Topics in Ergodic Theory

(4th October 2018, Thursday)

1. Measure preserving system

Measure preserving system : $(X, \mathcal{B}, \mu, T)$, where X is a set, $\mathcal{B}$ is a σ-algebra, μ is a probability measure with $\mu(A) \geq 0\ \forall A \in \mathcal{B}$, $\mu(X) = 1$, and T is a measure-preserving transformation. That is, $T : X \to X$ is measurable s.t. $\mu(T^{-1}(A)) = \mu(A)\ \forall A \in \mathcal{B}$.

If Y is a random element of X with distribution μ, then $T(Y)$ also has distribution μ.

Example)

- (Circle rotation) Let $X = \mathbb{R}/\mathbb{Z}$, $\mathcal{B}$ be the Borel sets, μ be the Lebesgue measure and $T = R_\alpha$ where $R_\alpha(x) = x + \alpha$, and $\alpha \in \mathbb{R}/\mathbb{Z}$ is parameter.
- (Times 2 map) $X = \mathbb{R}/\mathbb{Z}$, $\mathcal{B}$ be the Borel sets, μ is a Lebesgue measure, $T = T_2$ where $T_2(x) = 2x$.

  (proof that T_2 is measure preserving) First prove for intervals : let $I = (a, b)$. Then $\mu(I) = b-a$ and $\mu(T_2^{-1}I) = \mu((\frac{a}{2}, \frac{b}{2}) \cup (\frac{a+1}{2}, \frac{b+1}{2})) = b/2-a/2+b/2-a/2 = b-a$.(Just use Dynkin's lemma to conclude... Or,)
  Now let $U \subset \mathbb{R}/\mathbb{Z}$ be open. Then $U = \sqcup_i I_i$ is a disjoint union of intervals, so

  $$\mu(T^{-1}U) = \mu(\sqcup_j T^{-1}I_j) = \sum_j \mu(T^{-1}I_j) = \sum_j \mu(I_j) = \mu(U)$$

  Let $K \subset \mathbb{R}/\mathbb{Z}$ be a compact set. Then

  $$\mu(T^{-1}K) = 1 - \mu((T^{-1}K)^c) = 1 - \mu(T^{-1}(K^c)) = 1 - \mu(K^c) = \mu(K)$$

  Let A be an arbitrary Borel set and let $\epsilon > 0$. Then $\exists U$ open and $\exists K$ compact such that $K \subset A \subset A$ and $\mu(U \backslash K) < \epsilon$, so

  $$\mu(K) = \mu(T^{-1}K) \leq \mu(T^{-1}A) \leq \mu(T^{-1}U) = \mu(U)$$

  We also have $\mu(K) \leq \mu(A)\mu(U)$. Since $\mu(U) - \mu(K) < \epsilon$, $|\mu(A) - \mu(T^{-1}A)| < \epsilon$. Since ϵ was arbitrary, so $\mu(A) = \mu(T^{-1}A)$.

  (End of proof) $\square$

The **orbit** $x \in X$ is the sequence $x, Tx, T^2x, \cdots$.

Some Questions:

- Let $A \in \mathcal{B}$ and $x \in A$. Doest the orbit of x visit A infinitely often?
- What is the proportion of the times n such that $T^n x$ is in A?
- (Mixing property) What is $\mu(\{x \in A : T^n x \in A\})$ if n is large

Example) Let $A = [0, \frac{1}{4}) \subset \mathbb{R}/\mathbb{Z}$ and $T = T_2$. Then $T^n x \in A \Leftrightarrow (n+1)^{\text{st}}$ and $(n+2)^{\text{nd}}$ binary digits of x are 0.
For example, $x = 1/6 = 0.00101010\ldots_{(2)}$ never comes back to A.
Another interesting fact : $\mu(\{x : x \in A, T_2^n x \in A\}) = 1/16$ if $n \geq 2$.(Circle rotation has very different property.)

Markov Shift

- Let $(p_1, p_2, \cdots, p_n)^T$ be a probability vector. Let $A \in \mathbb{R}_{\geq 0}^{n \times n}$ be the **matrix of transition probabilities**.

  Assumptions : (1) $A(1, \cdots, 1)^T = (1, \cdots, 1)^T$; (2) $(p_1, \cdots, p_n)A = (p_1, \cdots, p_n)$
- Let $X = \{1, \cdots, n\}^{\mathbb{Z}}$, $\mathcal{B}$ be the Borel σ-algebra generated by the product topology of the discrete topology on $\{1, \cdots, n\}$, and $T = \sigma$ is the shift map $(\sigma X)_m = X_{m+1}$.
- Let $\mu(\{x \in X : x_m = i_0, \cdots, x_{m+n} = i_n\}) = p_{i_0} a_{i_0 i_1} \cdots a_{i_{n-1} i_n}$.

2. Furstenberg's correspondence principle

Theorem) (Szemerédi) Let $S \subset \mathbb{Z}$ of positive upper Banach density. That is:

$$\bar{d}(S) = \limsup_{N,M : M-N \to \infty} \frac{1}{M-N} |S \cap [N, M-1]| > 0.$$

Then S contains arbitrary long arithmetic progressions. That is, $\forall l$, $\exists a \in \mathbb{Z}$, $d \in \mathbb{Z}_{>0}$ such that $a, a+d, \cdots, a+(l-1)d \in S$.

Theorem) (Furstenberg) (Multiple recurrence) Let $(X, \mathcal{B}, \mu, T)$ be a MPS(Measure preserving system). Let $A \in \mathcal{B}$ be s.t. $\mu(A) > 0$. Let $l \in \mathbb{Z}_{>0}$. Then

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}(A) \cap \cdots \cap T^{-(l-1)n}(A)) > 0$$

(6th October 2018, Saturday)

Theorem) (Szemerédi) Let $S \subset \mathbb{Z}$ of positive upper Banach density. Then S contains arbitrary long arithmetic progressions.

Theorem) (Furstenberg) Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Let $A \in \mathcal{B}$ be s.t. $\mu(A) > 0$. Then for $\forall l \in \mathbb{Z}_{>0}$

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}(A) \cap \cdots \cap T^{-(l-1)n}(A)) > 0$$

Let $X = \{0,1\}^{\mathbb{Z}}$, $\mathcal{B}$ be the Borel σ-algebra, $T = \sigma$ be the shift map.
For a set $S \subset \mathbb{Z}_{\geq 0}$, Let $x^S \in X$ be defined by

$$x_n^S = \begin{cases} 1 & \text{if} \quad n \in S \\ 0 & \text{if} \quad \text{o/w} \end{cases}$$

Let $A \in \mathcal{B}$ and $A = \{x \in X : x_0 = 1\}$
Observation : $n \in S \Leftrightarrow \sigma^n x^S \in A \Leftrightarrow (\sigma^n x^S)_0 = 1 \Leftrightarrow x_n^S = 1$.
Let $\{M_m\}$ and $\{N_m\}$ be sequences s.t.

$$\bar{d}(S) = \lim_{m\to\infty} \frac{1}{M_m - N_m} \big| S \cap [N_m, M_m - 1] \big|$$

Let $\mu_m = \frac{1}{M_m - N_m} \sum_{n=N_m}^{M_m - 1} \delta_{\sigma^n x^S}$, where δ_x is a measure on X defined as

$$\delta_x(B) = \begin{cases} 1 & x \in B \\ 0 & \text{o/w} \end{cases}$$

Let μ be the weak limit of a subsequence of μ_m.

(Reminder)

- **Weak Limits) :** (In fact, weak-* limits) Let X be a compact metric space, Let μ_m be a sequence of Borel measures on X, and let μ be another Borel measure. Then μ_m weakly converges to μ. In notation,

  $$\lim -w_{m\to\infty} \mu_m = \mu$$

  if $\int f d\mu_m \to \int f d\mu \ \forall f \in C(X)$

- **Theorem)** (Banach-Alaoglu/Helly) Let X be a compact metric space. Then $\mathcal{M}(X)$, the set of Borel probability measures endowed with the topology of weak convergence, is compact and metrizable.

  In particular, there is a weakly convergent subsequence in any sequence of Borel probability measures.

Lemma) Let $(X, \mathcal{B}, \mu, \sigma)$ be as defined above is a measure preserving system.

proof sketch) Let $B \in \mathcal{B}$ Then:

$$\begin{aligned} \mu_m(B) &= \frac{1}{M_m - N_m} |\{n \in [N_m, M_m - 1] : \sigma^n x^S \in B\}| \\ \mu_m(\sigma^{-1}B) &= \frac{1}{M_m - N_m} |\{n \in [N_m, M_m - 1] : \sigma^n x^S \in \sigma^{-1}B\}| \\ &= \frac{1}{M_m - N_m} |\{n \in [N_m + 1, M_m] : \sigma^n x^S \in B\}| \\ |\mu_m(B) - \mu_m(\sigma^{-1}B)| &\leq \frac{1}{M_m - N_m} \to 0 \quad \text{as } m \to \infty \end{aligned}$$

It can be shown that we can pass to the limit and conclude $\mu(B) = \mu(\sigma^{-1}B)$.

•**Remark :** If B is a cylinder set, i.e. $\exists L \in \mathbb{Z}_{>0}$ and $\tilde{B} \subset \{0,1\}^{2L+1}$ s.t.

$$B = \{x \in X : (x_{-L}, \cdots, x_L) \in \tilde{B}\}$$

then B is both closed and open. Therefore χ_B, the characteristic function of B, is continuous. Hence, the limit

$$\lim_{m\to\infty} \mu_m(B) = \mu(B)$$

Proposition) Let $S \subset \mathbb{Z}$, let x^S, A, $(X, \mathcal{B}, \mu, \sigma)$ be as defined above. Let $l \in \mathbb{Z}_{>0}$. Suppose that $\exists n \in \mathbb{Z}_{>0}$ s.t.

$$\mu(A \cap \sigma^{-n}(A) \cap \cdots \cap \sigma^{-n(l-1)}(A)) > 0$$

Then S contains an arithmetic progression of length l.

> **proof)** Without loss of generality, we may assume that $\mu = \lim_{m\to\infty} \mu_m$(if this is not the case, we just replace μ_m with its converging subsequence). Let $B = A \cap \sigma^{-n}(A) \cap \cdots \cap \sigma^{-n(l-1)}(A)$ and observe that B is a cylinder set. Then $\mu(B) = \lim \mu_m(B)$ hence $\exists m$ s.t. $\mu_m(B) > 0$.
>
> By definition of μ_m, $\exists k \in [N_m, M_m - 1]$ such that $\sigma^k x^S \in B$. Hence
>
> $$\begin{aligned} &\sigma^k x^S \in A,\ \sigma^k x^S \in \sigma^{-n}(A), \cdots, \sigma^k x^S \in \sigma^{-n(l-1)}(A) \\ \Rightarrow &\sigma^k x^S \in A,\ \sigma^{k+n} x^S \in A, \cdots, \sigma^{k+n(l-1)} x^S \in A \end{aligned}$$
>
> and so $k, k+n, \cdots, k+n(l-1) \in S$ by earlier observation.
>
> *(End of proof)* $\square$

Note A is also a cylinder set. Then $\mu(A) = \lim_m \mu_m(A)$ and

$$\mu(A) = \lim_m \mu_m(A) = \lim_m \frac{1}{M_m - N_m} \big|\{n \in [N_m, M_m - 1] : n \in S\}\big| = \bar{d}(S) > 0$$

by assumption that S is of positive upper Banach density, and therefore we can prove Szemerédi when assuming Furstenberg.

