Ergodic Theory Notes

Thomas Chaplin

Contents

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

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

The following theorem will be very useful.

Theorem 1.1 (Hahn-Kolmogorov)**.** *Let $\mathcal{A}$ be an algebra on a space X. Suppose we have a function $\mu_0 : \mathcal{A} \to [0, \infty]$ which satisfies*

(i) ***Finite additivity:*** *Given $A_1, \ldots, A_n \in \mathcal{A}$ disjoint*

$$\mu_0\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu_0(A_i)$$

(ii) ***Sigma additivity:*** *Given $A_1, A_2, \cdots \in \mathcal{A}$ disjoint such that $\bigcup_i A_i \in \mathcal{A}$*

$$\mu_0\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_0(A_i)$$

Then μ_0 extends to a measure on $\sigma(\mathcal{A})$. Moreover, if μ_0 is σ-finite then this extension is unique.

2 Facts on Fourier Series

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

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

We define the partial Fourier sums by

$$S_N f(x) := \sum_{|n| \leq N} \hat{f}(n) e^{2\pi i (n \cdot x)}$$

Theorem 2.1 (Riemann-Lebesgue Lemma)**.** *For all $f \in L_1(\mathbb{T}^k)$,*

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

Theorem 2.2 (Riesz-Fisher Theorem)**.** $S_n f \to f$ *in* L^2 *for all* $f \in L^2(\mathbb{T}^k)$.

Theorem 2.3 (Fejér's Theorem)**.** *The average of the partial Fourier sums converges uniformly to f, i.e.*

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

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*

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

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

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 $N \geq 1$ 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 $n = (n_1, \ldots, n_k) \in \mathbb{Z}^k, x = (x_1, \ldots, x_k), |n| = |n_1| + \cdots + |n_k|$.

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

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

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.

Corollary 4.1. *The following are equivalent:*

1. *T is ergodic.*
2. *$B \in \mathcal{B}$ and $\mu(T^{-1}(B) \triangle B) = 0 \implies \mu(B) = 0$ or 1.*

Theorem 4.2. *Given a measure preserving transformation $T : X \to X$ 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.*

Proof. The difficult one to show is *(1)* $\implies$ *(2)*. Suppose that f is measurable and $f \circ T = f$ almost everywhere but f is not constant. So there is some $y \in \mathbb{R}$ such that $\mu\{f > y\} > 0$ and $\mu\{f < y\} > 0$. Then $T^{-1}\{f > y\} \triangle \{f > y\} = \{f \circ T > y\} \triangle \{f > y\} \subseteq \{f \circ T \neq f\}$. Hence $\mu(T^{-1}\{f > y\} \triangle \{f > y\}) \leq \mu(\{f \circ T \neq f\}) = 0$. So by the corollary, $\mu(\{f > y\}) = 0$ or 1. Either case leads to a contradiction. $\square$

Note: As a corollary to the Riesz-Fisher theorem, given any $f \in L^2$, if we have that $\hat{f}(n) = 0$ for all $n \neq 0$ then f must be constant. Therefore 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 4.3. *Given $f \in L^2$ which is invariant*

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

Example: Recipe for ergodicity:
We're going to show that the any invariant function must be constant by showing that all Fourier coefficients (away from $n = 0$) are 0.

1. Compute the partial sum $S_N f(Tx)$.
2. Use Theorem 4.3 to obtain a relationship between Fourier coefficients.
3. Use this relationship to equate absolute values of a sequences of Fourier coefficients.
4. Use the Riemann-Lebesgue Lemma to conclude that all coefficients away from $n = 0$ are vanishing.

Example: The doubling map is ergodic with respect to the Lebesgue Measure.

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*

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

Proof.

1. Define F to be the set of points in B who never return.

$$F := \{x \in B \mid T^n x \notin B \quad \forall n \geq 1\}$$

Then note the $T^{-n}F$ are all disjoint and of the same measure so we must have $\mu(F) = 0$.

2. Notice

$$x \notin E \iff \exists k \geq 1 \quad s.t. \quad T^k x \in F$$

and hence $B \setminus E = \cup_k T^{-k} F$ and so $\mu(B \setminus E) = 0$. We can conclude $\mu(E) = \mu(B)$.

$\square$

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*

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

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*

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

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

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

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

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

Then

$$\left|\left| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - f^* \right|\right|_p \leq \left|\left| \frac{1}{n} \sum_{j=0}^{n-1} (f-g) \circ T^j \right|\right|_p + \underbrace{\left|\left| \frac{1}{n} \sum_{j=0}^{n-1} g \circ T^j - g^* \right|\right|_p}_{<\epsilon/3 \text{ for } n \geq N} + \underbrace{||g^* - f^*||}_{<\epsilon/3}$$

$$\leq \frac{1}{n} \sum_{j=0}^{n-1} \left|\left| (f-g) \circ \cancel{T^j} \right|\right|_p^{\text{m.p.t}} + \frac{2\epsilon}{3}$$

$$= ||f - g|| + \frac{2\epsilon}{3} < \epsilon$$

□

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.
- $\forall k \in \{0, 1, \dots, b-1\}$

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

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

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*

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

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

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

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

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*

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

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:

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

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

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 (Riesz 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 $|\mu(f)| \leq ||f||_\infty$ 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 Riesz 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: $f \geq 0 \implies \mu_n(f) \geq 0$ for all n and hence $w(f) \geq 0$.

□

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 $x_1, c_2, \dots \in C$ 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 $||\mu_n - \mu_m|| \geq 1$. 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 $||\delta_{x_n} - \delta_{x_m}|| \geq 1$. □

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

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

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 Riesz 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

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

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

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

Also there is an N such that for $n \geq N$ 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 $n \geq N$ we have $d(\mu_n, \mu) \leq \epsilon$. 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

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

so there is an N such that for all $n \geq N$ 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)**.** $M(X,T) \neq \emptyset$.

Proof. We know that $M(X) \neq \emptyset$ (i.e. use a Dirac measure), so we can choose $\sigma \in M(X)$. Then also note that $T_*^j\sigma \in M(X)$ for all $j \geq 1$ Now define a sequence of measures by

$$\mu_n := \frac{1}{n}\sum_{j=0}^{n-1} T_*^J\sigma$$

Since $M(X)$ is convex we see that $\mu_n \in M(X)$ for every n. By weak $*$ compactness of $M(X)$ we get a convergent subsequence $\mu_{n_k} \to \mu \in M(X)$ weak $*$.
Now we need to show that $T^*\mu = \mu$ which we do via the Riesz Representation Theorem. Choose arbitrary $f \in C(X)$. Notice

$$|T_*\mu_n(f) - \mu_n(f)| = \left|\frac{1}{n}T_*^n\sigma(f) - \sigma(f)\right| = \frac{1}{n}\left|\sigma(f \circ T^n) - \sigma(f)\right| \leq \frac{2}{n}\left|\left|f\right|\right|_\infty$$

But then

$$|\mu_{n_k}(f \circ T) - \mu_{n_k}(f)| = |T_*\mu_{n_k}(f) - \mu_{n_k}(f)| \leq \frac{2}{n_k}\left|\left|f\right|\right|_\infty$$

Taking limits on both sides the left goes to $|T_*\mu(f) - \mu(f)|$ and the right goes to 0. $\square$

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.

