, μ) , T measure preserving and ergodic, $f \in L^1(x)$, then*

$$\underbrace{\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j}_{\text{Time average}} \rightarrow \underbrace{\int f d\mu}_{\text{Space average}} \text{ a.e.}$$

Theorem 6.3 (Mean Ergodic Theorems). $1 \leq p < \infty$, T measure preserving theorem, $f \in L^p(X)$. Define $f^* := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$ almost everywhere. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j = f^*$$

in L^p .

7 Examples

7.1 Linear toral automorphism

A **linear toral automorphism** is a map $Tx = Ax \pmod{1}$ with A a $k \times k$ matrix with integer entries and $\det(A) \neq 0$.

Such an automorphisms is **hyperbolic** if all eigenvalue for A have $|\lambda| \neq 1$.

Theorem 7.1. T ergodic \iff no eigenvalue of A is a root of unity.

7.2 Normality of real numbers

$x \in \mathbb{R}$ is **normal (base b)** if

- x has a unique expansion in that base.
- $\forall k \in \{0, 1, \dots, b-1\}$

$$\frac{1}{n} \# \{1 \leq i \leq n | x_i = k\} \rightarrow \frac{1}{10} \quad \text{as } n \rightarrow \infty$$

$x \in \mathbb{R}$ is **absolutely normal** if x is normal base b for all $b \geq 2$.

Theorem 7.2. Almost every $x \in \mathbb{R}$ is absolutely normal.

8 Existence of invariant/ergodic measures

Let $M(X)$ be the set of all probability measure on X .

We can view measures as linear functionals on the space of continuous functions as such:

$$\forall f \in C(X) \quad \mu(f) := \int_X f d\mu$$

$C(X)^* := \{\text{bounded linear functionals } w : C(X) \rightarrow \mathbb{R}\}$

A linear functional is called **normalised** if $\int 1 d\mu = 1$

A linear functional is called **positive** if $f \geq 0 \implies \int f d\mu \geq 0$

Theorem 8.1. Every $\mu \in M(X)$ defines a normalised, positive, bounded, linear functional in $C(X)^*$ defined by $\mu(f) = \int_X f d\mu$.

Theorem 8.2 (Reisz Representation Theorem). *Let $w \in C(X)^*$ be a bounded linear functional. Suppose that w is positive and normalised. Then $\exists! \mu \in M(X)$ such that $w(f) = \mu(f)$ for all $f \in C(X)$.*

9 Entropy

The motivation for a definition of entropy is as a vehicle to distinguish between dynamical systems. First we need to know how tell when two systems are identical.

Two probability spaces with measure preserving transformations, $(X, \mathcal{B}, \mu, T), (Y, \mathcal{C}, \nu, S)$ are **measure-theoretically isomorphic** if there exists a bijection $\pi : B \rightarrow C$ where $B \in \mathcal{B}$ and $C \in \mathcal{C}$ such that

- $\mu(B) = \nu(C) = 1$
- $T(B) \subseteq B, S(C) \subseteq C$
- $\pi : B \rightarrow C$ and $\pi^{-1} : C \rightarrow B$ are measure preserving transformations
- $\pi \circ T = S \circ \pi$

Assume (X, \mathcal{B}, μ) is a probability space and $\alpha = \{A_i\}$ a countable collection of subsets $A_i \subseteq B$.

- We say α is a **partition** of X if $\cup A_i = X$ and $A_i \cap A_j = \emptyset$ up to measure 0.
- The **join** of two partitions α, β is the partition $\alpha \vee \beta$ of all possible intersections $A_i \cap B_j$.
- A countable partition β is a **refinement** of α if every element of α is a union of element of β and write $\alpha \leq \beta$.
- α, β are **independent** if $\mu(A \cap B) = \mu(A)\mu(B)$ for all $A \in \alpha, B \in \beta$.
- The **information of a partition** α is

$$I(\alpha) := - \sum_{A \in \alpha} \mathbb{1}_A \log(\mu(A))$$

where $I(\alpha) : X \rightarrow [0, \infty]$.

- The **entropy of a partition** α is

$$H(\alpha) := \int_X I(\alpha) d\mu = - \sum_{A \in \alpha} \mu(A) \log(\mu(A))$$

using the convention $0 \cdot \log(0) = 0$.

- The **expectation given a partition** is

$$\mathbb{E}(\cdot \mid \alpha) := \mathbb{E}(\cdot \mid \sigma(\alpha))$$

- The **conditional probability** of $B \in \mathcal{B}$ given α is

$$\mathbb{P}(B \mid \alpha) := \mathbb{E}(\mathbb{1}_B \mid \alpha)$$

Suppose that \mathcal{C} is a sub σ -algebra of \mathcal{B} .