(9th October, Tuesday)

3. Poincaré recurrence, Ergodicity

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be MPS. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then $\exists n \in \mathbb{Z}_{>0}$ s.t. $\mu(A \cap T^{-n}A) > 0$.

> **proof)** Suppose $\mu(A \cap T^{-n}A) = 0$ for all $n > 0$. Then
>
> $$\mu(T^{-k}A \cap T^{-n}A) = \mu(A \cap T^{-(n-l)}A) = 0$$

for all $n > k \geq 0$. Hence the sets $A, T^{-1}A, \cdots$ are "almost pairwise disjoint". Then

$$\begin{aligned}\mu(A \cup T^{-1}A \cup \cdots \cup T^{-n}A) &= \mu(A) + \big(\mu(T^{-1}A) - \mu(T^{-1}A \cap A)\big) \\ &+ \big(\mu(T^{-2}A) - \mu(T^{-2}A \cap (A \cup T^{-1}A)) + \cdots \\ &+ \big(\mu(T^{-n}A) - \mu(T^{-n}A \cap (A \cup T^{-1}A \cup \cdots \cup T^{-(n-1)}A))\big) \\ &= (n+1)\mu(A),\end{aligned}$$

a contradiction if $n + 1 > \mu(A)^{-1}$.

(End of proof) $\square$

Theorem) (Poincaré recurrence) Let $(X, \mathcal{B}, \mu, T)$ be MPS. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then a.e. $x \in A$ returns to A infinitely often. That is,

$$\mu(A \backslash \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A) = 0$$

Remark : $x \in T^{-n}A \Leftrightarrow T^n x \in A$. So $\cup_{n=N}^{\infty} T^{-n}A$ are the points that visit A at least once after time N.

proof) Let A_0 be the set of point in A that never returns to A. We first show $\mu(A_0) = 0$. Note that $\mu(A_0 \cap T^{-n}A_0) \leq \mu(A_0 \cap T^{-n}A) = \mu(\phi) = 0$ for all $n > 0$. By the previous lemma, we have $\mu(A_0) = 0$. Note that if $x \in A \backslash \big(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A\big)$, then there is a maximal $m \in \mathbb{Z}_{\geq 0}$ such that $T^m x \in A_0$. This means that

$$A \backslash \cap \cup T^{-n}A \subset \cup_{m=0}^{\infty} T^{-m}A_0$$

and since $T^{-m}A_0$ has measure 0 for each $m \geq 0$, $A \backslash \cap \cup T^{-n}A$ also has measure 0.

(End of proof) $\square$

However, if we are aim to show that any point of X(or almost every) visits a set A with $\mu(A)$ infinitely often, we should prevent elements of X being partitioned by orbits of T^{-1}. Assumption of ergodicity turns out to be enough for this.(In fact, we can make 'ergodic decomposition' for T to to satisfy ergodicity on each partition - but not lecturing on this; bit tricky)

Definition) A MPS $(X, \mathcal{B}, \mu, T)$ is called **ergodic** if $A = T^{-1}A$ implies $\mu(A) = 0$ or 1 for all $A \in \mathcal{B}$.

If the MPS is not ergodic, and $A \in \mathcal{B}$ with $\mu(A) \in (0, 1)$ s.t. $T^{-1}A = A$, then we can restrict the MPS to A. That is, we consider the MPS:
$(A, \mathcal{B}_A, \mu_A, T\big|_A)$ where $\mathcal{B}_A = \{B \in \mathcal{B} : B \subset A\}$, $\mu_A(B) = \mu(B)/\mu(A)$ for all $B \in \mathcal{B}_A$.

Theorem) The following are equivalent for a MPS $(X, \mathcal{B}, \mu, T)$:

(1) $(X, \mathcal{B}, \mu, T)$ is ergodic.

(2) $\mu(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A) = 1$ for all $A \in \mathcal{B}$ with $\mu(A) > 0$.

(3) $\mu(A \triangle T^{-1}A) = 0$ implies $\mu(A) = 0$ or 1 for all $A \in \mathcal{B}$.

(4) For all bounded measurable functions $f : X \to \mathbb{R}$, $f = f \circ T$ a.e. implies f is constant a.e.

(5) For all measurable functions $f : X \to \mathbb{C}$, $f = f \circ T$ a.e. implies f is constant a.e.

Each condition show different perspective to view ergodicity. The second item shows that for ergodic systems Poincaré recurrence holds in a stronger form: not only almost every point in A but also almost every point in X visits A infinitely often. The last three conditions are often used in practice.

proof)

(1)⇒(2) Let $A \in \mathcal{B}$ with $\mu(A) > 0$ Let $B = \bigcap \bigcup T^{-n}A$, the set of points that visit A infinitely often. By Poicaré recurrence(or P-recurrence), $\mu(B) \geq \mu(A) > 0$. So if we show that $B = T^{-1}B$, then $\mu(B) = 1$ follows by ergodicity.

While, $x \in B \Leftrightarrow x$ visits A i.o. $\Leftrightarrow Tx$ visits A i.o. $\Leftrightarrow Tx \in B$. So we proved $B = T^{-1}B$.

(2)⇒(3) Let $A \in \mathcal{B}$ s.t. $\mu(A \triangle T^{-1}A) = 0$. If $\mu(A) = 0$, there is nothing to prove, so assume $\mu(A) > 0$. Let $B = \cap \cup T^{-n}A$. By (2), we know that $\mu(B) = 1$. We show $\mu(B \backslash A) = 0$, which completes the proof.

Let $x \in B \backslash A$, then there is a first time $m > 0$ s.t. $T^m x \in A$, hence $x \in T^{-m}A \backslash T^{-(m-1)}A$. This shows $B \backslash A \subset \bigcup T^{-m}A \backslash T^{-(m-1)A}$. But $T^{-m}A \backslash T^{-(m-1)A}$ has measure 0 because $\mu(T^{-m}A \backslash T^{-(m-1)}A) = \mu(T^{-1}A \backslash A) = 0$.

So we conclude $\mu(B \backslash A) = 0$.

(3)⇒(4) Let $f : X \to \mathbb{R}$ be a bounded measurable function s.t. $f = f \circ T$ almost everywhere. For any $t \in \mathbb{R}$, define $A_t = \{x \in A : f(x) \leq t\}$. Then

$$\mu(A_t \triangle T^{-1}A_t) = \mu(\{x \in A : f(x) \leq t\} \triangle \{x \in A : f \circ T(x) \leq t\}) = 0$$

By (3), we have $\mu(A_t) \in \{0, 1\}$ for all t. Since f was bounded, if t is very small, then $\mu(A_t) = 0$ and if t is very large $\mu(A_t) = 1$. But $t \mapsto \mu(A_t)$ is a monotone function, we have $\exists c \in \mathbb{R}$ s.t. $\mu(A_t) = 0$ for all $t < c$ and $\mu(A_t) = 1$ for all $t > c$. Therefore we have $f(x) = c$.

(4)⇒(1) Let $A \in \mathcal{B}$ with $A = T^{-1}A$. Then $\chi_A = \chi_A \circ T$ everywhere, so χ_A is constant a.e.

Example : The circle rotation $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, R_\alpha)$ is ergodic if and only if α is irrational.

proof) Let $f : X \to \mathbb{R}$ be measurable, and let $f(x) = \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n x)$. Then

$$\begin{aligned} f \circ R_\alpha(x) = f(x + \alpha) &= \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n(x + \alpha)) \\ &= \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n \alpha) \exp(2\pi i n x) \end{aligned}$$

so $f = f \circ R_\alpha$ is equivalent to having $a_n = a_n \exp(2\pi i n \alpha)$ for all n. If α is irrational, then $\exp(2\pi i n \alpha) \neq 1$ for all $n \neq 0$ so $a_n = 0$ for all $n \neq 0$.

(End of proof) □

(11th October, Thursday)

4. Ergodic theorems

Theorem) (Mean ergodic theorem, von Neumann) Let $(x, \mathcal{B}, \mu, T)$ be a MPS. Write

$$I = \{f \in L^2(X) : f \circ T = f \text{ a.e.}\} \subset L^2(X)$$

for the closed subspace of T-invariant functions. Write $P_T : L^2(X) \to I$ for the orthogonal projection. Then for every $f \in L^2(X)$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \to P_T f \quad \text{in } L^2(X)$$

Here, $\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n$ called the **ergodic average**.

There are two proofs for this theorem : one uses spectral theory and the the other does not. We would prove using the second approach, and sketch the first proof in the example sheet.

Theorem) (Pointwise ergodic theorem, Birkhoff) Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Then for all $f \in L^1(X)$, $\exists f^* \in L^1(X)$ s.t. $f^* = f^* \circ T$ a.e. and

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n(x) \to f^*(x) \quad \text{a.e. in } X$$

Comments

(1) If $f \in L^2 \cap L^1$, then $f^* = P_T f$.

(2) There is an L^p version of convergence in norm. That is, if $f \in L^p$, then

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \to f^* \quad \text{in } L^p \text{ norm}$$

This will be proved in the example sheet.

(3) If $(X, \mathcal{B}, \mu, T)$ is ergodic, then f^*(or $P_T f$) is constant a.e., because it is T-invariant.
Note : $f^*(x) = \int f^* d\mu$ a.e. By L^1 norm convergence, we also have

$$\lim_{N\to\infty} \int \frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n d\mu \to \int f^* d\mu$$

By a lemma that would follow,

$$\int f \circ T^n d\mu = \int f d\mu \quad \forall n, \text{ hence } \quad \int f d\mu = \int f^* d\mu$$

Then

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int f d\mu$$

Can be interpreted as ”time average(LHS) converges to spatial average(RHS)”.

Lemma) Let $T : X \to X$ be a measurable transformation and let μ be a probability measure. Then μ is T-invariant *if and only if*

$$\int f \circ T d\mu = \int f d\mu \quad \forall f \in L^1(X, \mu) \tag{1}$$

proof) (*(1) $\Rightarrow$ measure preserving property*) : Let $A \in \mathcal{B}$. Then

$$\mu(T^{-1}A) = \int \chi_{T^{-1}A} d\mu = \int \chi_A \circ T d\mu = \int \chi_A d\mu = \mu(A)$$

(*MPP $\Rightarrow$ (1)*) : Let $f \in L^1(X)$. If $f = \chi_A$ for some $A \in \mathcal{B}$, then

$$\int f \circ T d\mu = \mu(T^{-1}A) = \mu(A) = \int f d\mu$$

(1) hold for such f. Then (1) also holds for simple functions by linearity of integration. In the case where f is non-negative, let f_n be a monotone increasing sequence of simple functions such that $\lim_n f_n = f$(e.g. $f_n = f \wedge n$),

$$\int f \circ T d\mu = \lim_n \int f_n \circ T d\mu = \lim_n \int f_n d\mu = \int f d\mu$$

In the general case, separate f into positive and negative parts and conclude the proof.

(End of proof) □

Definition) Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Then the **Koopman operator** is defined as : $U_T f = f \circ T$ acting on functions on X.