Given measures μ, ν on a measure space $(X, \mathcal{B})$ we say that ν is absolutely continuous with respect to ν if

$$B \in \mathcal{B}, \quad \mu(B) = 0 \implies \nu(B) = 0$$

Theorem 9.12 (Radon-Nikodym Theorem)**.** *Given a measurable space and measures μ, ν. Suppose μ is σ-finite and $\nu << \mu$. Then there is a unique $h : X \to [0, +\infty]$ such that*

$$\forall B \in \mathcal{B} \quad \nu(B) = \int_B h \, d\mu$$

Then we write $h = \frac{d\nu}{d\mu}$.

Proposition 9.13. *Given $\mu, \nu \in M(X,T)$ and $\nu << \mu$ write $h = \frac{d\nu}{d\mu}$. Then h is a T-invariant function, i.e. $h \circ T = h$.*

Theorem 9.14. *$\mu \in M(X,T)$ is ergodic $\implies$ μ is extremal.*

Proof. Suppose that $\mu = (1-t)\mu_0 + t\mu_1$ for some $\mu_0, \mu_1 \in M(X,T)$ and $0 < t < 1$. We aim to show that in fact $\mu = \mu_0 = \mu_1$. Notice that for any $B \in \mathcal{B}$ we have that $\mu(B) \geq (1-t)\mu_0$ and hence $\mu_0 << \mu$. The Radon-Nikodym Theorem gives us a unique h such that

$$\mu_0(B) = \int_b h \, d\mu$$

Notice that μ_0 is a probability measure and hence $\int_X h \, d\mu = 1$. Also we have seen that $h \circ T = h$ and so by ergodicity we must have that h is constant. But since $\int_X h \, d\mu = 1$ we must have that $h \equiv 1$ almost everywhere and hence $\mu_0 = \mu$. $\square$

Theorem 9.15. *Suppose $T : X \to X$ is a continuous map on a compact metric space. Then there is a $\mu \in M(X,T)$ which is extremal and hence ergodic.*

9.4 Abundance and Uniqueness of Ergodic Measures

Example: Suppose that $T : X \to X$ is continuous and $x_0 \in X$ is a fixed point then $\delta_{x_0} \in M(X,T)$ is an ergodic, T-invariant measure.
Suppose that x_0 is a periodic points so that $T^q x_0 = x_0$ then define

$$\mu := \frac{1}{q} \sum_{j=0}^{k-1} \delta_{T^j x_0}$$

Then this is also T-invariant and ergodic. So the doubling map on S^1 has a countable infinity of ergodic, invariant probability measures arising in this way.

So the doubling map has infinitely many ergodic invariant probability measures. It also has one ergodic absolutely continuous invariant probability measure (a.c.i.p.) which is absolutely continuous wrt *Leb*, namely the Lebesgue measure itself. Are there any others?

Proposition 9.16. *Given $T : X \ toX$ a measure preserving transformation and $\mu, \nu \in M(X,T)$*

$$\nu << \mu \implies \nu = \mu$$

Proof. Choose some arbitrary $B \in \mathcal{B}$, we aim to show that $\nu(B) = \mu(B)$. By the ergodic theorem, there is a set E_μ such that $\mu(E_\mu) = 1$ and

$$\forall x \in E_\mu \quad \frac{1}{n}\sum_j \chi_b(T^j x) \to \mu(B)$$

And likewise there is some set E_ν with $\nu(E_\nu) = 1$ and

$$\forall x \in E_\nu \quad \frac{1}{n}\sum_j \chi_b(T^j x) \to \nu(B)$$

If we can find a point in $E_\mu \cap E_\nu$ then by the uniqueness of limits we are done.
Let $h = \frac{d\nu}{d\mu}$. Then note

$$\nu(E_\nu) = \int_{E_\nu} h \, d\mu = \int_X h \, d\mu = 1$$

because E_μ has full measure under μ and also $\nu(X) = 1$. So we see that $\nu(E_\mu) = \mu(E_\nu) = 1$ and hence their intersection has full measure and so is non-empty. □

Corollary 9.17. *The doubling map has a unique acip (with respect to Leb), namely the Lebesgue measure itself.*

We say a measurable $T : X \to X$ is uniquely ergodic if $|M(X,T)| = 1$.

Note: If a transformation T is uniquely ergodic then the unique T-invariant measure is certainly extremal in $M(X,T)$ and hence must be ergodic, justifying the name.

Proposition 9.18. *Irrational rotation $T : \mathbb{T} \to \mathbb{T}$, $x \mapsto x_\alpha$, for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ are uniquely ergodic.*

Proof. We're in a compact metric space and so we can use the Riesz representation theorem. We take some arbitrary $\mu \in M(X,T)$ and aim to show that $\int f \, d\mu = \int f \, dLeb$ for all $f \in C(X)$. We proceed by a density argument using the space of trigonometric polynomials.
Consider $e^{2\pi i n x}$

$$\int e^{2\pi i n x} \, d\mu = \int e^{e\pi i n (Tx)} \, d\mu = e^{2\pi i n \alpha} \int e^{2\pi i n x} \, d\mu$$

Notice $\alpha \notin \mathbb{Q}$ and hence so long as $n \neq 0$ then we have $e^{2\pi i n \alpha} \neq 1$. Therefore $\int e^{2\pi i n x} = 0$.
Now choose some arbitrary trig polynomial $P(x) = \sum_{|n| \leq q} c_n e^{e\pi i n x}$ Then we have just shown that $\int P \, d\mu = c_0 = \int P \, dLeb$. For arbitrary $f \in C(X)$ we proceed by the usual density argument. □

Theorem 9.19. *Given a continuous map $T : X \to X$ on a compact metric space, the following are equivalent:*

(a) For every $f \in C(X)$ there is some $c = c(f) \in \mathbb{R}$ such that

$$\frac{1}{n}\sum_j f \circ T^j \to c \quad \textit{uniformly on } X$$

(b) For every $f \in C(X)$ there is some $c = c(f) \in \mathbb{R}$ such that

$$\frac{1}{n}\sum_j f \circ T^j \to c \quad \textit{pointwise on } X$$

(c) There is some $\mu \in M(X,T)$ such that

$$\frac{1}{n}\sum_j f \circ T^j \to \int f \, d\mu \quad \textit{pointwise} \quad \forall f \in C(X)$$

(d) T is uniquely ergodic.

