Ergodic Theory Notes

1 Basic Definitions

For this section we fix a probability space $(X, \mathcal{B}, \mu)$and we have a transformation $T : X \to X$ 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 defiend 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.*

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

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

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

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

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

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

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

3 Criteria for measure preserving

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

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

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

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

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

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 $\mu = leb$ and $\mathcal{B} = Borel$.

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

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

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

Note:

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

and hence

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

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

1. *T is m.p.t*
2. *$\int P \circ T d\mu = \int P d\mu \qquad \forall$ trigonometric polynomials P.*

4 Criteria for Ergodicity

First another few definitions.

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

$$A \triangle B := (A \backslash B) \cup (B \backslash A)$$

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 $T : X \to X$ and some $1 \leqslant p \leqslant \infty$. 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*

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

5 Theorems using Measure Preserving

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

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

6 Theorems using Ergodicity

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

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

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 $(X, \mathcal{B}, \mu)$, T measure preserving and ergodic, $f \in L^1(x)$, then*

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

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

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

in L^p*.*

Proof. **Special Case:** $1 \leqslant p < \infty$ but $f \in L^\infty(X)$.
Then by the ergodic theorem and the DCT with dominator $2^p \,||f||^p$ we have

$$\left|\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j - f*\right|^p \to 0$$

General Case: Take $f \in L^p$
Given $\epsilon > 0$ then there is a $g \in L^\infty$ such that $||f-g||_p < \frac{\epsilon}{3}$. Then we get f^* associated to f and g^* associated to g. Then $(f-g)^*$ is associated to $f-g$ and $(f-g)^* = f^* - g^*$. By a previous proposition we can see

$$||f^* - g^*||_p = ||(f-g)^*||_p \leqslant ||f-g||_p < \frac{\epsilon}{3}$$

Also since $g \in L^\infty$ there must be an N such that

$$n \geqslant N \implies \left|\left|\frac{1}{n}\sum_{j=0}^{n-1} g \circ T^j - g^*\right|\right|_p < \frac{\epsilon}{3}$$

Then

$$\begin{aligned}
\left|\left|\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j - f^*\right|\right|_p &\leqslant \left|\left|\frac{1}{n}\sum_{j=0}^{n-1} (f-g) \circ T^j\right|\right|_p + \underbrace{\left|\left|\frac{1}{n}\sum_{j=0}^{n-1} g \circ T^j - g^*\right|\right|_p}_{<\epsilon/3 \text{ for } n\geqslant N} + \underbrace{||g^* - f^*||}_{<\epsilon/3} \\
&\leqslant \frac{1}{n}\sum_{j=0}^{n-1} \left|\left|(f-g)\circ \cancelto{\text{m.p.t}}{T^j}\right|\right|_p + \frac{2\epsilon}{3} \\
&= ||f-g|| + \frac{2\epsilon}{3} < \epsilon
\end{aligned}$$

□

7 Examples

7.1 Linear toral automorphism

A linear toral automorphism is a map $Tx = Ax(mod 1)$ 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.
- $\forall k \in \{0, 1, \ldots, b-1\}$

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

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

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

8 Von Neumann's Ergodic Theorem & The Adjoint

Given $T : X \to X$ a measure preserving transformation on a probability space (X, μ), the Koopman operator is given by

$$Uf := f \circ T$$

for any $f : X \to \mathbb{R}$ measurable.
Suppose H is a complex Hilbert space with inner product $\langle \cdot, \cdot \rangle$ then a linear operator $U : H \to H$ is an isometry if

$$||Uf|| = ||f|| \quad \forall f \in H$$

where $||f|| = \sqrt{\langle f, f \rangle}$. Equivalently $\langle Uf, Ug \rangle = \langle f, g \rangle$ for all $f, g \in H$.
Given a linear operator $U : H \to H$, the adjoint $U* : H \to H$ is the unique bounded linear operator satisfying

$$\langle U^* f, g \rangle = \langle f, Ug \rangle \quad \forall f, g \in H$$

Let $V \subseteq H$ be a subspace then the orthogonal complement is

$$V^\perp := \{f \in H \mid \langle f, v \rangle = 0 \quad \forall v \in V\}$$

Lemma 8.1 (Properties of the adjoint)**.** *If U is an isometry then*

- $||U^* f|| \leqslant ||f|| \quad \forall f \in H$
- $U^*U = id$ *because*

$$\langle U^*Uf, g \rangle = \langle Uf, Ug \rangle = \langle f, g \rangle \quad \forall f, g \in H$$