Lemma) The Koopman operator is an isometry on $L^2(X)$. That is,

$$\langle f, g \rangle = \langle U_T f, U_T g \rangle$$

proof) Apply the previous lemma for the function $f \circ \bar{g}$.

$$\mu(U_T f \cdot U_T \bar{g}) = \mu(U_T(f\bar{g})) = \mu(f\bar{g})$$

(End of proof) □

Definition) A MPS $(X, \mathcal{B}, \mu, T)$ is called **invertible** if $\exists S : X \to X$, measure preserving, s.t.

$$S \circ T = T \circ S = id_X \quad \text{a.e.}$$

If such a map exists, we denote it by $T^{-1} = S$. (such operator is unique up to a.s. equality)

Lemma) If $(X, \mathcal{B}, \mu, T)$ is invertible, then U_T is unitary, and $U_T^* = U_{T^{-1}}$.

proof) Note : $U_{T^{-1}} \circ U_T = U_T \circ U_{T^{-1}} = id_{L^2(X)}$, so it is enough to show that $U_T^* = U_{T^{-1}}$. To do this, we need to show :

$$\langle U_{T^{-1}} f, g \rangle = \langle f, U_T g \rangle \quad \forall f, g, \in L^2$$

and

$$\langle U_{T^{-1}} f, g \rangle = \int f \circ T^{-1} \cdot \bar{g} d\mu = \int \left(f \circ T^{-1} \cdot \bar{g} \right) \circ T d\mu = \int f \cdot \bar{g} \circ T d\mu = \langle f, U_T g \rangle$$

(End of proof) □

Both von Neumann's and Birkhoff's theorems are easy for certain special kinds of functions. For instance, if $f \in I$:

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n = f$$

Also, if $f = g \circ T - g$ for some g, then

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n = \frac{1}{N}\big(g \circ T^N - g\big)$$

It turns out that these two are the only functions that we have to worry about (in the case of von Neumann's theorem) - as presented in the following lemma.

Lemma) Write $B = \{g \circ T - g : g \in L^2(x)\}$. Then $B^\perp = I$.

Caution ! : B is not close in L^2. So we get $L^2 = I \oplus \bar{B}$, but not $L^2 = I \oplus B$.

proof) Let $f \in L^2(X)$. Then

$$\begin{aligned} & f \in B^\perp && \Leftrightarrow \quad \langle f, g \circ T - g\rangle = 0 \quad \forall g \in L^2 \\ \Leftrightarrow \quad & \langle f, g \circ T\rangle = \langle f, g\rangle \quad \forall g \in L^2 \\ \Leftrightarrow \quad & \langle U_T^* f, g\rangle = \langle f, g\rangle \quad \forall g \in L^2 \\ \Leftrightarrow \quad & U_T^* f = f \end{aligned}$$

Now we only need to see that $U_T^* f = f \Leftrightarrow U_T f = f$:

$$\begin{aligned} & U_T f = f \\ \Leftrightarrow \quad & \left\|f - U_T f\right\|^2 = 0 \\ \Leftrightarrow \quad & \left\|f\right\|^2 + \left\|U_T f\right\|^2 - \langle f, U_T f\rangle - \langle U_T f, f\rangle = 0 \\ \Leftrightarrow \quad & \left\|f\right\|^2 + \left\|U_T^* f\right\|^2 - \langle f, U_T^* f\rangle - \langle U_T^* f, f\rangle + \left(\left\|U_T f\right\|_{-}\left\|U_T^* f\right\|^2\right) = 0 \\ \Leftrightarrow \quad & \left\|f - U_T^* f\right\|^2 + \left(\left\|U_T f\right\|^2 - \left\|U_T^* f\right\|^2\right) = 0 \end{aligned}$$

Since $\left\|f - U_T^* f\right\|^2 \geq 0$, $\left\|U_T f\right\|^2 - \left\|U_T^* f\right\|^2 \geq 0$(note that we do not know that U_T^* is unitary, since we do not know if T is invertible, but we know that $\left\|U_T^*\right\|_{op} \leq 1$), this statement is equivalent to having $f = U_T^* f$.

Now we are ready to prove the mean ergodic theorem.

proof of MET) Fix $\epsilon > 0$. Let $f \in L^2$. By the lemma, $\exists g, e \in L^2$ s.t.

$$f = P_T f + (g \circ T - g) + e$$

with $\left\|e\right\| < \epsilon$ and

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n = P_T f + \frac{1}{N}(g \circ T^N - g) + \frac{1}{N}\sum_{n=0}^{N-1} e \circ T^n$$

This gives bound

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n - P_T f\Big\| \leq \frac{2\|g\|}{N} + \epsilon$$

Taking $N\to\infty$ gives

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n - P_T f\Big\|_{\leq}\epsilon$$

(End of proof) □

(13th October, Saturday)

Now we start proving mean ergodic theorem, starting with the following theorem.

Theorem) *(Maximal Ergodic Theorem, Wiener)* Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Let $f \in L^1$, $\alpha \in \mathbb{R}_{>0}$. Let

$$E_\alpha = \{x\in X : \sup_N \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) > \alpha\}$$

Then $\mu(E_\alpha) \leq \frac{1}{\alpha}\|f\|_1$.

-the theorem is useful, because we can bound some set of particular irregularity depending on the parameter α.
-usually, these kinds of maximal inequalities are prove using covering lemmas, e.g. using Vitalli covering lemma. This is also possible in this case, but the proof gets too long.

The proof of the theorem depends on the following proposition

Proposition) Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Let $f \in L^1$. Let

$$\begin{aligned}
&f_0 = 0, \quad f_1 = f, \quad f_2 = f\circ T + f, \quad \cdots \\
&f_n = f\circ T^{n-1} + \cdots + f\circ T + f = f_{n-1}\circ T + f \\
&F_N = \max_{n=0,\cdots,N} f_n
\end{aligned}$$

Then $\int_{x:F_N(x)>0} f(x)d\mu(x) \geq 0$ for all N.

proof) Suppose that $F_N(x) > 0$. Then $F_N(x) = f_n(x)$ for some $n \in \{1, \cdots N\}$. Then $F_N(x) = f_{n-1}(Tx) + f(x) \leq F_N(Tx) + f(x)$, hence $f(x) \leq F_N(x) - F_N(Tx)$.

$$\int_{\{x:F_N(x)>0\}} f(x)d\mu \geq \int_{\{x:F_N(x)>0\}} (F_N(x) - F_N\circ T(x))d\mu(x)$$

Note, if $F_N(x) \leq 0$, then we have $F_N(x) = 0$ and $F_N(x) - F_N(Tx) \leq 0$, so we have $F_N(x) - F_N\circ T(x) \geq 0$ on the domain $\{x : F_N(x) > 0\}$.

(End of proof) □

We now prove maximal ergodic theorem.

proof of Maximal E.T.) Define

$$E_{\alpha,N} = \{x \in X : \max_{m=0\cdots,N} \frac{1}{m}\sum_{n=0}^{m-1} f(T^n x) > \alpha\}$$
$$= \{x \in X : \max_{m=0,\cdots,N} \sum_{n=0}^{m-1} \big(f(T^n x) - \alpha\big) > 0\}$$

(with convention that the sum is just 0 in the case $m = 0$) We apply the proposition for the function $f - \alpha$. Then

$$\int_{E_{\alpha,N}} \big(f(x) - \alpha\big) d\mu \geq 0$$

Then

$$\|f\|_1 \geq \int_{E_{\alpha,N}} f(x) \geq \alpha\mu(E_{\alpha,N})$$

Note that $E_\alpha = \bigcup_M E_{\alpha,M}$ is an increasing union and the inequality holds for any N, so $\|f\|_1 \geq \alpha\mu(E_\alpha)$.

(End of proof) □

Note that, in fact the proof in showing a somewhat stronger version of maximal ergodic theorem. Namely,

Theorem) *(Maximal Ergodic Theorem, version 2)* Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Let $f \in L^1$, $\alpha \in \mathbb{R}_{>0}$. Let

$$E_\alpha = \{x \in X : \sup_N \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) > \alpha\}$$

Then $\mu(E_\alpha) \leq \frac{1}{\alpha}\mu(f 1_{E_\alpha})$.

proof) It follows from the fact $\int_{E_{\alpha,N}} f(x) \geq \alpha\mu(E_{\alpha,N})$ for all $N \geq 0$.

(End of proof) □

Theorem) (Pointwise ergodic theorem) Let $(X, \mathcal{B}, \mu, T)$ be a MPS. Let $f \in L^1$. Then $\exists f^* \in L^1$, T-invariant s.t.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to f^*(x) \quad \text{pointwise a.e.}$$

proof) Fix $\epsilon > 0$. Then $\exists f_\epsilon \in L^2$, $e_{\epsilon,1} \in L^1$ s.t.

$$f = f_\epsilon + e_{\epsilon,1} \quad \text{and } \|e_{\epsilon,1}\|_1 < \epsilon.$$

Also $\exists g_\epsilon \in L^2$, $e_{\epsilon,2} \in L^2$ s.t.

$$f_\epsilon = P_T f_\epsilon + g_\epsilon \circ T - g_\epsilon + e_{\epsilon,2} \quad \text{and } \|e_{\epsilon,2}\|_1 < \epsilon$$

and $\exists h_\epsilon \in L^\infty$, $e_{\epsilon,3} \in L^1$ s.t.

$$g_\epsilon = h_\epsilon + e_{\epsilon,3} \quad \text{and } \left\|e_{\epsilon,3}\right\|_1 < \epsilon$$

So $f = P_T f_\epsilon + h_\epsilon \circ T - h_\epsilon + e_\epsilon$, where $e_\epsilon \in L^1$ with$\left\|e_\epsilon\right\|_1 < \epsilon$.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) = P_T f_\epsilon(x) + \frac{1}{N}\big(h_\epsilon(T^n x) - h_\epsilon(x)\big) + \frac{1}{N}\sum_{n=0}^{N_1} e_\epsilon(T^n x)$$

Let

$$E_{\epsilon,\alpha} = \{x \in : \limsup_{N\to\infty} \Big|\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) - P_T f_\epsilon(x) > \alpha\Big|\}$$

(Not same as $E_{\alpha,N}$ defined earlier) Applying the Maximal ergodic theorem for the f_n gives

$$\mu(E_{\epsilon,\alpha}) \leq \frac{1}{\alpha}\left\|e_\epsilon\right\|_1 \leq \frac{h\epsilon}{\alpha}$$

Let F be the set of points x s.t. $\frac{1}{N}\sum_{n=0}^{N_1} f(T^n x)$ does not converge at x. Then $F \subset \bigcup_\alpha F_\alpha$, where

$$F_\alpha = \{x \in X : \limsup_{N_1,N_2\to\infty} \Big|\frac{1}{N}\sum_{n=0}^{N_1-1} f(T^n x) - \frac{1}{N_2}\sum_{n=0}^{N_2-1} f(T^n x)\Big| > 2\alpha\}$$

Notice, $F_\alpha \subset E_{\epsilon,\alpha}$ for all $\epsilon > 0$(?????), so $\mu(F_\alpha) \leq \mu(E_{\epsilon,\alpha}) \leq \frac{h\epsilon}{\alpha}$. Therefore $\mu(F_\alpha) = 0$. We can take a countable sequence of α's(e.g. $(1/k)_{k\in\mathbb{N}}$) and conclude $\mu(F) = 0$.

We proved that $\frac{1}{N}\sum_{n=0} f(T^n x) \to f^*(x)$ for some function f^*. By Fatou's lemma, we have $f^* \in L^1$, and it remains to prove $f^*(x) = f^*(Tx)$ a.e.

For almost every x,

$$\begin{aligned}
f^*(x) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \\
f^*(Tx) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T^{n-1} x) \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \\
&= \lim_{N\to\infty}\frac{1}{N-1}\sum_{n=1}^{N-1} f(T^n x) \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N-1} f(T^n x)
\end{aligned}$$

Therefore $f^*(x) - f^*(Tx) = \lim_{N\to\infty}\frac{1}{N} f(x) = 0$

(End of proof) $\square$

a more elegant proof) For simplicity, let $S_N(x)\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n(x)$.