Proof. *(a) $\Longrightarrow$ (b).* Trivial.
(b) $\Longrightarrow$ (c). We let $c(f)$ be as in *(b)* and define $w : C(X) \to \mathbb{R}$ by $w(f) = c(f)$. Our aim is hence to show that w is a positve, normalised, bounded linear functional and so must correspond to a measure.
Note that w is linear. w is bounded because

$$\lim_{n\to\infty}\left|\left|\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j\right|\right|_\infty \leq ||f||_\infty$$

by the triangle inequality. So certainly $w \in C(X)^*$. We also have positivity and normalised fairly trivially. So by the Riesz Representation theorem there is a $\mu \in M(X)$ such that $w(f) = \int f \, d\mu$. We just need to show that μ is T-invariant.

$$\int f \circ T \, d\mu = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=0}^{n-1}(f \circ T) \circ T^j\right] = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=0}^{n-1}(f \circ T^j) + \cancelto{0}{\frac{f \circ T^n}{n}} - \cancelto{0}{\frac{f}{n}}\right] = \int f \, d\mu$$

(c) $\Longrightarrow$ (d). We want to show $M(X,T) = \{\mu\}$. Suppose $\nu \in M(X,T)$ so our aim is to show $\nu(f) = \mu(f)$ for all $f \in C(X)$ and then apply the Riesz Representation Theorem.

$$\begin{aligned}
\nu(f) &= \textstyle\int f \, d\nu = \frac{1}{n}\sum_{j=0}^{n-1}\int (f \circ T^j) \, d\nu \qquad \text{since } \nu \in M(X,T) \\
&= \textstyle\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int (f \circ T^j) \, d\nu \\
& \qquad\qquad \text{(1)} \\
&= \textstyle\int \left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j\right) \, d\nu \\
& \qquad\qquad \textit{by (c)} \\
&= \textstyle\int (\int f \, d\mu) \, d\nu \\
& \qquad\qquad \textit{const over prob measure} \\
&= \textstyle\int f \, d\mu = \mu(f)
\end{aligned}$$

Note in (1) we switched the limit and the integral. This is justified because $\frac{1}{n}\sum_{j=0}^{n-1}(f \circ T^j)$ is bounded above by $||f||_\infty$ and converges pointwise so we can use the DCT.

(d) $\implies$ *(a)*. Suppose that our unique ergodic measure is called μ and that *(a)* does not hold. So in particular *(a)* fails with $c(f) = \int_X f \, d\mu$, i.e. there is some $f \in C(X)$ such that

$$\frac{1}{n}\sum_j f \circ T^j \not\to \int f \, d\mu \quad \text{uniformly}$$

Then there are sequences $n_k \to \infty$ and $x_k \in X$ such that

$$\frac{1}{n_k}\sum_{j=0}^{n_k-1} f(T^j x_k) \not\to \int f \, d\mu$$

We aim to contradict the uniqueness of μ so we set

$$\nu_k := \frac{1}{n_k}\sum_{j=0}^{n_k-1} \delta_{T^j x_k}$$

Note that $\nu_k(f) = \frac{1}{n_k}\sum_{j=0}^{n_k-1} f(T^j x_k)$. Then $\nu_k \in M(X)$ and $M(X)$ is weak $*$ compact so by passing to a subsequence we can assume that $\nu_k \to \nu$ weak $*$ for some $\nu \in M(X)$. By the same proof as the Krylov-Bogulyvhov Theorem $\nu \in M(X,T)$. It remains to show that $\nu \neq \mu$.

$$\nu(f) = \int_f d\nu = \lim_{k\to\infty}\int f \, d\nu_k = \lim_{k\to\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1} f(T^j x_k) \neq \int f \, d\mu = \mu(f)$$

□

10 Shifts

Most of this is verbatim from Dynamical Systems. There is something on the cylinder generating the Borel σ-algebra which is probably worth reading.

10.1 Bernoulli Measures

Take $\Sigma^+ := \{1,\dots,k\}^{\mathbb{N}}$ and let σ denote the natural shift map. Fix some $p = (p_1,\dots,p_k) \in \mathbb{R}^q$ such that each $p_i \geq 0$ and $\sum p_i = 1$. Now we define the measure of cylinders by the formula

$$\mu[y_0,\dots,y_m] := p_{y_0} p_{y_1} \dots p_{y_m}$$

We claim that this extends to a measure on the Borel σ-algebra. This will come to be known as a Bernoulli measure.

Proposition 10.1. *μ extends uniquely to a probability measure on Σ^+.*

Proof. Define $\mathcal{A} := \{\text{finite unions of cylinder sets}\}$. This is an algebra and it generates the Borel σ-algebra. Given $A \in \mathcal{A}$ we can write A uniquely as a finite disjoint union of cylinders C_i

$$A = \bigcup_{i=1}^{p} C_i \quad \implies \quad \mu(A) = \sum_{i=1}^{p} \mu(C_i)$$

We would like to use Hahn-Kolmogorov Theorem and hence we need to show that given $A_1, A_2, \dots \in \mathcal{A}$ which are disjoint such that $\cup_i A_i \in \mathcal{A}$ then $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ But in fact any such union must be finite and hence the condition trivial holds. So we get a unique probability measure on all of Σ^+, it is certainly a probability measure because

$$\mu(\Sigma^+) = \mu(\bigcup_{i=1}^{k}[i] = \sum_{i=1}^{k}\mu\,[i] = \sum_{i=1}^{k}p_i = 1$$

$\square$

Proposition 10.2. *μ is σ-invariant.*

Proof. Note by the uniqueness on $H - K$ it suffices to prove $\mu(\sigma^{-1}A) = \mu(A)$ for each $A \in \mathcal{A}$. $\square$

Proposition 10.3. *If $A \in \mathcal{A}$ then there is an $n \geq 0$ such that $\mu(A \cap \sigma^{-n}A) = \mu(A)^2$.*

Theorem 10.4. *Bernoulli measures are ergodic.*

Proof. This is definitely worth proving. $\square$

Corollary 10.5. *There are uncountably many distinct ergodic, invariant probability measures on Σ^+.*

We have the standard semi-conjugacy $\pi : \Sigma_k^+ \to [0,1)$ from the full one-sided shift on k symbols to $[0,1)$ given by

$$\pi(x_0, x_1, x_2, \dots) := \sum_{k=0}^{\infty}\frac{x_j}{k^{j+1}}$$

We can use this to push Bernoulli measures on Σ_k^+ forward into measures for $[0,1)$ by

$$\pi_*\mu(A) := \mu(\pi^{-1}A)$$

One can easily show that these will be invariant and ergodic measures for $[0,1)$ with the map $Tx = kx \mod 1$. But do different Bernoulli measures push forward to different measures on $[0,1)$?