Example: Computing the adjoint. $X = [0, 1]$, $\mu = Leb$, $Tx = 2x \mod 1$ and $Uf = f \circ T$ where $U : L^2(X) \circlearrowleft$ and our inner product is

$$\langle f, g \rangle := \int_0^1 f\overline{g} \, d\mu$$

$$
\begin{aligned}
\langle U^* f, g\rangle = \langle f, Ug\rangle &= \int_0^1 f\overline{Ug}\, dx \\
&= \int_0^1 f(x)\overline{g(Tx)}\, dx \\
&= \int_0^{\frac{1}{2}} f(x)\overline{g(2x)}\, dx + \int_{\frac{1}{2}}^1 f(x)\overline{g(2x-1)}\, dx \\
&= \frac{1}{2}\int_0^1 f\left(\frac{x}{2}\right)\overline{g(x)}\, dx + \frac{1}{2}\int_0^1 f\left(\frac{x+1}{2}\right)\overline{g(x)}\, dx
\end{aligned}
$$

Hence we can conclude

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

Proposition 8.2. *Suppose U is an isometry then*

$$
Uf = f \iff U^* f = f
$$

Given a bounded linear operator $A : H \to H$ we can define the kernel to be

$$
\ker(A) := \{f \in H \mid Af = 0\}
$$

then this a closed subspace in H. Moreover, if U is an isometry then the above proposition tells us that $\ker(U - I) = \ker(U^* - I)$.
Fact: For every closed subspace $V \subseteq H$ we can write $H = V \oplus V^\perp$ and hence

$$
\forall f \in H \quad \exists! v \in V, w \in V^\perp \; s.t. \; f = v + w
$$

then we can define orthogonal projection $\pi : H \to V$ by

$$
\pi(f) = \pi(v + w) = v
$$

Theorem 8.3 (Von Neumann)**.** *If H is a Hilbert space and $U : H \circlearrowleft$ is an isometry. Let π denote orthogonal projection into $V = \ker(U - I)$ then*

$$
\frac{1}{n}\sum_{j=0}^{n-1} U^j f \to \pi(f) \quad in\ H \quad as\ n \to \infty
$$

that is

$$
\lim_{n\to\infty}\left\|\frac{1}{n}\sum U^j f - \pi(f)\right\| = 0
$$

Proof. The proof of this is about a page long and definitely warrants a read. □

Corollary 8.4. *Given a measure preserving transformation and $Uf = f \circ T$ and $H = L^2(x)$. Then*

$$
\left\|\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j - \pi f\right\|_2 \to 0 \quad as\ n \to \infty
$$

If T is ergodic then $\pi f = \int f\, d\mu$.

9 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:

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

$C(X)^* :=$ {bounded linear functionals $\quad w : C(X) \to \mathbb{R}$}
A linear functional is called normalised if $\int 1 d\mu = 1$
A linear functional is called positive if $f \geqslant 0 \implies \int f d\mu \geqslant 0$

Theorem 9.1. *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 9.2 (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 $f \in C(X)$.*

We would like to give the space $M(X)$ a topology. Our first idea is the strong/norm topology. We view $M(X) \subseteq C(X)^*$ and inherit the operator norm from $C(X)^*$. That is, given $\mu, \nu \in M(X)$

$$d_s(\mu, \nu) := ||\mu - \nu|| = \sup_{f \in C(X),\, ||f||_\infty = 1} |\mu(f) - \lambda(f)| = \sup_{f \in C(X),\, ||f||_\infty = 1} \left| \int f \, d\mu - \int f \, d\lambda \right|$$

Note:

$$||\mu|| = 1 \quad \forall \mu \in M(X) \subseteq C(X)^*$$

since $|\mu(f)| \leqslant ||f||_\infty$ for all $f \in C(X)$ and $\mu(1) = 1$. Therefore $M(X)$ is a bounded subset of $C(X)^*$.

Lemma 9.3. *$M(X)$ is closed.*

Proof. Suppose we have some sequence $(\mu_n) \subseteq M(X)$ such that $\mu_n \to w \in C(X)^*$. We aim to show that $w = \mu \in M(X)$ We check that the Reisz Representation Theorem is satisfied

- Certainly $w \in C(X)^*$.
- Normalised : $w(1) = \lim_{n \to \infty} \mu_n(1) = \lim_{n \to \infty} 1 = 1$.
- Positive: $f \geqslant 0 \implies \mu_n(f) \geqslant 0$ for all n and hence $w(f) \geqslant 0$.