(1) First, let f be a positive L^1 function. Then we may find a positive measurable function $\bar{f}(x)$ such that

$$\bar{f}(x)=\liminf_{N\to\infty} S_N(x) \quad \text{a.e.}$$

Note that $\bar{f}$ is T-invariant, since

$$S_N\circ T=\frac{1}{N}(f\circ T+\cdots+f\circ T^N)=\frac{N+1}{N}S_{N+1}-\frac{1}{N}f$$

Also, by Fatou's lemma,

$$\mu(|\bar{f}|)\leq \liminf_N \mu(|\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n|)=\left\|f\right\|_1<\infty$$

and therefore $\bar{f}\in L^1$ with $\int \bar{f}d\mu=\int fd\mu$. Now let $g(x)=f(x)-\bar{f}(x)$, then again $g\in L^1$, with $\int gd\mu=0$ and

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} g\circ T^n(x)=0 \quad \text{a.e.}$$

Now consider the set F_q, defined for $q\in\mathbb{Q}_{>0}$.

$$F_q=\{x:\limsup_N \frac{1}{N}\sum_{n=0}^{N-1} g\circ T^n(x)>q\}$$

Observe that F_q is a T-invariant set, since

$$\begin{aligned}
Tx\in F_q \quad &\Leftrightarrow \quad \limsup_N \frac{1}{N}\sum_{n=0}^{N-1} g\circ T^{n+1}(x)>q \\
&\Leftrightarrow \quad \limsup_N \Big(\frac{1}{N}\sum_{n=0}^{N} g\circ T^n(x)-\frac{1}{N}g\Big)>q \quad \Leftrightarrow \quad x\in F_q
\end{aligned}$$

So we may repeat our arguments above to the restricted MPS $(F_q,\mathcal{B}\big|_{F_q},\mu\big|_{F_q},T\big|_{F_q})$ and hence show that $\mu(g1_{F_q})=\mu(f1_{F_q})-\mu(\bar{f}1_{F_q})=0$. But by the maximal ergodic theorem(version 2), we have

$$q\mu(F_q)\leq \int_{F_q} gd\mu=0$$

hence $\mu(F_q)=0$, and

$$\mu(\{x:|\limsup_N \frac{1}{N}\sum_{n=0}^{N-1} g\circ T^n(x)|>0\})=\mu(\cap_{q\in\mathbb{Q}_{>0}}F_q)=0$$

We may conclude that

$$\limsup_N \frac{1}{N}\sum_{n=0}^{N-1} g\circ T^n(x)=\liminf_N \frac{1}{N}\sum_{n=0}^{N-1} g\circ T^n(x)=0 \quad \text{a.e.}$$

and we see that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x) \to \bar{f} \quad \text{a.e.}$$

with $\|\bar{f}\|_1 \le \|f\|_1$.

(2) For the general case, just divide f into a non-negative part and a negative part, e.g. $f = f^+ - f^-$ and apply part (1) to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f^+ \circ T^n(x) \to \bar{f}^+ \quad \text{a.e.}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f^- \circ T^n(x) \to \bar{f}^- \quad \text{a.e.}$$

and put $f^* = \bar{f}^+ - \bar{f}^-$, then we have the desired result.

(End of proof) □

(not done in the lecture. a question in example sheet.)

Theorem) (Pointwise ergodic theorem, L^p-version) Let $(X, \mathcal{B}, \mu, T)$ be a MPS, that is σ-finite. Let $f \in L^p$. Then $\exists f^* \in L^p$, T-invariant s.t.

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \to f^*(x) \quad \text{pointwise a.e.}$$

proof) First, assume that f is a positive function. Let $(f_n)_n$ be a increasing sequence of L^1 functions s.t. $f_n \to f$ in L^p and almost everywhere. Then by pointwise ergodic theorem for L^1 functions, we may find $(f_n^*)_n \subset L^1$ s.t.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_n \circ T^n(x) \to f_n^* \quad \text{a.e.}$$

Then $(f_n^*)_n$ also forms an increasing sequence, and therefore converges almost everywhere, say $f_n^* \to f^*$ a.e. Now, by Fatou's lemma, we have, for $n \ge m$,

$$\mu((f_n^* - f_m^*)^p)^{1/p} \le \liminf_n \left(\mu(\frac{1}{N} \sum_{k=0}^{N-1} (f_n - f_m) \circ T^k) \right)^{1/p} \le \liminf_n \mu((f_n - f_m)^p)^{1/p}$$

where the last inequality follows from Minkowski's inequality. Therefore, $(f_n^*)_n$ forms a Cauchy sequence in L^p, and in fact converges to f^* in L^p. Also, again by Minkowski's inequality, we have $\|f_n^*\|_p \le \|f_n\|_p$ and

$$\left\|\frac{1}{N} \sum_{m=0}^{N-1} f_n \circ T^m\right\|_p \le \|f_n\|_p$$

so by dominated convergence theorem, we realize that $\frac{1}{N}\sum_{n=0}^{N-1} f_n \to f^*$ is in fact in L^p. Putting these results together, we conclude that

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \to f^* \quad \text{in } L^p \text{ and a.e.}$$

$$\begin{array}{ccc} f_n & \xrightarrow{L^p} & f \\ \Big\downarrow{\scriptstyle L^p} & & \Big\downarrow \\ f_n^* & \xrightarrow{L^p} & f^* \end{array}$$

For general functions(not necessarily positive), divide it into a non-negative part and a negative part, and find a.e. and L^p converging functions separately and add them.

(16th October, Tuesday)

Definition) A number $x \in [0,1)$ is called **normal** in base K, if for every $b_1, b_2, \cdots, b_M \in \{0, \cdots, K\}$, we have :

$$\frac{1}{N}\left|\{n \in \{0, \cdots, N-1\} : x_{n+1} = b+1, \cdots, x_{n+M} = b_M\}\right| \to \frac{1}{K^M}$$

where $x = 0.x_1 x_2 \cdots_{(K)}$ is a base K expansion.

Theorem) Almost every number (w.r.t. Lebesgue measure) is normal in any base $K \geq 2$.

proof) Consider the MPS $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, m, T_K)$($\mathcal{B}$ the Borel σ-algebra, m the Lebesgue measure) where $T_K(x) = K \cdot x$. From the example sheet, this is an ergodic MPS. Now fix M and $b_1, \cdots, b_M$ as in the definition and consider the set

$$A = \left[(0.b_1 b_2 \cdots b_M)_{(K)}, (0.b_1 \cdots b_M)_{(K)} + \frac{1}{K^M}\right)$$

$\bullet$Note : $T^n x \in A \quad \Leftrightarrow \quad x_{n+1} = b_1, c \ldots, x_{n+M} = b_M$

To see that x is normal, we need

$$\frac{1}{N}\sum_{n=0}^{N-1} \chi_A(T^n x) \to \frac{1}{K^M}$$

This holds by the pointwise ergodic theorem for almost every x. Since there are countably many choices for K, M, and $b, \cdots, b_M$, the theorem follows.

(End of proof) $\square$

6. Unique ergodicity

The problem with these ergodic theorems is that the converging point might differ depending on the selection of measure. To study in which cases this can be prevented, we study the uniqueness of measures that is preserved under a fixed map T.

Definition) A **topological dynamical system** is a tuple (X, T), where X is a compact metric space and $T : X \to X$ is a continuous map. We say that (X, T) is **uniquely ergodic**, if there is only one *T-invariant Borel probability measure* on X.

Theorem) Let (X, T) be a topological dynamical system. The followings are equivalent :

(1) (X, T) is uniquely ergodic.

(2) For every $f \in C(X)$, there is $c_f \in \mathbb{C}$ s.t.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to c_f \quad \text{uniformly on } X$$

(3) There is a dense $A \subset C(X)$ and for each $f \in A$ there is $c_f \in \mathbb{C}$ s.t.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to c_f \quad \text{not necessarily uniformly } \forall x \in X$$

Theorem) *(Riesz representation theorem)* Let X be a compact metric space. Then to each finite Borel measure on X, we associate bounded linear functional on $C(X)$ as follows :

$$L_\mu f = \int f d\mu$$

Then $\mu \mapsto L_\mu$ is a bijection from the space of finite Borel measures on X, $\mathcal{M}(X)$, to bounded linear functional on $C(X)$.

Corollary) Let $\mu_1 \neq \mu_2$ be two Borel measures on a compact metric space. Then $\mu_1 = \mu_2$ if and only if

$$\int f d\mu_1 = \int f d\mu_2 \quad \forall f \in C(X)$$

Definition) Let X, T be as above, let μ be a Borel measure. The push-forward of μ via T is the measure

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

-This indeed defines a measure.

Lemma) Let X, T, μ be as above. Then

$$\int f dT_*\mu = \int f \circ T d\mu$$

for every bounded measurable function f.

proof) First prove this for characteristic functions of sets. Let $A \in \mathcal{B}$.

$$\int \chi_A dT_*\mu = T_*\mu(A) = \mu(T^{-1}A) = \int \chi_{T^{-1}A} d\mu = \int \chi_A \circ T d\mu$$

Now use uniform class theorem to complete the proof.

(End of proof) $\square$

•**Remark :** μ is T-invariant *iff* $\mu = T_*\mu$.

Lemma) Let X, T, μ be as above. Then μ is T-invariant *iff*

$$\int f d\mu = \int f \circ T d\mu \quad \forall f \in C(X) \qquad \cdots\cdots (\star)$$

(we are talking about continuous functions in place of measurable functions - so is in fact enough to work with only continuous functions.)

proof) We have already seen that μ being T-invariant implies $(\star)$.

For the other direction, note the following : suppose that $(\star)$ holds. Then $\int f dT_* d\mu = \int f \circ T d\mu = \int f d\mu$ for all $f \in C(X)$. Now by the corollary before, we have $\mu = T_*\mu$.

(End of proof) $\square$

Theorem) Let (X, T) be a topological dynamical system. Let $(\nu_j)_j$ be a sequence of Borel probability measures on X. Let $(N_j) \subset \mathbb{Z}_{>0}$ be sequence s.t. $N_j \to \infty$ as $j \to \infty$. Let μ be the weak limit of a subsequence of

$$\frac{1}{N_j} \sum_{n=0}^{N_j-1} T_*^n \nu_j$$

Then μ is T-invariant.

proof) Fix $f \in C(X)$. Wlog, assume $w - \lim \frac{1}{N_j} \sum_{n=0}^{N_j-1} T_*^n \nu_j = \mu$.

$$\begin{aligned}
\int f \circ T d\mu &= \lim_{j\to\infty} \int f \circ T d\Big(\frac{1}{N_j} \sum_{n=0}^{N_j-1} T^n) * \nu_j\Big) \\
&= \lim_{j\to\infty} \frac{1}{N_j} \sum_{n=0}^{N_j-1} \int f \circ T dT_*^n \nu_j \\
&= \lim_{j\to\infty} \frac{1}{N_j} \sum_{n=0}^{N_j-1} \int f \circ T^{n+1} d\nu_j \\
&= \lim_{j\to\infty} \frac{1}{N_j} \sum_{n=1}^{N_j} \int f \circ T^n d\nu_j
\end{aligned}$$

Now we can expand $\int f d\mu$ similarly

$$\int f d\mu = \lim_{j\to\infty} \frac{1}{N_j} \sum_{n=0}^{N_j-1} \int f \circ T^n d\nu_j$$

then

$$\Big| \int f d\mu - \int f \circ T d\mu \Big| \le \limsup_{j\to\infty} \frac{\|f\|_\infty + \|f\|_\infty}{N_j} = 0$$

(18 October, Thursday)

(Example sheet distributed. Example Class at 27 Oct, 10, 24 Nov. 2pm-4pm)
(Problem 9 and 10 to be submitted before Thursday 3pm)

Theorem) Let (X, T) be a topological dynamical system. The followings are equivalent :

(1) (X, T) is uniquely ergodic.