Proof. Yes! Take two Bernoulli measures associated to starting distributions $\mathbf{p}$ and $\mathbf{q}$ such that for some specific j we know $p_j \neq q_j$. Then in the associated Bernoulli measures we notice

$$\mu_p([j]) = p_j \neq q_j = \mu_q([j])$$

Take $B := \left[\frac{j}{k}, \frac{j+1}{k}\right)$. Then $\pi^{-1}(B) = [j]$ and hence

$$\pi_*\mu_p(B) = p_j \neq q_j = \pi_*\mu_q(B)$$

$\square$

10.2 Markov Measures

We will need the Perron-Frobenius Theorem.

Theorem 10.6 (Perron-Frobenius)**.** *For an aperiodic matrix B*

- *$\exists\lambda > 0$ eigenvalue of B such that for all remaining eigenvalues $\lambda' \leq |\lambda|$.*
- *λ is simple so there is a unique eigenvector $\boldsymbol{v}$ such that $B\boldsymbol{v} = \lambda\boldsymbol{v}$ and $\sum_i v_i = 1$.*

- *This $\boldsymbol{v}$ is positive, i.e. $v_i > 0$ for every i.*
- *If w is another eigenvector then some of its coordinates are positive and others are negative.*

We call such λ the maximal eigenvalue.

Consider a subshift of finite type associated to an aperiodic matrix $A \in \{0,1\}^{k\times k}$. Choose some $k \times k$ matrix P with non-negative entries such that

- P is a row-stochastic if for all i, $\sum_{j=1}^{k} p_{ij} = 1$.
- P is compatible with A, i.e. $p_{ij} > 0 \iff A_{ij} = 1$.

Note that we get the following desirable properties:

- A aperiodic $\implies$ P aperiodic since the same choice of n will work.
- $(1,1,\ldots,1)^T$ is an eigenvector with eigenvalue 1.
- By the Perron-Frobenius Theorem, 1 must be the maximum eigenvalue. Therefore, there must be a unique left eigenvector q such that

$$qP = d \quad \sum_i q_i = 1 \quad q_i > 0$$

We can use this to define a new measure by first defining on cylinder sets

$$\mu([y_0,\ldots,y_m]) := q_{y_0} P_{y_0 y_1} \ldots P_{y_{m-1} y_m}$$

By the Hahn-Kilogram Theorem this extends to a unique measure on Σ_A^+ and note

$$\mu(\Sigma_A^+) = \mu\left(\bigcup_{i=1}^{k}[j]\right) = \sum_{j=1}^{k} \mu([j]) = \sum_{j=1}^{k} q_j = 1$$

So μ is a probability measure and μ is the unique such extension. We call μ the Markov measure corresponding to P.

Proposition 10.7. *Markov measures are invariant under the shift map.*

Proof.

$$\begin{aligned}
\mu(\sigma^{-1}[y_0,\ldots,y_m]) &= \mu\left(\cup_{i=1}^{k}[j,y_0,\ldots,y_m]\right) \\
&= \sum_{j=1}^{k} \mu[j,y_0,y_1,\ldots,y_m] \\
&= \sum_{j=1}^{k} q_j P_{jy_0} P_{y_0y_1} \ldots P_{y_{m-1}y_m} \\
&= \underbrace{\left(\sum_{j=1}^{k} q_j P_{jy_0}\right)}_{=q y_0} P_{y_0y_1} \ldots P_{y_{m-1}y_m} = \mu[y_0,\ldots,y_m]
\end{aligned}$$

□

Note: Markov measure are in fact ergodic but we are not going to prove it.

We can recover Bernoulli measures as a special case of Markov measures. Given our initial distribution $q = (q_1, \dots, q_k)$ we form the matrix P by

$$P := \begin{pmatrix} q_1 & \cdots & q_k \\ \vdots & \ddots & \vdots \\ q_1 & \cdots & q_k \end{pmatrix}$$

and then we can easily see that $qP = q$ and the corresponding Markov measure coincides with the corresponding Bernoulli measure.
Given any aperiodic matrix A with maximum eigenvalue $\lambda > 0$ we know that there are unique vectors u, v such that

$$\begin{aligned} Av &= \lambda v \sum_i v_i = 1 & v_i > 0 \\ uA^T &= \lambda u \sum_i u_i = 1 & u_i > 0 \end{aligned}$$

Then we define a new matrix P by $P_{ij} := \frac{A_{ij} v_i}{\lambda v_i}$. We also define a distribution $q_i := \frac{u_i v_i}{c}$ where $c := \sum_i u_i v_i$. Then we can see that P is compatible with A and is now row-stochastic. Moreover, $qP = q$ and $\sum_i q_i = 1$ and $q_i > 0$ for all i. The corresponding Markov measure μ is then called the Parry measure.

11 Mixing

Theorem 11.1. *Suppose $T : X \to X$ is a measure preserving transformation on $(X, \mathcal{B}, \mu)$. The following are equivalent:*

(a) T is ergodic.

(b) For all $A, B \in \mathcal{B}$ we have
$$\frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j} A \cap B) \to \mu(A)\mu(B)$$

Proof. *(b)* $\implies$ *(a)*. Choose $B \in \mathcal{B}$ such that $T^{-1}B = B$. Then in the formula take $A = B$ so that $T^{-j} A \cap B = B$. Then we can see that $\mu(B) = \mu(B)^2$ and hence $\mu(B) \in \{0, 1\}$.
(a) $\implies$ *(b)*. We apply the ergodic theorem for $f = \mathbb{1}_A$. We get that

$$\frac{1}{n} \sum_{0}^{n-1} \mathbb{1}_A \circ T^j \to \mu(A) \quad a.e.$$

Let's multiply this expression by $\mathbb{1}_B$.

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{T^{-j} A \cap B} = \frac{1}{n} \sum_{j=0}^{n-1} (\mathbb{1}_A \circ T^j) \mathbb{1}_B \to \mu(A) \mathbb{1}_B \quad a.e.$$

Then we integrated both sides and use the dominated convergence theorem since the left hand side is bounded by 1 which has finite integral. Hence

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)=\int\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_{T^{-j}A\cap B}\to\int\mu(A)\mathbb{1}_B=\mu(A)\mu(B)$$

□

We say that T is weak mixing if

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

and say that T is (strong) mixing if

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

each for all $A, B\in\mathcal{B}$.
A subset $J\subseteq\mathbb{N}$ is of density d if

$$\frac{|J\cap\{0,1,\ldots,n-1\}|}{n}\to d\quad\text{as } n\to\infty$$

We say J has

- full density if $d=1$,
- zero density if $d=0$, and
- positive density if $d>0$.