□

Lemma 9.4. *Unfortunately, $M(X)$ is not compact in the strong topology.*

Proof. Recall that in a metric space compactness is equivalent to sequential compactness. So it suffices to find a sequence with no convergent subsequence. Let $x_1, c_2, \cdots \in C$ such that $x_i \neq x_j$ and for all n let $\mu_n = \delta_{x_n}$.
Now take $n \neq m$ we want to show that $||\mu_n - \mu_m|| \geqslant 1$. For this we define the function

$$f(x) = \frac{d(x, x_n}{d(x, x_n) + d(x, x_m)}$$

Note since $x_n \neq x_m$ this is well-defined and $f(x_n) = 0$ and $f(x_m) = 1$. Moreover, f is continuous and $||f||_\infty = 1$ Therefore $||\delta_{x_n} - \delta_{x_m}|| \geqslant 1$. □

9.1 Weak $*$ topology on $M(X)$

Let $\mu_n \in M(X)$ and $\mu \in M(X)$. We say that $\mu_n \to \mu$ weak $*$ if

$$\mu_n(f) \to \mu(f) \quad \forall f \in C(X)$$

We can then given $M(X)$ a metric by fixing some countable dense subset $\{f_1, f_2, \dots\} \subseteq C(X)$ and defining

$$d(\lambda, \mu) := \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{1}{||f_i||_\infty} \underbrace{|\lambda(f_i) - \mu(f_i)|}_{\leqslant ||f_i||_\infty \int 1 d(\lambda-\mu) \leqslant ||f_i||} \quad \in [0,1]$$

Proposition 9.5. *d is a metric.*

Proof. The difficult thing to prove here is that $\lambda \neq \mu \implies d(\lambda, \mu) > 0$. Suppose that we have measures $\lambda \neq \mu$. By the Reisz Representation Theorem, they must constitute different element of $C(X)^*$. So there is an $f \in C(X)$ such that $\lambda(f) \neq \mu(f)$. Since the f_i are dense there is some i such that

$$||f_i - f||_\infty < \frac{|\lambda(f) - \mu(f)|}{3}$$

Now

$$\begin{aligned} ||\lambda(f) - \mu(f)|| &\leqslant |\lambda(f) - \lambda(f_i)| + |\lambda(f_i) - \mu(f_i)| + |\mu(f_i) + \mu(f)| \\ &\leqslant 2\, ||f_i - f||_\infty + |\lambda(f_i) - \mu(f_i)| \\ &< \frac{2\,|\lambda(f) - \mu(f)|}{3} + |\lambda(f_i) - \mu(f_i)| \end{aligned}$$

Therefore $|\lambda(f_i) - \mu(f_i)| > \frac{1}{3}\,|\lambda(f) - \mu(f)| > 0$ So one term of the sum is non-zero and therefore $d(\lambda, \mu) > 0$. □

Proposition 9.6. *$\mu_n \to \mu$ weak $*$ $\iff$ $d(\mu_n, \mu) \to 0$ as $n \to \infty$.*

Proof. Suppose that $\mu_n \to \mu$ weak $*$ and choose $\epsilon > 0$. There exists M such that

$$\sum_{i=M}^{\infty} \frac{1}{2^i} < \frac{\epsilon}{2}$$

Then

$$d(\mu_n, \mu) \leqslant \sum_{i=1}^{M} \left[\frac{1}{2^i} \frac{1}{||f_i||_\infty} |\mu_n(f_i) - \mu(f_i)| \right] + \frac{\epsilon}{2}$$

Also there is an N such that for $n \geqslant N$ we can be sure each summand is less than $\frac{\epsilon}{2M}$ since $\mu_n \to \mu$ weak $*$ and we only have finitely many i to deal with. Therefore for any $n \geqslant N$ we have $d(\mu_n, \mu) \leqslant \epsilon$. Conversely, suppose that $d(\mu_n, \mu) \to 0$ then choose $f \in C(X)$ and $\epsilon > 0$. Then there is an i such that $||f_i - f||_\infty < \frac{\epsilon}{3}$. Also

$$|\mu_n(f_i) - \mu(f_i)| \leqslant 2^i\, ||f_i||\, d(\mu_n, u) \to 0 \quad \text{as } n \to \infty$$

so there is an N such that for all $n \geqslant N$ we have $|\mu_n(f_i) - \mu(f_i)| < \frac{\epsilon}{3}$. Then we can do the normal to trick to show that $|\mu_n - \mu(f)| < \epsilon$. □

Theorem 9.7. *$M(X)$ is weak $*$ compact.*

9.2 Existence of Invariant Measures

Given X a compact metric space, let $\mathcal{B}$ denote the Borel σ-algebra and $M(X)$ be defined as before. Let $T : X \to X$ be a continuous map. Define

$$M(X,T) := \{\mu \in M(X) \mid T_*\mu = \mu\}$$

One can show that for any $f \in C(X)$ we have $T_*\mu(f) = \mu(f \circ T)$. This is proven first for simple functions and then slowly built up.