(2) $\forall f \in C(X)$, $\exists c_f \in \mathbb{C}$ s.t. $\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to c_f$ uniformly.

(3) $\exists A \subset C(f)$ dense s.t. $\forall f \in A$, $\exists c_f \in \mathbb{C}$ s.t.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to c_f \quad \forall x \in X$$

but not necessarily uniformly.

proof)

$(1) \Rightarrow (2)$ Suppose that (2) fails with $c_f = \int f d\mu$, where μ is the unique invariant measure. Then $\exists \epsilon > 0$, $\exists x_1, x_2, \cdots \in X$, $\exists (N_j)_j \subset \mathbb{Z}$ s.t.

$$\Big|\frac{1}{N_j}\sum_{n=0}^{N_j-1} f(T^n x_j) - \int f d\mu\Big| > \epsilon \quad \cdots\cdots(\star)$$

By using Bolzanno-Weierstrass theorem to restrict to a converging subsequence whenever necessary(and using diagonal argument)(noting that), we may suppose that

$$\frac{1}{N_j}\sum_{n=0}^{N_j-1} f(T^n x_j) \to a$$

for some $a \in \mathbb{C}$. Moreover, we can also assume that

$$\frac{1}{N_j}\sum_{n=0}^{N_j-1} T_*^n \delta_{x_j} \to \nu$$

for some probability measure ν. By the theorem from the previous lecture, ν is T-invariant. Also,

$$\begin{aligned}\int f d\nu &= \lim_{j\to\infty}\frac{1}{N_j}\sum_{n=0}^{N_j-1}\int f dT_*^n\delta_{x_j} \\ &= \lim_{j\to\infty}\frac{1}{N_j}\sum_{n=0}^{N_j-1} f(T^n x_j) = a\end{aligned}$$

By $(\star)$, $|a - \int f d\mu| > \epsilon$, so $\int f d\mu \neq f d\nu$, hence $\mu \neq \nu$, a contradiction.

$(2) \Rightarrow (3)$ This implication is trivial.

$(3) \Rightarrow (1)$ Let μ, ν be T-invariant probability measures. We will show that $\int f d\mu = \int f d\nu$ for all $f \in A$. Since A is dense, this also holds for all $f \in C(X)$. By the corollary to Riesz representation theorem, this implies $\mu = \nu$.

We know

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to c_f \quad \forall x \in X$$

By dominated convergence, has

$$\int f d\mu = \int \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) d\mu \to c_f$$

Thus $\int f d\mu = c_f$.
The same argument gives $\int f d\mu = c_f$.

(End of proof) $\square$

Example : Let $\alpha \in \mathbb{R}/\mathbb{Z}$ be irrational. Then the circle rotation $(\mathbb{R}/\mathbb{Z}, R_\alpha)$ is uniquely ergodic. Indeed, let μ be an R_α-invariant measure. Then

$$\int \exp(2\pi inx) d\mu = \int \exp(2\pi in R_\alpha(x)) d\mu$$
$$= \int \exp(2\pi in(x+\alpha)) d\mu = \exp(2\pi in\alpha) \int \exp(2\pi inx) d\mu$$

Since α is irrational, $\exp(2\pi in\alpha) \neq 1$ if $n \neq 0$. Then

$$(\dagger) \cdots\cdots \begin{cases} \int \exp(2\pi inx) d\mu = 0 & \forall n \neq 0 \\ \int 1 d\mu = 1 \end{cases}$$

Let f be a trigonometric polynomial, i.e. a finite lienar combination of the functions $\exp(2\pi inx)$, where $n \in \mathbb{Z}$. Then $(\dagger)$ implies

$$\int f d\mu = \int f(x) dx$$

Fact : Trigonometric polynomials is dense in $C(X)$ - use Stone-Weirestrass theorem.

Therefore, μ is just a Lebesgue measure.

Definition) A sequence $x_1, x_2, \cdots \in [0,1)$ is said to be **equidistributed** if

$$\frac{1}{N} \sum_{n=0}^{N-1} f(x_n) \to \int_{\mathbb{R}/\mathbb{Z}} f(x) dx \quad \forall f \in C(\mathbb{R}/\mathbb{Z})$$

Remark : Let $0 \leq a < b < 1$. Then $x_1, x_2, \cdots$ is equidistributed *if and only if*

$$\frac{1}{N} \Big| \{n \in [0, N-1] : x_n \in [a,b]\} \Big| \to a - b \quad \forall 0 \leq a < b < 1$$

Corollary) $\{n\alpha + x \bmod 1 : n \geq 0\}$ is equidistributed for all α irrational and $x \in [0,1)$.

This is the difference between pointwise ergodic theorem and unique ergodicity - unique ergodicity shows results for all α irrational, while pointwise erodic theorem shows for almost every points.

proof) This follows from the previous theorem and the example.

Open Problem : Classify the Borel probability measures on $\mathbb{R}/\mathbb{Z}$ that are invariant under both T_2 and T_3.

If only invariant under T_2, there are too many of them, so it is hopeless to classify them. However, if invariant under both T_2 and T_3, it is expected that the result is a combination of Lebesgue measure and measures derived from it.

7. Equidistribution of polynomials

Definition) Let $(X, \mathcal{B}, \mu, T)$ be a MPS, with X a compact metric space, and $T : X \to X$ is continuous. Then $x \in X$ is called **generic w.r.t.** μ if the following holds :

$$\begin{aligned} &\frac{1}{N}\sum_{n=0}^{N-1} T_*^n \delta_x \to \mu \quad \text{weakly} \\ \Leftrightarrow \quad &\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int f d\mu \quad \forall f \in C(X) \quad \cdots\cdots (\star) \end{aligned}$$

Lemma) μ-almost every $x \in X$ is μ-generic.

> **proof)** By the pointwise ergodic theorem, for $\forall f \in C(X)$, there is a set X_f with $\mu(X_f) = 1$ such that $(\star)$ holds. Observe that every point in $\bigcap_{f \in A} X_f$ where $A \subset C(X)$ is dense and countable is μ-generic.
>
> *(End of proof)* $\square$

(20th October, Saturday)

We seek for a generalized version of the previous corollary.

Theorem) *(Furstenbeg)* Let (X, T) be a uniquely ergodic topological dynamical system. Denote by μ the invariant measure. Write

$$\begin{aligned} S : X \times \mathbb{R}/\mathbb{Z} &\to X \times \mathbb{R}/\mathbb{Z} \\ (x, y) &\mapsto (Tx, y + c(x)) \end{aligned}$$

where $c : X \to \mathbb{R}/\mathbb{Z}$ is a fixed continuous function. Then $\mu * m$, where m is the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$ is S-invariant. If $\mu \otimes m$(the product measure) is S-ergodic, then $(X \times \mathbb{R}/\mathbb{Z}, S)$ is uniquely ergodic.

This has name of **skew-product**.

> **proof)** Let $f \in C(X \times \mathbb{R}/\mathbb{Z})$. Then
>
> $$\begin{aligned} \iint f \circ S(x, y) d\mu(x) dy &= \int_X \int_0^1 f(Tx, y + c(x)) dy d\mu(x) = \int_X \int_{-c(x)}^{1-c(x)} f(Tx, y) dy d\mu(x) \\ &= \int_{\mathbb{R}/\mathbb{Z}} \int_X f(Tx, y) d\mu(x) dy = \iint f(x, y) d\mu(x) dy \end{aligned}$$
>
> So $\mu \otimes m$ is indeed S-invariant.
>
> Now we assume that $\mu * m$ is S-ergodic. We show that $(X \times \mathbb{R}/\mathbb{Z}, S)$ is uniquely ergodic.
>
> Recall the definition :

Definition) A point $(x,y) \in X \times \mathbb{R}/\mathbb{Z}$ is **generic** if

$$\frac{1}{N}\sum_{n=0}^{N-1} S_*^n \delta(x,y) \xrightarrow{\text{weakly}} \mu \otimes m$$

$$\Leftrightarrow \quad \frac{1}{N}\sum_{n=0}^{N-1} f(S^n(x,y)) \to \int f d\mu dm \quad \forall f \in C(X \times \mathbb{R}/\mathbb{Z})$$

Let E be the set of $\mu \otimes m$-generic points. We showed last time, ergodicity implies $\mu \otimes m(E) = 1$.

Claim : If $(x,y) \in E$, then $(x, t+y) \in E$ for all $t \in \mathbb{R}/\mathbb{Z}$.
proof) Observation : $S \circ U_t = U_t \circ S$, where $U_t(x,y) = (x, t+y)$. Indeed,

$$S \circ U_t(x,y) = S(x,t+y) = (Tx, t+y+c(x)) = U_t(Tx, y+c(x)) = U_t \circ S(x,y)$$

Let $f \in C(X \times \mathbb{R}/\mathbb{Z})$. Write

$$\frac{1}{N}\sum_{n=0}^{N-1} f(S^n(x+ty)) = \frac{1}{N}\sum_{n=0}^{N-1} f(S^n \circ U_t(x,t))$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} f(U_t \circ S^n(x,t)) \xrightarrow{N\to\infty} \iint f \circ U_t dm d\mu = \iint f dm d\mu \quad \text{since } (x,y) \in E$$

So $(x,t+y)$ is indeed generic, i.e. $(x,t+y) \in E$.

This means $E = A \times \mathbb{R}/\mathbb{Z}$ for some $A \subset X$ Borel set. Note, $\mu(A) = \mu \otimes m(E) = 1$. Let ν be an S-invariant measure on $X \times \mathbb{R}/\mathbb{Z}$.

We aim to prove that $\nu(E) = 1$: Write P for the projection $X \times \mathbb{R}/\mathbb{Z} \to X$. We show $P_*\nu = \mu$. It is enough to show that $P_*\nu$ is T-invariant. Let $B \subset X$, a Borel set. Then

$$\begin{aligned} P_*\nu(T^{-1}(B)) &= \nu(T^{-1}(B) \times \mathbb{R}/\mathbb{Z}) = \nu(S^{-1}(B \times \mathbb{R}/\mathbb{Z})) \\ &= \nu(B \times \mathbb{R}/\mathbb{Z}) = P_*\nu(B) \end{aligned}$$

Since $(X, \mathcal{B}, \mu, T)$ was assumed to be uniquely ergodic, this forces us to have $P_*\nu = \mu$. Then $P_*\nu(A) = \mu(A) = 1$, and therefore $\nu(E) = \nu(A \times \mathbb{R}/\mathbb{Z}) = 1$.

Finally, we show that $\int f d\nu = \iint f d\mu dm$ for all $f \in C(X \times \mathbb{R}/\mathbb{Z})$ and this proves $\nu = \mu \otimes m$: If $(x,y) \in E$, then

$$\frac{1}{N}\sum_{n=0}^{N-1} f(S^n(x,y)) \to \iint f d\mu dm$$

But since $\nu(E) = 1$, this holds ν-a.e. By dominated convergence,

$$\int f d\nu = \int \frac{1}{N}\sum_{n=0}^{N-1} f(S^n(x,y)) d\nu \to \iint f d\mu dm$$

So we have $\int f d\nu = \iint f d\mu dm$.