Notes on density:

- If d exists then certainly $d\in[0,1]$.
- If J has density d then $\mathbb{N}\setminus J$ has density $1-d$.
- If T is ergodic and $B\in\mathcal{B}$ is a Borel set then

$$J:=\{n\geq 0\mid T^n x\in B\}\ \text{ has density } \mu(B) \text{ for a.e. } x$$

Proof.

$$\frac{|J\cap\{0,1,\ldots,n-1\}|}{n}=\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_B(T^j x)\to\mu(B)$$

by the ergodic theorem. □

Lemma 11.2. *Let $a_n\in\mathbb{R}$ be a bounded sequence. The following are equivalent:*

1. $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}|a_j|=0$
2. *There is a subset $J\subseteq\mathbb{N}$ of full density such that* $\lim_{n\to\infty,\, n\in J}a_n=0$.

Proof. The proof of this is unreasonably long. $\square$

Corollary 11.3. *T is weak-mixing if and only if for all $A, B \in \mathcal{B}$ there is some $J \subseteq \mathbb{N}$ of full density such that*

$$\lim_{n\to\infty,\, n\in J} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$$

Proof. Take $a_n = \mu(A \cap T^{-n}B) - \mu(A)\mu(B)$ in the above lemma. $\square$

Theorem 11.4. *T is weak mixing $\implies$ $T \times T$ is ergodic (in fact weak mixing). Note $T \times T$ is taking place under the product measure on the product σ-algebra.*

Corollary 11.5. *Rotations of the circle are not weak mixing.*

Proof. It suffices to check that $T \times T$ is not ergodic and hence we will try and find a non-constant invariant function for $T \times T$. Let's take

$$f(x, y) := e^{2\pi i x} e^{-2\pi i y}$$

then $f : \mathbb{T} \times \mathbb{T} \to \mathbb{C}$ is measurable in L^∞ and is not constant. Now notice

$$f(x + \alpha, y + \alpha) = e^{2\pi i x} e^{-2\pi i y} e^{2\pi i \alpha} e^{-2\pi i \alpha} = f(x, y)$$

and hence $f \circ (T \times T) = f$. $\square$

We now know

$$\begin{tikzcd}
\text{Mixing} \arrow[r, Rightarrow, bend left] & \text{Weak mixing} \arrow[l, Rightarrow, bend left, "/" marking, "\text{Hard to prove}"] \arrow[r, Rightarrow, bend left] & \text{Ergodic} \arrow[l, Rightarrow, bend left, "/" marking, "\text{irrational rotation}"]
\end{tikzcd}$$

Proposition 11.6. *The shift map $\sigma : \Sigma_+ \to \Sigma_+$ with a Bernoulli measure is mixing.*

Proof. When we proved ergodicity we say that for sufficiently large n

$$\mu(A \cap \sigma^{-n}A) = \mu(A)^2$$

It is an exercise to show that the same argument gives $\forall A, A' \in \mathcal{A}$

$$\mu(A \cap \sigma^{-n}A') = \mu(A)\mu(A') \quad \forall n \text{ sufficiently large}$$

Then given arbitrary $B, B' \in \mathcal{B}$ we can find $A, A' \in \mathcal{A}$ such that $\mu(A \triangle B) < \epsilon)$ and $\mu(A' \triangle B') < \epsilon$. Then we can see

$$\left|\mu(B \cap \sigma^{-n}B' - \mu(B)\mu(B')\right| \le \left|\mu(A \cap \sigma^{-n}A') - \mu(A)\mu(A')\right| + 4\epsilon \to 0$$

$\square$

11.1 Correlation Function / Covariance

We would like to say something about the rate of mixing and to this end we define the following:

The correlation function / covariance is

$$\mathrm{Cov}(f, g \circ T^n) := \int_X (f)(g \circ T^n)\, d\mu - \int_X f\, d\mu \int_X g \circ T^n\, d\mu$$

We will also need the following for a function $f : \Sigma^+ \to \mathbb{R}$

$$\mathrm{lip}(f) := \sup_{x,y \in \Sigma^+,\, x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}$$

Given $T : X \to X$ a measure preserving transformation then

$$\begin{aligned} Cov(f, g \circ T^n) &= \int +X f \cdot g \circ T^n\, d\mu - \textstyle\int_X f\, d\mu \int_X g \circ T^n\, d\mu \\ &= \textstyle\int_X \left(f - \int_X f\, d\mu\right) g \circ T^n\, d\mu \\ &= \textstyle\int_X \left(f - \int_X f\, d\mu\right) U^n g\, d\mu \\ &= \textstyle\int_X L^n \left(f - \int_X f\, d\mu\right) g\, d\mu \end{aligned}$$

Koopman operator

$L := U^*$

Now notice the following about L

1. $$\int L1\overline{g} = \int 1\,(\overline{g} \circ T) = \int \overline{g} \circ T = \int \overline{g} = \int 1\overline{g}$$

   since T is measure preserving. Hence $L1 = 1$.

2. $$\int Lf = \int Lf\ 1 = \int (f)(1 \circ T) = \int f\ 1 = \int f$$

   and hence

   $$L^n \int f\, d\mu = \int f\, d\mu = \int L^n f\, d\mu$$

Therefore we can conclude that

$$\mathrm{Cov}(f, g \circ T^n) = \int_X \left(L^n f - \int L^n f\right) f\, d\mu$$

So by applying the Cauchy-Schwarz inequality we can see

Proposition 11.7. *For all $f, g \in L^2$, and T a measure preserving transformation*

$$|\mathrm{Cov}(g, g \circ T^n)| \leq \left|\left| L^n f - \int L^n f \right|\right|_2 ||g||_2$$

Proposition 11.8.

$$\left|\left| f - \int f \right|\right|_\infty \leq \mathrm{lip}(f)\, \mathrm{diam}(X)$$

for $f : X \to \mathbb{R}$ Lipschitz on a metric space (X, d) where

$$\mathrm{diam}(X) := \sup_{x,y \in X} d(x,y)$$

Proof. Choose some $x \in X$. Suppose $f(x) > \int f$. Then there must be some y such that $f(y) \leq \int f$. Hence $f(x) - f(y) \geq f(x) - \int f > 0$. Similarly if $f(x) < \int f$. So we have

$$\begin{aligned} \left| f(x) - \int f \right| &\leq \sup_{y \in X} |f(x) - f(y)| \\ &\leq \sup_{y \in X} \operatorname{lip}(f) d(x,y) \\ &\leq \operatorname{lip}(f) \operatorname{diam}(X) \end{aligned}$$

□

Now take some Bernoulli measure μ associated to a distribution $(p_1, \ldots, p_k)$. One can show, using a similar method as for the doubling map that

Proposition 11.9. *For any $f \in L^2(X)$ we have*

$$(Lf)(x) = \sum_{j=1}^{k} p_j f(jx)$$

where $jx = (j, x_0, x_1, \ldots)$.