- The **conditional information of α given \mathcal{C}** is

$$I(\alpha \mid \mathcal{C}) := - \sum_{A \in \alpha} \mathbb{1}_A \log(\mu(A \mid \mathcal{C}))$$

where $\mu(A \mid \mathcal{C}) := \mathbb{E}(\mathbb{1}_A \mid \mathcal{C})$

- The **conditional entropy of α given \mathcal{C}** is

$$H(\alpha \mid \mathcal{C}) := \int_X I(\alpha \mid \mathcal{C}) d\mu$$

We have the following desirable properties:

- If α and β are independent then $I(\alpha \vee \beta) = I(\alpha) + I(\beta)$.
- If $\alpha = \{X\}$ then $I(\alpha) = 0$ so $H(\alpha) = 0$.
- If T is a measure preserving transformation then $H(T^{-1}\alpha) = H(\alpha)$.
- Given $A \in \alpha$, $\mathbb{E}(f \mid \alpha) \big|_A = \frac{\int_A f d\mu}{\mu(A)}$ and hence

$$\mathbb{E}(f \mid \alpha) = \sum_{A \in \alpha} \mathbb{1}_A \frac{\int_A f d\mu}{\mu(A)}$$

- Conditional probability and expectation are constant on partition elements.
- For $A \in \alpha$,

$$\mathbb{P}(B \mid \alpha) \big|_A = \mathbb{E}(\mathbb{1}_B \mid \alpha) \big|_A = \frac{\int_A \mathbb{1}_B d\mu}{\mu(A)} = \frac{\mu(A \cap B)}{\mu(A)}$$

- If $\mathcal{C} = \{X, \emptyset\}$ then $I(\alpha \mid \mathcal{C}) = I(\alpha)$ and $H(\alpha \mid \mathcal{C}) = H(\alpha)$.
- If $g \geq 0$ is $\sigma(\alpha)$ -measurable then $\mathbb{E}(fg \mid \sigma(\alpha)) = g \cdot \mathbb{E}(f \mid \sigma(\alpha))$.
- If T is a measure preserving transformation then $I(T^{-1}\alpha \mid T^{-1}\mathcal{C}) = I(\alpha \mid \mathcal{C}) \circ T$.
- Integrating this gives $H(T^{-1}\alpha \mid T^{-1}\mathcal{C}) = H(\alpha \mid \mathcal{C})$.
- $\alpha \leq \beta \implies I(\alpha \mid \beta) = 0$.

Proposition 9.1.

$$H(\alpha \mid \mathcal{C}) = - \int_X \sum_{A \in \alpha} \mu(A \mid \mathcal{C}) \log(\mu(A \mid \mathcal{C})) d\mu$$

Lemma 9.2 (Basic Identity). *Given α, β, γ partitions of X then*

$$I(\alpha \vee \beta \mid \gamma) = I(\alpha \mid \gamma) + I(\beta \mid \alpha \vee \gamma)$$

$$H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) + H(\beta \mid \alpha \vee \gamma)$$

Corollary 9.3.

$$\beta \leq \gamma \implies I(\alpha \vee \beta \mid \gamma) = I(\alpha \mid \gamma)$$

Corollary 9.4 (Monotonicity of information of entropy).

$$\alpha \leq \beta \implies I(\alpha \mid \gamma) \leq I(\beta \mid \gamma)$$

Corollary 9.5 (Anti-monotonicity of entropy).

$$\beta \leq \gamma \implies H(\alpha \mid \beta) \geq H(\alpha \mid \gamma)$$

Corollary 9.6. *We have the two following properties as well:*

- $H(\alpha \mid \gamma) \leq H(\alpha)$ (because always $\gamma \leq \{X, \emptyset\}$)
- $H(\alpha \vee \beta \mid \gamma) \leq H(\alpha \mid \gamma) + H(\beta \mid \gamma)$

So far this does not encapsulate any dynamics of the system and so we must use these concepts to arrive at a definition of entropy which depends on the transformation. For convenience define the following set:

$$\mathcal{P} := \{\alpha \text{ countable partitions} \mid H(\alpha) < \infty\}$$

Now choose $\alpha \in \mathcal{P}$. Then we define the following:

$$H_n(\alpha) := H(\alpha^n) \quad \text{where} \quad \alpha^n := \bigvee_{j=0}^{n-1} T^{-j} \alpha$$

This has the convenient property that $H_{n+m}(\alpha) \leq H_n(\alpha) + H_m(\alpha)$, i.e. these H_n form a sub-additive sequence \mathbb{R} -valued sequence and hence the limit $h(T, \alpha) := \lim_{n \rightarrow \infty} \frac{1}{n} H_n(\alpha)$ exists. We call this the **entropy of T relative to α** . We can then define the **entropy of T** by taking the supremum:

$$h(T) := \sup_{\alpha \in \mathcal{P}} h(T, \alpha)$$

Having done all this work, this had better be a measure-theoretic isomorphism invariant.