(End of proof) $\square$

From the theorem, we can prove a generalized version of the example from the previous lecture.

Corollary) Let $S : (\mathbb{R}/\mathbb{Z})^d \to (\mathbb{R}/\mathbb{Z})^d$, defined by

$$S(x_1, x_2, \cdots, x_d) = (x_1 + \alpha, x_2 + x_1, \cdots, x_d + x_{d-1})$$

where $\alpha \in (\mathbb{R}/\mathbb{Z})\backslash\mathbb{Q}$ is a fixed irrational number. Then $((\mathbb{R}/\mathbb{Z})^d, S)$ is uniquely invariant.

proof) Prove by induction on d.

- $d = 1$ case is the circle rotation that we already discussed.
- Suppose $d \geq 2$ and the claim holds for $d - 1$. By Furstenberg's theorem, it is enough to show that $((\mathbb{R}/\mathbb{Z})^d, \mathcal{B}, m^d, S)$ is ergodic (m is the Lebesgue measure).

  Let f be a bounded measurable function on $(\mathbb{R}/\mathbb{Z})^d$. Then

$$\begin{aligned} f(x) &= \sum_{n \in \mathbb{Z}^d} a_n \exp(2\pi i n \cdot x) \quad \text{a.e.} \\ f(S(x)) &= \sum_{n \in \mathbb{Z}^d} a_n \exp(2\pi i (n_1(x_1 + \alpha) + n_2(x_2 + x_1) + \cdots + n_d(x_d + x_{d-1}))) \\ &= \sum_{n \in \mathbb{Z}^d} \exp(2\pi i n_1 \alpha) a_n \exp(2\pi i ((n_1 + n_2)x_1 + \cdots + (n_{d-1} + n_d)x_{d-1} + n_d x_d)) \\ &= \sum_{n \in \mathbb{Z}^d} \exp(2\pi i n_1 \alpha) a_n \exp(2\pi i \hat{S}(n) \cdot x) \end{aligned}$$

  where $\hat{S}(n) = (n_1 + n_2, n_2 + n_3, \cdots, n_{d-1} + n_d, n_d)$.

  Suppose $f = f \circ S$ a.e. Then $a_{\hat{S}(n)} = \exp(2\pi i \alpha n_1) a_n$. Suppose that $a_m \neq 0$ for some $m \in \mathbb{Z}^d$. We aim to show $m = 0$, which implies that f is constant : By Parseval's formula,

$$\sum_{n \in \mathbb{Z}} |a_n|^2 = \|f\|_2^2 < \infty$$

  This means that there are at most finite n's such that $|a_m| = |a_n|$. In particular, the orbit $m, \hat{S}(m), \hat{S}^2(m), \cdots$ must be periodic. Note $(\hat{S}^j(m))_{d-1} = m_{d-1} + j m_d$. Thus $m_d = 0$. Similar argument gives $m_j = 0$ for all $j = 2, 3, \cdots, d$. Hence we should have $m = (m_1, 0, \cdots, 0)$ and $\hat{S}^j(m) = (m)$. We now use $a_m = \exp(2\pi i \alpha m_1) a_m$. Since $a_m \neq 0$, we must have $\exp(2\pi i \alpha m_1) = 1$. As α is irrational, this implies $m_1 = 0$.

(End of proof) $\square$

(23 October, Tuesday)

(Examples Class : This Saturday 2-4pm, MR12)

We meet the main theorem of the chapter.

Theorem) *(Weyl)* Let $P(x) = a_d x^d + \cdots + a_1 x + a_0$ be polynomial such that a_j is irrational for at least one $j \neq 0$. Then the sequence $\{P(n)\}_{n>0}$ is equdistributed in $[0, 1)$ mod $\mathbb{Z}$.

proof) First consider the case when a_d is irrational. Recall that the system $(\mathbb{R}/\mathbb{Z}, S)$ is uniquely ergodic, where $S(x_1, \cdots, x_d) = (x_1 + \alpha, x_2 + x_1, \cdots, x_d + x_{d-1})$ and $\alpha \in \mathbb{R}/\mathbb{Z}$ irrational.

Note :

$$S^n \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} x_1 + n\alpha \\ x_2 + nx_1 + \binom{n}{2}\alpha \\ \vdots \\ x_d + nx_{d-1} + \cdots + \binom{n}{d-1} x_1 + \binom{n}{d}\alpha \end{pmatrix}$$

This can be proved by induction on n.

Consider the polynomials $q_j = t(t-1)\cdots(t-j+1)/j!$. The polynomials $q_0, q_1, \cdots, q_d$ form a basis in the vector space of polynomials of degree $\leq d$. In particular, there are $x_1, \cdots, x_d, \alpha \in \mathbb{R}$ such that

$$p(t) = \alpha q_d(t) + x_1 q_{d-1}(t) + \cdots + x_d q_0(t) \quad \cdots\cdots\cdots (\star)$$

with $\alpha = a_d \cdot d!$ is irrational. Then there are $\alpha, x_1, \cdots, x_d \in \mathbb{R}/\mathbb{Z}$ such that

$$p(n) = \alpha \binom{n}{d} + x_1 \binom{n}{d-1} + \cdots + x_d \binom{n}{0} \quad \mod \mathbb{Z}$$

for all $n \in \mathbb{Z}$. Let $f \in C(\mathbb{R}/\mathbb{Z})$ and let $g(x_1, \cdots, x_d) = f(x_d) \in (\mathbb{R}/\mathbb{Z})^d$. Now we have :

$$\frac{1}{N}\sum_{n=0}^{N-1} f(p(n)) = \frac{1}{N}\sum_{n=0}^{N-1} g(S^n(x_1, \cdots, x_d)) \quad \cdots\cdots (\dagger)$$

where $x_1, \cdots, x_d$ are as in $(\star)$ and α in the definition of S is also coming from $(\star)$. By unique ergodicity, $(\dagger)$ converges to

$$\int \cdots \int g(t_1, \cdots, t_d) dt_1 \cdots dt_d = \int f(t) dt$$

for all $x \in [0, 1)$ uniformly. This proves equidistribution.

General case : Let j be maximal such that a_j is irrational. Let $q \in \mathbb{Z}_{>0}$ be such that $qa_d, \cdots, qa_{d-1}, \cdots, qa_{j+1} \in \mathbb{Z}$. Fix $b \in \{0, 1, \cdots, q-1\}$.

Note :

$$p(qn + b) = a_d b^d + \cdots a_{j+1} b^{j+1} + a_j (qn + b)^j + \cdots a_1 (qn + b) + a_0 \quad \mod \mathbb{Z}$$

This is a polynomial in n with irrational leading coefficient. By the special case proved earlier, $\{p(qn + b)\}$ equidistributes for each fixed b.

(End of proof) $\square$

This theorem was proved by number theorists way before ergodic theory was founded. There are in fact many ways to prove this theorem, e.g. using Harmonic analysis. The proof using Harmonic analysis even gives the rate of convergence to the equidistributed state, while no proof with ergodic theory does. However, there are yet more sophisticated versions of this theorem, which cannot be proved (at least as long as it is known) without using ergodic theory.

8. Mixing Properties

Definition) An MPS $(X, \mathcal{B}, \mu, T)$ is called **mixing** if $\forall A, B \in \mathcal{B}$ and $\forall \epsilon > 0$, there is $N > 0$ such that

$$\Big|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\Big| < \epsilon \quad \forall n > N$$

Definition) An MPS $(X, \mathcal{B}, \mu, T)$ is called a **mixing on k sets** if $\forall A_0, A_1, \cdots, A_{k-1} \in \mathcal{B}$ and $\forall \epsilon > 0$, there is N such that

$$\Big|\mu(A_0 \cap T^{-n_1}A \cap \cdots \cap T^{-nk-1}A_{k-1}) - \mu(A_0) - \cdots - \mu(A_{k-1})\Big| < \epsilon$$

for all $n_1, \cdots, n_{k-1}$ if $n_1 > N$, $n_2 - n_1 > N$, $\cdots$, $n_{k-1} - n_{k-2} > N$.

Open Problem : Is there an MPS that is mixing on 2 sets but not on 3 sets?

Definition)

- A subset $S \subset \mathbb{Z}_{>0}$ has **full density** if

$$\frac{|S \cap [1, N]|}{N} \to 1 \quad \text{as } N \to \infty$$

- We say that the sequence of complex numbers (a_n) **converge in density** to $a \in \mathbb{C}$ if $\{n : |a_n - a| < \epsilon\}$ has full density for all $\epsilon > 0$.

  In notation, write $D - \lim_{n\to\infty} a_n = a$.

- We say that (a_n) **Cesàro-converges** to a if

$$\frac{1}{N}\sum_{n=1}^{N} a_n \to a \quad \text{as } N \to \infty$$

  Denote $C - \lim_{n\to\infty} a_n = a$

Definition) An MPS $(X, \mathcal{B}, \mu, T)$ is **weak mixing** if $\forall A, B \in \mathcal{B}$, we have

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

(25th October, Thursday)

Lemma) Let $(a_n) \subset \mathbb{R}$ be a *bounded sequence*. Let $a \in \mathbb{R}$. Then the following are equivalent.

(1) $D - \lim_{n\to\infty} a_n = a$.

(2) $C - \lim_{n\to\infty} |a_n - a| = 0$.

(3) $C - \lim_{n\to\infty} |a_n - a|^2 = 0$.

(4) $C - \lim_{n\to\infty} a_n = a$ and $C - \lim a_n^2 = a^2$.

proof)

(1) $\Rightarrow$ (2) Fix $\epsilon > 0$. Let $M = \sup |a_n|$. By assumption we may pick N that is large enough so that

$$\frac{1}{N}\Big|\{n \in [1, N] : |a_n - a| > \epsilon\}\Big| < \epsilon$$

We estimate

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}|a_n - a| &\leq \frac{1}{N}\Big(\epsilon N + 2M \cdot \epsilon N\Big)\\ &= \epsilon(1 + 2M)\end{aligned}$$

Since M is constant and ϵ is arbitrary, we have

$$\frac{1}{N}\sum_{n=1}^{N}|a_n - a| \to 0 \quad \text{as } N \to \infty$$

(2) $\Rightarrow$ (1) Fix $\epsilon > 0$. Then

$$\Big|\{n \in [1, N] : |a_n - a| > \epsilon\}\Big|\frac{1}{\epsilon} \leq \sum_{n=1}^{N}|a_n - a|$$

By (2), one gets

$$\frac{1}{N}|\{n \in [1, N] : |a_n - a| > \epsilon\}| \to 0 \quad \text{as } N \to \infty$$

for each ϵ. This proves (1).

(1) $\Leftrightarrow$ (3) Use same arguments.

(1),(2) $\Rightarrow$ (4)

$$\Big|\frac{1}{N}\sum_{n=1}^{N}(a_n - a)\Big| \leq \frac{1}{N}\sum_{n=1}^{N}|a_n - a| \to 0 \quad \text{by (2)}$$

This implies $C - \lim a_n = a$.

Now by the definition of $D-\lim$, we see that $D-\lim a_n = a$ implies $D-\lim a_n^2 = a^2$ and this implies $C - \lim |a_n^2 - a^2| \to 0$. But by our previous part of the proof, this also implies $C - \lim a_n^2 = a^2$.