Lemma 11.10. *Given $f : \Sigma^+ \to \mathbb{R}$ Lipschitz we have*

$$\operatorname{lip}(L^n f) \leq \frac{1}{2^n} \operatorname{lip}(f)$$

Proof. Note because of our choice of metric we have that $d(jx, jy) = \frac{1}{n} d(x,y)$ for any $x, y \in \Sigma^+$.

$$\begin{aligned} |(Lf)(x) - (Lf)(y)| &\leq \sum_{j=1}^{k} p_j |f(jx) - f(jy)| \leq \sum_{j=1}^{k} p_j \operatorname{lip}(f) d(jx, jy) \\ &= \sum_{j=1}^{k} p_j \operatorname{lip}(f) \frac{1}{2} d(x,y) = \frac{1}{2} \operatorname{lip}(f) d(x,y) \sum_{j=1}^{k} p_j = \frac{\operatorname{lip}(f)}{2} d(x,y) \end{aligned}$$

Then the result follows by induction. □

Then we get our big theorem on mixing

Theorem 11.11. *On Σ^+ with a Bernoulli measure, we have exponential mixing rates. That is given $f : \Sigma^+ \to \mathbb{R}$ Lipschitz, $g \in L^2(\Sigma^+)$ and $n \geq 1$ we have*

$$|\operatorname{Cov}(f, g \circ \sigma^n)| \leq \left(\frac{1}{2}\right)^n \operatorname{lip}(f) \, ||g||_2$$

Proof.

$$\begin{aligned} |\operatorname{Cov}(f, g \circ \sigma^n)| &\leq \left|\left| L^n f - \int_{\Sigma^+} L^n f \, d\mu \right|\right|_\infty ||g||_2 \leq \operatorname{lip}(L^n f) \underbrace{\operatorname{diam}(\Sigma^+)}_{=1} ||g||_2 \\ &= \operatorname{lip}(L^n f) \, ||g||_2 \leq \left(\frac{1}{2}\right)^n \operatorname{lip}(f) \, ||g||_2 \end{aligned}$$

□

It was an exercise to show the same thing for $g \in L^1(\Sigma^+)$, using the 1-norm on g. It was also an exercise to estimate mixing rates for the doubling map when f is Lipschitz or Hölder continuous.

12 Entropy

12.1 Conditional Expectation

Suppose we have a probability space $(X, \mathcal{B}, \mu)$ and a sub σ-algebra $\mathcal{A} \subseteq \mathcal{B}$. Our aim is to define a conditional expectation operator

$$\mathbb{E}(\bullet|\mathcal{A}) : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$$

with the property that $\mathbb{E}(\bullet|\mathcal{A})^2 = \mathbb{E}(\bullet|\mathcal{A})$ We would also like that

$$\int_A \mathbb{E}(f|\mathcal{A}) \, d\mu = \int_A f \, d\mu \quad \forall A \in \mathcal{A}$$

1. First take $f \in L^1(X, \mathcal{B}, \mu)$ such that $f \geq 0$ and define

$$\nu(A0 := \int_A f \, d\mu \quad \forall A \in \mathcal{A}$$

   This gives us a measure on $\mathcal{A}$. Notice that $\mu\big|_{\mathcal{A}}$ is also a measure on $\mathcal{A}$ and in fact we claim that $\nu << \mu\big|_{\mathcal{A}}$. So by the Radon-Nikodym theorem there is a unique $h_f \geq 0$ such that h_f is $\mathcal{A}$-measurable and

$$\nu(A) = \int_A h_f \, d\mu \quad \forall A \in \mathcal{A}$$

   We then define $\mathbb{E}(f|\mathcal{A}) := h_f$ which has the property that it is $\mathcal{A}$-measurable and has the same integral as f. This can be seen as the "best approximation to f that is $\mathcal{A}$-measurable".

2. For arbitrary $f \in L^1(X, \mathcal{B}, \mu)$ we write $f = f^+ - f^-$ and then define

$$\mathbb{E}(f|\mathcal{A}) := \mathbb{E}(f^+|\mathcal{A}) - \mathbb{E}(f^-|\mathcal{A})$$

Proposition 12.1. *The conditional expectation operator is uniquely defined by the properties*

(i) $\mathbb{E}(f|\mathcal{A})$ *is* $\mathcal{A}$ *measurable.*

(ii) $\int_A \mathbb{E}(f|\mathcal{A}) \, d\mu = \int_A f \, d\mu$ *for all* $f \in \mathcal{A}$.

Lemma 12.2. *The conditional expectation operator is linear.*

Note that if f is already $\mathcal{A}$ measurable then $\mathbb{E}(f|\mathcal{A})$ because f already satisfies *(i)* and *(ii)*. Also if $\mathcal{A} = \{X, \emptyset\}$ then

$$\mathbb{E}(f| \{X, \emptyset\}) = \int_X f \, d\mu$$

Proposition 12.3. *Given probability spaces* $(X, \mathcal{B}_X, \mu_X)$ *and* $(Y, \mathcal{B}_Y, \mu_Y)$ *and a measure preserving transformation* $T : X \to Y$, *let* $\mathcal{A} \subseteq \mathcal{B}_Y$ *be a sub* σ*-algebra. Then for all* $f \in L^1(Y, \mathcal{B}_Y, \mu_Y)$ *we have*

$$\mathbb{E}\left(f \mid \mathcal{A}\right) \circ T = \mathbb{E}\left(f \circ T \,\middle|\, T^{-1}\mathcal{A}\right)$$

Proof. It suffices to check that the left hand side is $T^{-1}\mathcal{A}$-measurable and that

$$\int_{T^{-1}A} \mathbb{E}\left(f \mid \mathcal{A}\right) \circ T = \int_{T^{-1}A} f \circ T \qquad \forall A \in \mathcal{A}$$

(i) Note that $\mathbb{E}(f \mid \mathcal{A})$ is already $\mathcal{A}$-measurable and hence the left hand side is $T^{-1}\mathcal{A}$-measurable.

(ii)

$$\begin{aligned}
\textstyle\int_{T^{-1}A} \mathbb{E}(f \mid \mathcal{A}) \circ T \, d\mu_X &= \textstyle\int_X (\mathbb{1}_A \mathbb{E}(f \mid \mathcal{A})) \circ T \, d\mu_X && \text{\textit{T m.p.t.}} \\
&= \textstyle\int_Y \mathbb{1}_A \mathbb{E}(f \mid \mathcal{A}) \, d\mu_Y \\
&= \textstyle\int_A \mathbb{E}(f \mid \mathcal{A}) \, d\mu_Y && A \in \mathcal{A} \\
&= \textstyle\int_A f \, d\mu_Y && \text{\textit{same tricks}} \\
&= \textstyle\int_X (\mathbb{1}_A f) \circ T \, d\mu_X \\
&= \textstyle\int_{T^{-1}A} f \circ T \, d\mu_X
\end{aligned}$$

□

We can also prove the following by the same technique.