Theorem 9.8 (Krylov-Bogulyvhov)**.** $M(X,T) \neq \varnothing$.

Proposition 9.9. *$M(X,T)$ is convex.*

Proposition 9.10. *$M(X,T)$ is weak $*$ compact.*

Proof. Note it suffices to prove that $M(X,T)$ is closed since $M(X,T) \subseteq M(X)$ and $M(X)$ is compact. In metric spaces we can just make sure all sequences have limits in $M(X,T)$. Let $\mu_n \in M(X,T)$ such that $\mu_n \to \mu \in M(X)$ weak $*$.
Take $f \in C(X)$ then notice

$$T_*\mu(f) = \mu(f \circ T) \leftarrow \mu_n(f \circ T) = \mu_n(f) \to \mu(f)$$

By uniqueness of limits and since f was arbitrary we have that $\mu = T_*\mu$ and hence $\mu \in M(X,T)$. $\square$

9.3 Existence of Ergodic Measures

Let Y be a convex set then $y \in Y$ is called extremal if

$$\exists y_0, y_1 \in Y \quad \text{and} \quad t \in (0,1) \; s.t. \quad y = (1-t)y_0 + ty_1 \implies y = y_0 = y_1$$

Proposition 9.11. *$\mu \in M(X,T)$ is extremal $\implies$ μ is ergodic.*

Proof. Suppose that μ is not ergodic so there exists $B \in \mathcal{B}$ such that $T^{-1}B = B$ and $\mu(B) \in [0,1)$. Then we let

$$\mu_0(A) := \frac{\mu(A \cap B)}{\mu(B)} \qquad \mu_1(A) := \frac{\mu(A \cap B^{\mathsf{c}})}{\mu(B^{\mathsf{c}})}$$

One can show that these define T-invariant measures and satisfy

$$\mu(B)\mu_0 + \mu(B^{\mathsf{c}})\mu_1 = \mu$$

Note that $\mu(B)$ and $\mu(B^{\mathsf{c}})$ are both in $(0,1)$ and $\mu_i \neq \mu$ for either i. Hence μ cannot be extremal. $\square$

For the opposite direction we need the Radon-Nikodym Theorem.

10 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 $\pi : B \to C$ where $B \in \mathcal{B}$ and $C \in \mathcal{C}$ such that

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

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\pi"'] & X \arrow[d, "\pi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

Assume $(X, \mathcal{B}, \mu)$ 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 $A_i \cap A_j = \varnothing$ 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 $\alpha \leqslant \beta$.
- 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 $I(\alpha) : X \to [0, \infty]$.
- 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

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

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

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

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

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 $A \in \alpha$, $\mathbb{E}\left(f \mid \alpha\right)\big|_A = \frac{\int_A f\, d\mu}{\mu(A)}$ and hence

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

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

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

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

Proposition 10.1.

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

Lemma 10.2 (Basic Identity)**.** *Given α, β, γ partitions of X then*

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

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

Corollary 10.3.

$$\beta \leqslant \gamma \implies I\left(\alpha \vee \beta \mid \gamma\right) = I\left(\alpha \mid \gamma\right)$$

Corollary 10.4 (Monotonicity of information of entropy)**.**

$$\alpha \leqslant \beta \implies I\left(\alpha \mid \gamma\right) \leqslant I\left(\beta \mid \gamma\right)$$

Corollary 10.5 (Anti-monotonicity of entropy)**.**

$$\beta \leqslant \gamma \implies H\left(\alpha \mid \beta\right) \geqslant H\left(\alpha \mid \gamma\right)$$

Corollary 10.6. *We have the two following properties as well:*

- $H\left(\alpha \mid \gamma\right) \leqslant H(\alpha)$ *(because always* $\gamma \leqslant \{X, \varnothing\}$*)*
- $H\left(\alpha \vee \beta \mid \gamma\right) \leqslant H\left(\alpha \mid \gamma\right) + H\left(\beta \mid \gamma\right)$

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:

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

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 $H_{n+m}(\alpha) \leqslant H_n(\alpha) + H_m(\alpha)$, i.e. these H_n form a sub-additive sequence $\mathbb{R}$-valued sequence and hence the limit $h(T, \alpha) := \lim_{n\to\infty} \frac{1}{n} H_n(\alpha)$ 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.