(4) $\Rightarrow$ (3)

$$\frac{1}{N}\sum_{n=1}^{N}(a_n - a)^2 = \frac{1}{N}\sum_{n=1}^{N}a_n^2 + \frac{1}{N}\sum_{n=1}^{N}a^2 - 2\frac{1}{N}\sum_{n=1}^{N}aa_n \to a^2 + a^2 - 2a \cdot a = 0 \quad \text{as } N \to \infty$$

(End of proof) $\square$

Theorem) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. The followings are equivalent :

(1) $(X, \mathcal{B}, \mu, T)$ is weak mixing.

(2) $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu, T \times S)$ is ergodic for any ergodic MPS $(Y, \mathcal{C}, \nu, S)$.

(3) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu, T \times T)$ is ergodic.

(4) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu, T \times T)$ is weak mixing.

(5) U_T has no non-constant eigenfunction, i.e. if $f : X \to \mathbb{C}$ is measurable and $\lambda \in \mathbb{C}$ such that $f \circ T = \lambda f$ almost everywhere, then f is constant almost everywhere.

Implication from (5) to the others involves some functional analysis, which we do not assume in this course. So we will not prove the theorem in full in the lecture. But can find a guide to the proof in the example sheet.

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\mathcal{S} \subset \mathcal{B}$ be a semi-algebra(π-system) that generates $\mathcal{B}$. Then

(i) $(X, \mathcal{B}, \mu, T)$ is weak mixing *if and only if*

$$D-\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \quad \forall A, B \in \mathcal{S}$$

(ii) $(X, \mathcal{B}, \mu, T)$ is ergodic *if and only if*

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

proof) (In example sheet) We aim to use Dynkin's lemma. Recall :

Dynkin's lemma : A d-system containing a π-system Π also contains $\sigma(\Pi)$, the σ-algebra generated by Π.

The backward implications are trivial, so we just prove forward implications here.

(i) First,

$$\mathcal{D} = \{A \in \mathcal{B} : D-\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \quad \forall B \in \mathcal{S}\}$$

Then for any $B \in \mathcal{S}$, $A \in D$ and $(A_n)_n \subset D$ disjoint, we have

$$D-\lim_{n\to\infty} \mu(T^{-n}A^c \cap B) = \mu(B) - D-\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A^c)\mu(B)$$

and

$$D-\lim_{n\to\infty} \mu(T^{-n} \cup_n A_n \cap B) = \sum_n D-\lim_{n\to\infty} \mu(T^{-n}A_n \cap B) = \mu(\cup_n A_n)\mu(B)$$

and hence $A^c \in \mathcal{D}$ and $\cup_n A_n \in \mathcal{D}$. Therefore, D is a d-system and by Dynkin's lemma, $\mathcal{D} = \sigma(\mathcal{S}) = \mathcal{B}$.

Next, let

$$\mathcal{D}' = \{B \in \mathcal{B} : D-\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \quad \forall A \in \mathcal{B}\}$$

then we can show accordingly that $\mathcal{D}'$ is a d-system, and hence is in fact equal to $\mathcal{B}$.

(ii) We may use exactly the same method to show that

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

Now suppose A is a T-invariant set, so $T^{-n}A=A$ for all $n\geq 0$. Application of the above formula with $B=A$ gives

$$\lim_{N\to\infty}\frac{1}{N}\sum_n\mu(A\cap A)=\mu(A)=\mu(A)\mu(A)$$

which says that $\mu(A)\in\{0,1\}$. This proves that the MPS is ergodic.

(End of proof) $\square$

We prove the theorem using this lemma.

proof of the theorem)

(1) $\Rightarrow$ (2) Let $\mathcal{S}$ be the set of measurable rectangles, i.e. the set of the form $B\times C$, where $B\in\mathcal{B}$, $C\in\mathcal{C}$. We write for $B_1\times C_1$, $B_2\times C_2\in\mathcal{S}$. Then

$$\begin{aligned}
&\Big|\frac{1}{N}\sum_{n=1}^{N}\Big[\mu\otimes\nu\big((T\times S)^{-n}(B_1\times C_1)\cap(B_2\times C_2)\big)-\mu\otimes\nu(B_1\times C_1)\mu\otimes\nu(B_2\times C_2)\Big]\Big|\\
=&\Big|\frac{1}{N}\sum_{n=1}^{N}\Big[\mu(T^{-n}B_1\cap B_2)\nu(S^{-n}C_1\cap C_2)-\mu(B_1)\mu(B_2)\nu(C_1)\nu(C_2)\Big]\Big|\\
\leq&\frac{1}{N}\Big|\sum_{n=1}^{N}\Big[\mu(T^{-n}B_1\cap B_2)\nu(S^{-n}C_1\cap C_2)-\mu(B_1)\mu(B_2)\nu(S^{-n}C_1\cap C_2)\Big]\Big|\\
&+\frac{1}{N}\Big|\sum_{n=1}^{N}\Big[\mu(B_1)\mu(B_2)\nu(S^{-n}C_1\cap C_2)-\mu(B_1)\mu(B_2)\nu(C_1)\nu(C_2)\Big]\Big|\\
\leq&\frac{1}{N}\Big|\sum_{n=1}^{N}\Big[\mu(T^{-n}B_1\cap B_2)-\mu(B_1)\mu(B_2)\Big]\Big|\qquad(\text{as }\nu(S^{-n}C_1\cap C_2)\leq 1)\\
&+\frac{1}{N}\mu(B_1)\mu(B_2)\Big|\sum_{n=1}^{N}\Big[\nu(S^{-n}C_1\cap C_2)-\nu(C_1)\nu(C_2)\Big]\Big|\to 0\quad\text{by ergodicity}
\end{aligned}$$

Therefore this converges to 0 as $N\to 0$. So by the previous lemma, $(X\times Y,\mathcal{B}\otimes\mathcal{C},\mu\otimes\nu,T\times S)$ is ergodic.

(2) $\Rightarrow$ (3) (3) is a special case of (2) if we show that $(X,\mathcal{B},\mu,T)$ is ergodic. This can be seen by applying (2) with $|Y|=1$ and $S=1_Y$.

(3) $\Rightarrow$ (1) Let $A,B\in\mathcal{B}$. We have

$$\begin{aligned}
&C-\lim\mu\otimes\mu((T\times T)^{-n}(A\times X)\cap(B\times X))=\mu\otimes\mu(A\times X)\cdot\mu\otimes\mu(B\times X)\\
=&C-\lim\mu(T^{-n}A\cap B)\mu(T^{-n}X\cap X)=\mu(A)\mu(B)\mu(X)^2\\
=&C-\lim\mu(T^{-n}A\cap B)=\mu(A)\mu(B)
\end{aligned}$$

Same argument with $A\times A$ and $B\times B$ in place of $A\times X$ and $B\times X$ gives :

$$C-\lim\mu(T^{-n}A\cap B)^2=(\mu(A)\cdot\mu(B))^2$$

Then by last last lemma (4) $\Rightarrow$ (1), we have

$$D-\lim \mu(T^{-n}A\cap B)=\mu(A)\mu(b)$$

This proves (1).

(1) $\Leftrightarrow$ (4) Use same arguments.

(27th October, Saturday)

proof continued)

(3) $\Rightarrow$ (5) Suppose that $f\circ T=\lambda f$ μ-a.e. for $f\in L^2$ and $\lambda\in\mathbb{C}$. Consider the function $\tilde{f}(x,y)=f(x)\bar{f}(u)$. We will show that $\tilde{f}$ is $T\times T$ invariant. We have

$$\tilde{f}(Tx,Ty)=f(Tx)\cdot\bar{f}(Ty)=\lambda f(x)\bar{\lambda}\bar{f}(y)=|\lambda|^2\tilde{f}(x,y)$$

Since U_T is an isometry,

$$\langle f,f\rangle=\langle U_Tf,U_Tf\rangle=\langle\lambda f,\lambda f\rangle=|\lambda|^2\langle f,f\rangle$$

so $|\lambda|=1$ and $\tilde{f}\circ(T\times T)=\tilde{f}$ indeed. However, if f is not constant, then $\tilde{f}$ is also not constant, so $(X\times X,\cdots)$ is not ergodic. This gives a contradiction, so f need be a constant

(5) $\Rightarrow$ (3) This implication requires some knowledge in functional analysis. Will not be dealt here.

(End of proof) $\square$

We can in fact do better then this.

9. Multiple recurrence for weak mixing systems

Theorem) Let $(X,\mathcal{B},\mu,T)$ be a weak mixing MPS, and let $k\in\mathbb{Z}_{>0}$. Let $f_1,\cdots,f_k\in L^\infty$. Then :

$$\frac{1}{N}\sum_{n=1}^{N}U_T^nf_1\cdots U_T^{2n}f_2\cdots U_T^{kn}f_k\xrightarrow{L^2}\int f_1d\mu\cdots\int f_kd\mu$$

Corollary) $(X,\mathcal{B},\mu,T)$ be a weak mixing MPS, and let $k\in\mathbb{Z}_{>0}$. Let $f_0\in L^\infty$. Then :

$$\frac{1}{N}\sum_{n=1}^{N}\int f_0U_T^nf_1\cdots U_T^{kn}f_kd\mu\to\int f_0d\mu\int f_1d\mu\cdots\int f_kd\mu$$

In particular, let $f_0=f_1=\cdots=f_k=\chi_A$ for some $A\in\mathcal{B}$. Then

$$\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-n}A\cap\cdots\cap T^{-nk}A)\to\mu(A)^{k+1}$$

proof) Let $g_N = \frac{1}{N}\sum_{n=1}^N U_T^n f_1 \cdots U_T^{2n} f_2 \cdots U_T^{kn} f_k$. Then the above theorem says that $g_n \to \gamma$ in L^2, where γ is as in the theorem. Then it also follows that $\langle f_0, \bar{g}_n\rangle \to \langle f_0, \bar{\gamma}\rangle$, which is exactly the statement of the corollary.

(End of proof) $\square$

We need a lemma to prove the theorem.