Proposition 12.4. *Given a measure space $(X, \mathcal{B}, \mu)$ and nested sub σ-algebras $\widetilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{B}$*

$$\mathbb{E}\left(\mathbb{E}(f \mid \mathcal{A}) \,\middle|\, \widetilde{\mathcal{A}}\right) = \mathbb{E}\left(f \,\middle|\, \widetilde{\mathcal{A}}\right) \qquad \forall f \in L^1(X, \mathcal{B}, \mu)$$

Given a measure preserving transformation $T : X \to X$ we know that $U^*U = I$ but it is an exercise to show that

$$UU^* = \mathbb{E}\left(\bullet \,\middle|\, T^{-1}\mathcal{B}\right)$$

12.2 An Ergodic Aside

Recall the ergodic theorem. Suppose we have a measure preserving transformation $T : X \to X$ on a measure space $(X, \mathcal{B}, \mu)$ which is not necessarily ergodic. IF $f \in L^1(X)$ then almost everywhere we have

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

where $\widetilde{f} \in L^1(X)$ and $\int \widetilde{f} \, d\mu = \int f \, d\mu$ and $\widetilde{f} \circ T = \widetilde{f}$.
When $f \in L^2(X)$ we found out that $\widetilde{f} = \pi f$ where π is the orthogonal projection onto

$$\ker(U - I) = \{T\text{-invariant functions}\}$$

We would like a formula for arbitrary $f \in L^1(X)$. To this end define the set

$$I := \left\{B \in \mathcal{B} \,\middle|\, T^{-1}B = B\right\}$$

then we can see that $I \subseteq \mathcal{B}$ is a sub σ-algebra of $\mathcal{B}$.

Theorem 12.5. *In the ergodic theorem $\widetilde{f} = \mathbb{E}(f \mid I)$ for all $f \in L^1(X)$.*

Proof. We do the same style of proof.

(i) We wish to show that $\widetilde{f}$ is I-measurable. We know that it is $\mathcal{B}$-measurable so take $U \in \mathcal{B}(R)$ then

$$\widetilde{f}^{-1}(U) = (\widetilde{f} \circ T)^{-1}(U) = T^{-1}(\widetilde{f}^{-1}U)$$

and hence $\widetilde{f}^{-1}U \in I$.

(ii) Take any $A \in I$ then we have that

$$\frac{1}{n}\sum_{j=0}^{n-1}(\mathbb{1}_A f)\circ T^j = \mathbb{1}_A \cdot \frac{1}{n}\sum_{j=0}^{n-1} f\circ T^j \to \mathbb{1}_A \widetilde{f} \quad a.e.$$

because $A \in I$ means we can move the indicator function outside. So we can conclude

$$\int_A f = \int_X \mathbb{1}_A f = \int_X \mathbb{1}_A \widetilde{f} = \int_A \widetilde{f}$$

□

12.3 Defining 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
- $\pi \circ T = S \circ \pi$

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\ Y & \xrightarrow[S]{} & Y \end{array}$$

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 $A_i \cap A_j = \emptyset$ 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 $I(\alpha) : X \to [0, \infty]$.

- 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

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

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

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

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 $\mathcal{C} = \{X, \emptyset\}$ then $I\left(\alpha \mid \mathcal{C}\right) = I(\alpha)$ and $H\left(\alpha \mid \mathcal{C}\right) = H(\alpha)$.
- If $g \geq 0$ 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 $I\left(T^{-1}\alpha \;\middle|\; T^{-1}\mathcal{C}\right) = I\left(\alpha \mid \mathcal{C}\right) \circ T$.

- Integrating this gives $H\left(T^{-1}\alpha \mid T^{-1}\mathcal{C}\right) = H(\alpha \mid \mathcal{C})$.
- $\alpha \leq \beta \implies I(\alpha \mid \beta) = 0$.

Proposition 12.6.

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

Proof. Use the definition and then write out $I(\alpha \mid \mathcal{C})$ explicit. Use the monotone convergence theorem to swap sum and limit. Change the integrand to the conditional expectation on $\mathcal{C}$ and then pull out the $\log(\mu(A \mid \mathcal{C}))$ term. □

Lemma 12.7 (Basic Identity)**.** *Given α, β, γ partitions of X then*

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

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

Proof. This can be proved by showing equality on each partition element. That is, choose $A \in \alpha$, $B \in \beta$, $C \in \gamma$ and then show equality for $x \in A \cap B \cap C$. □

Corollary 12.8.

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

Corollary 12.9 (Monotonicity of information of entropy)**.**

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

Corollary 12.10 (Anti-monotonicity of entropy)**.**

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

Corollary 12.11. *We have the two following properties as well:*

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

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:

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

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 $h(T, \alpha) := \lim_{n\to\infty} \frac{1}{n} H_n(\alpha)$ 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.

Theorem 12.12. *Given two measure preserving transformations $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ that are measure-theoretically isomorphic*

$$h(T) = h(S)$$

Proof. Choose $\alpha \in \mathcal{P}_Y$ some countable partition over Y then we get a partition over C. Then $\pi^{-1}\alpha \in \mathcal{P}_B$ which gives us a countable partition for X, $\pi^{-1}\alpha \in \mathcal{P}_X$.

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\pi^{-1}\alpha)\right) = H_\mu\left(\pi^{-1}\bigvee_{j=0}^{n-1} S^{-j}\alpha\right) = H_\nu\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right)$$

with the last equality holding because π is measure preserving. Dividing by n and taking $n \to \infty$ we see that

$$h_\mu(T, \pi^{-1}\alpha) = h\nu(S, \alpha)$$

Then taking supremums over $\alpha \in \mathcal{P}_Y$ corresponds to taking supremums over $\pi^{-1}\alpha \in \mathcal{P}_X$. Hence $h\mu(T) = h_\nu(S)$. □

Note: Given a countable partition α we can apparently write

$$h(T, \alpha) = \int I\left(\alpha \,\middle|\, \bigvee_{j=1}^{\infty} T^{-j}\alpha\right) d\mu$$

Theorem 12.13. *(i) For any $k \geq 0$ we have $h(T^k) = kh(T)$.*

(ii) Moreover if T is invertible with measure preserving inverse then for any $k \in \mathbb{Z}$ we have $k(T^k) = |k|\, h(T)$.

12.4 Abramov and Sinai Theorems

Given countable partitions α_n and a σ-algebra $\mathcal{B}$, we say $\alpha_n \uparrow \mathcal{B}$ if

$$\alpha_1 \leq \alpha_2 \leq \dots \quad \text{and} \quad \sigma\left(\bigcup_{n=1}^{\infty} \alpha_n\right) = \mathcal{B}$$

Lemma 12.14. *Given countable partitions $\alpha, \beta \in \mathcal{P}$*

$$h(T, \alpha) \leq h(T, \beta) + H(A \mid B)$$

Proof. By monotonicity we can see that $H(\alpha^n) \leq H(\alpha^n \vee \beta^n) = H(\beta^n) + H(alpha^n \mid \beta^n)$

$$\begin{aligned} H(\alpha^n \mid \beta^n) &= H\left(\textstyle\bigvee_{j=0}^{n-1} T^{-j}\alpha \,\middle|\, \bigvee_{i=0}^{n-1} T^{-i}\beta\right) \leq \textstyle\sum_{j=0}^{n-1} H\left(T^{-j}\alpha \,\middle|\, \bigvee_{i=0}^{n-1} T^{-i}\beta\right) && \textit{anti-monotonicity} \\ &\leq \textstyle\sum_{j=0}^{n-1} H\left(T^{-j}\alpha \,\middle|\, T^{-j}\beta\right) && \textit{m.p.t} \\ &= nH(\alpha \mid \beta) \end{aligned}$$

□

Lemma 12.15. *Suppose $\alpha, \alpha_n \in \mathcal{P}$ and $\alpha_n \uparrow \mathcal{B}$ then*

$$\lim_{n\to\infty} H(\alpha \mid \alpha_n) = 0$$

Theorem 12.16 (Abramov's Theorem)**.** *Given a probability space* $(X, \mathcal{B}, \mu)$ *and* $T : X \to X$ *a measure preserving transformation, let* $\alpha_n \in \mathcal{P}$ *such that* $\alpha_n \uparrow \mathcal{B}$*. Then*

$$\lim_{n\to\infty} h(T, \alpha_n) = h(T)$$

Proof. Well certainly $h(T) \geq \limsup_{n\to\infty} h(T, \alpha_n)$. Conversely let $\alpha \in \mathcal{P}$, then by the first Lemma we see

$$h(T, \alpha) \leq h(T, \alpha_n) + H(\alpha \mid \alpha_n)$$

Then taking $n \to \infty$ and using the second lemma we get

$$h(T, \alpha) \leq \liminf_{n\to\infty} h(T, \alpha_n)$$

□

If T is invertible and $\bigvee_{j=-n}^{n} T^{-j}\alpha \uparrow \mathcal{B}$ then we say α is a generator.
If $\bigvee_{j=0}^{n} T^{-j}\alpha \uparrow \mathcal{B}$ then we say α is a strong generator.

Theorem 12.17 (Sinai's Theorem)**.** *Let* $(X, \mathcal{B}, \mu)$ *be a probability space and* $T : X \to X$ *a measure preserving transformation. Given* $\alpha \in \mathcal{P}$*, if*

1. *T is invertible and α is a generator OR*
2. *α is a strong generator*

then $h(T) = h(T, \alpha)$*.*

Example: The doubling map has entropy $\log(2)$
Define $\alpha = \left\{\left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right]\right\}$. Then notice

$$\alpha^n = \left\{\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right] \,\middle|\, 0 \leq j \leq 2^n - 1\right\}$$

is a strong generator and $h(T) = h(T, \alpha)$
But then

$$H(\alpha^n) = \sum_{j=1}^{2^n} -\frac{1}{2^n}\log\left(\frac{1}{2^n}\right) = \log(2^n) = n\log(2)$$

and hence $h(T, \alpha) = \log(2)$.

Theorem 12.18 (Entropy of a Markov Measure)**.** *Given an aperiodic* $k \times k$ *matrix* A *with entries on* $\{0, 1\}$ *take some row-stochastic matrix* P *compatible with* A*. By the Perron-Frobenius Theorem we get a unique vector* $\boldsymbol{q}$ *with positive entries who sum to* 1 *and such that* $\boldsymbol{q}P = \boldsymbol{q}$*. The entropy of the shift map under the corresponding Markov measure is then*

$$h(T) = -\sum_{i,j=1}^{k} q_i P_{ij} \log P_{ij}$$

Proof. We're going to define a strong generator

$$\alpha := \{[1], [2], \ldots, [k]\}$$

Then $\sigma^{-1}\alpha = \{[*, 1], [*, 2], \ldots, [*, k]\}$. So therefore we see that

$$\alpha^n = \left\{[x_0, x_1, \ldots, x_{n-1}] \subseteq \Sigma^+\right\}$$

that is the set of all admissible cylinders of length k. Therefore

$$\begin{aligned}
H(\alpha^{n+1}) &= -\sum_{x_0=1}^{k} \cdots \sum_{x_n=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n} (\log q_{x_0} + \log P_{x_0 x_1} + \cdots + \log P_{x_{n-1} x_n}) \\
&= -\sum_{x_0=1}^{k} \cdots \sum_{x_n=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n} \log q_{x_0} \\
&\quad - \sum_{x_0=1}^{k} \cdots \sum_{x_n=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n} (\log P_{x_0 x_1} + \cdots + \log P_{x_{n-1} x_n})
\end{aligned}$$

Now we're going to deal with each of these terms in turn. In the first term we can use the row-stochasticty to peel off each sum

$$\begin{aligned}
&\textstyle\sum_{x_0=1}^{k} \cdots \sum_{x_n=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n} \log q_{x_0} \qquad \textit{Sum last } P \textit{ term} \\
&= \textstyle\sum_{x_0=1}^{k} \cdots \sum_{x_{n-1}=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-2} x_{n-1}} \log q_{x_0} \\
&\vdots \\
&= \textstyle\sum_{x_0=1}^{k} q_{x_0} \log q_{x_0} = \sum_{i=1}^{k} q_i \log q_i
\end{aligned}$$

For the second term we split of each log terms up. We cancel sums from the right as far as possible using row-stochasticity and then cancel from the left using $\mathbf{q}P = \mathbf{q}$. For example

$$\begin{aligned}
&\textstyle\sum_{x_0=1}^{k} \cdots \sum_{x_n=1}^{k} q_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n} \log P_{x_1 x_2} \qquad \textit{row-stochastic} \\
&= \textstyle\sum_{x_0} \sum_{x_1} \sum_{x_2} q_{x_0} P_{x_0 x_1} P_{x_1 x_2} \log P_{x_1 x_2} \qquad \boldsymbol{q}P = \boldsymbol{q} \\
&= \textstyle\sum_{x_1} \sum_{x_2} q_{x_1} P_{x_1 x_2} \log P_{x_1 x_2} \\
&= \textstyle\sum_{i,j=1}^{k} q_i P_{ij} \log P_{ij}
\end{aligned}$$

We can do the same thing for each other log term to see

$$H(\alpha^{n+1}) = -\sum_{i=1}^{k} q_i \log(q_i) - n \sum_{i,j=1}^{k} q_i P_{ij} \log(P_{ij})$$

So when we divide by $n+1$ and take $n \to \infty$ we get

$$h(T) = h(T, \alpha) = \lim_{n\to\infty} \frac{H(\alpha^{n+1})}{n+1} = -\sum_{i,j=1}^{k} q_i P_{ij} \log P_{ij}$$

$\square$