Lemma) *(van der Corput)* Let $u_1, u_2, \cdots$ be a bounded sequence in a Hilbert space $\mathcal{H}$. For each $h \in \mathbb{Z}_{\geq 0}$, write

$$s_h = \limsup_{N\to\infty} \Big| \frac{1}{N}\sum_{n=1}^N \langle u_n, u_{n+h}\rangle \Big|$$

Suppose that $D - \lim_{h\to\infty} s_h = 0$. Then

$$\Big\| \frac{1}{N}\sum_{n=1}^N u_n \Big\| \to 0 \quad \text{as } n \to \infty$$

(Idea)

$$\begin{aligned}
\Big\|\frac{1}{N}\sum_{n=1}^N u_n\Big\|^2 &= \frac{1}{N^2}\sum_{n_1=1}^N\sum_{n_2=1}^N \langle u_{n_1}, u_{n_2}\rangle \\
&= \frac{1}{N^2}\Big(\sum_{n=1}^N \|u_n\|^2 + \sum_{h=1}^{N-1}\sum_{n=1}^{N-k} 2\Re(\langle u_n, u_{n+h}\rangle)\Big)
\end{aligned}$$

proof) Fix $\epsilon > 0$. By the assumption, we can find H such that

$$\frac{1}{H}\sum_{h=0}^{H-1} s_h < \epsilon$$

Write

$$\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^N u_n - \frac{1}{NH}\sum_{n=1}^N\sum_{h=1}^H u_{n-h}\Big\| \\
\leq &\frac{1}{N}\Big[\sum_{n=1}^H \|u_h\| + \sum_{n=1}^{N+H}\|u_n\|\Big] < \epsilon \quad \text{if } N \text{ is large enough.}
\end{aligned}$$

and

$$\frac{1}{NH}\Big\|\sum_{n=1}^N\sum_{h=1}^H u_{n+h}\Big\| \leq \frac{1}{NH}\sum_{n=1}^N\Big\|\sum_{h=1}^H u_{n+h}\Big\|$$

Using Cauchy-Schwarz on the expression on the right hand side of the above equation, we have

$$\begin{aligned}
\frac{1}{N^2H^2}\Big\|\sum_{n=1}^N\sum_{h=1}^H u_{n+h}\Big\|^2 &\leq \frac{1}{NH^2}\sum_{n=1}^N\Big\|\sum_{h=1}^H u_{n+h}\Big\|^2 \\
&= \frac{1}{NH^2}\sum_{n=1}^N\sum_{h_1=1}^H\sum_{h_2=1}^H \langle u_{n+h_1}, u_{n+h_2}\rangle \\
&\leq \frac{1}{H^2}\sum_{h_1=1}^H\sum_{h_2=1}^H \frac{1}{N}\Big|\sum_{n=1}^N \langle u_{n+h_1}, u_{n+h_2}\rangle\Big|
\end{aligned}$$

Take $N \to \infty$, then we get

$$\limsup_{N\to\infty} \frac{1}{N^2H^2}\Big\|\sum_{n=1}^{N}\sum_{h=1}^{H} u_{n+h}\Big\|^2 \leq \frac{1}{H^2}\sum_{h_1=1}^{H}\sum_{h_2=1}^{H} s_{|h_1-h_2|}$$
$$\leq \frac{1}{H^2}\sum_{h=0}^{H-1} 2H\cdot s_h < 2\epsilon$$

Combining what we have,

$$\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=1}^{N} u_n\Big\| \leq \epsilon + \sqrt{2\epsilon}$$

ϵ was chosen arbitrary, so we have the desired result.

(End of proof) $\square$

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing measure preserving system. Let $f, g \in L^2$. Then

$$D-\lim\langle U_T^n f, g\rangle = \int f d\mu \cdot \int \bar{g} d\mu$$

If we plug in characteristic functions in f and g, this is precisely the definition of weak mixing. So we can think of this as an extension of weak mixing definition.

Lemma) If $(X, \mathcal{B}, \mu, T)$ is weak mixing, then so is $(X, \mathcal{B}, \mu, T^k)$.

proof) Proof of two lemmas are exercises.

(30th October, Tuesday)

We will prove the main theorem of the section using van der Corput lemma from the previous lecture. In the proof, the statement of its corollary would also be useful.

Corollary) $(X, \mathcal{B}, \mu, T)$ be a weak mixing MPS, and let $k \in \mathbb{Z}_{>0}$. Let $f_0 \in L^\infty$. Then :

$$\frac{1}{N}\sum_{n=1}^{N}\int f_0 U_T^n f_1 \cdots U_T^{kn} f_k d\mu \to \int f_0 d\mu \int f_1 d\mu \cdots \int f_k d\mu$$

Theorem) Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing MPS, and let $k \in \mathbb{Z}_{>0}$. Let $f_1, \cdots, f_k \in L^\infty$. Then :

$$\frac{1}{N}\sum_{n=1}^{N} U_T^n f_1 \cdots U_T^{2n} f_2 \cdots U_T^{kn} f_k \xrightarrow{L^2} \int f_1 d\mu \cdots \int f_k d\mu$$

proof) Proof is by induction on k. If $k = 1$, then this reduces to the mean ergodic theorem. Suppose that $k > 1$ and the statement of the theorem and the statement of the corollary holds for $k - 1$.

We first consider the special case $\int f_k d\mu = 0$. Aim to apply van der Corput's lemma for

$$u_n = U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{kn} f_k \in L^2(X)$$

We compute s_k.

$$\begin{aligned}\langle u_n, u_{n+h}\rangle &= \int U_T^n f_1 \cdots U_T^{kn} f_k \cdot U_T^{n+h} \bar{f}_1 \cdots U_T^{k(n+h)} \bar{f}_k d\mu \\ &= \int f_1 \cdot U_T^h \bar{f}_1 \cdot U_T^n (f_2 U_T^{2h} \bar{f}_2) \cdots U_T^{(k-1)n} (f_k \cdot U^{kh} \bar{f}_k) d\mu\end{aligned}$$

Apply the **Corollary** for $k-1$ and $f_j U_T^{jh} \bar{f}_j$ in the role of f_{j-1}. Then we get :

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+h}\rangle = \int f_1 U_T^k \bar{f}_1 d\mu \cdots \int f_k U_T^{kh} \bar{f}_k d\mu$$

Also as $|f_1 U_T^k \bar{f}_1| \leq \|f_1\|_\infty^2$ and similarly for the other terms,

$$\limsup_{N\to\infty} \frac{1}{N} \Big| \sum_{n=1}^{N} \langle u_n, u_{n+h}\rangle \Big| \leq \|f_1\|_\infty^2 \cdots \|f_{k-1}\|_\infty^2 \Big| \int f_k U_T^{kh} \bar{f}_k d\mu \Big|$$

By a lemma stated last time $(X, \mathcal{B}, \mu, T^k)$ is a weak mixing. By the other lemma, we then have :

$$D - \lim_{h\to\infty} \int f_k U_T^{kh} \bar{f}_k d\mu = \int f_k \cdot \int f_k = 0$$

and therefore $D - \lim s_k = 0$. Then van der Corput's lemma applies and we get the claimed result for this special case.

In the general case, we can write $f_k = \int f_k d\mu + f_k'$, where $\int f_k' d\mu = 0$. Then

$$\begin{aligned}\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 U_T^{2n} f_2 \cdots U^{nk} f_k =& \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 U_T^{2n} f_2 \cdots U^{nk} f_k' \\ &+ \int f_k d\mu \cdot \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 U_T^{2n} f_2 \cdots U^{n(k-1)} f_{k-1}\end{aligned}$$

The first term converges to 0 by our special case showed earlier and the second term converges to

$$\int f_1 d\mu \cdot \int f_{k-1} d\mu$$

by induction hypothesis.

(End of proof) $\square$

Cutting and Stacking

In this small sub-section, we will be constructing an example showing that weak mixing is not equal to mixing. The construction is going to be bit technical.

- For each $n \in \mathbb{Z}_{>0}$, let $I_1^{(n)}, \cdots, I_{N(n)}^{(n)}$ be a collection of subintervals $I_j^{(n)} = [a_j^{(n)}, b_j^{(n)}) \subset [0, 1)$. Suppose they satisfy the following properties.

(1) For each fixed n, $I_j^{(n)}$ are pairwise disjoint and they have the same length.

(2) $\text{length}(I_1^{(n)}) \to 0$ as $n \to \infty$ and $\mu([0,1)\backslash \bigcup_{j=1}^{N(n)} I_j^{(n)}) \to 0$ as $n \to \infty$.

(3) $\bigcup_{j=1}^{N(n)} I_j^{(n)} \subset \bigcup_{j=1}^{N(n+1)} I_j^{(n+1)}$.

(4) If $I_{j'}^{(n+1)} \cap I_j^{(n)} \neq \phi$ for some n, j, j' then $I_j^{(n+1)} \subset I_j^{(n)}$ and if $j \neq N(n)$ then $I_{j'}^{(n+1)} \subset I_j^{(n)}$ and if $j \neq N(n)$ then $a_{j'+1}^{n+1} - a_{j'}^{n+1} = a_{j+1}^{n+1} - a_j^{n+1}$.

Lemma) Given $I_j^{(n)}$ satisfy the above conditions, there is a unique (up to a set of measure 0) transformation $T : [0,1) \to [0,1)$ such that T maps $I_j^{(n)}$ onto $I_{j+1}^{(n)}$ by a translation. That is,

$$T\big|_{I_j^{(n)}} : x \mapsto x + a_{j+1}^{(n)} - a_j^{(n)}$$

This holds for all n and for all $j = 1, \cdots N(n) - 1$. Moreover T preserves the Lebesgue measure on $[0,1)$.

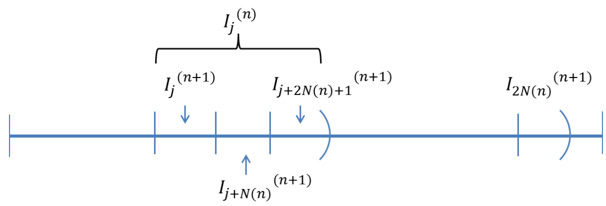
proof) See example sheets. (Note that the condition (4) above is imposed precisely in order to state this lemma)

The Chacón map

Let $I_1^{(1)} = [0, \frac{2}{3})$ and $N(1) = 1$. Suppose that $I_j^{(n)}$ are defined for some n. Cut $I_j^{(n)}$ into 3 equal pieces. Define $I_j^{(n+1)}$ to be the piece on the left hand side, $I_{j+N(n)}^{(n+1)}$ be the piece in the middle and $I_{j+2N(n)+1}^{(n+1)}$ be the piece on the right hand side. Also, let $N(n+1) = 3N(n) + 1$. Now cut off an interval of the same length as the other I_j^{n+1}'s from $[0,1)\backslash \bigcup_{j=1}^{N(n)} I_j^{(n)}$ and let this be $I_{2N(n)+1}^{(n+1)}$.

****

A more detailed explanation of the Chacón map,
Reference : http://math.uchicago.edu/ may/REU2015/REUPapers/Chen,Wenyu.pdf



The construction is basically "cutting the column into three, putting a spacer above the middle column, and stacking from left to right." We inductively construct the columns $C_0, C_1, C_2 \cdots$.

Let **column**(C) be a finite sequence of disjoint intervals of the same length. Each interval(I) is called a **level** and the number of intervals in a column is called the **height**(h) of the column.

First, let C_0 denote the column consisting of a single interval $I_{0,0} = [0, \frac{2}{3})$. Then $h_0 = 1$.

Construction of C_1 : divide C_0 into three disjoint subintervals and put a spacer $S_0 = [2/3, 8/9)$. Then the union of all the intervals is $[0, 8/9)$, and $h_1 = 4$.

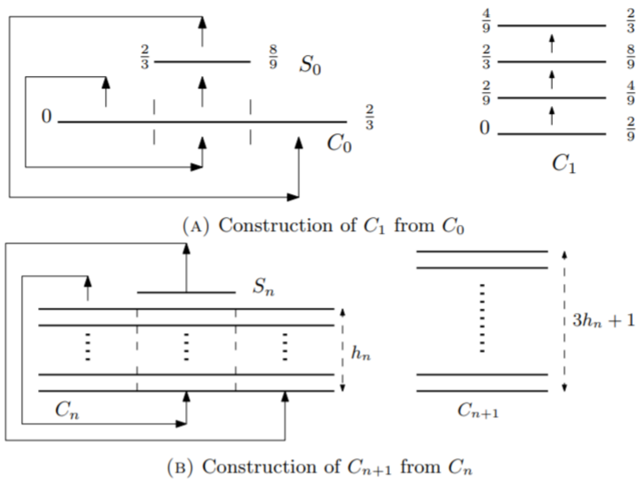
We obtain $C_{n+1} = C_n$. C_n has $(3^{n+1}-1)/2$ intervals, each of length $2 \cdot 3^{-n-1}$, and the union of each levels is $[0, 1-3^{-n-1}$. To obtain C_{n+1}, subdivide each interval equally into three and define spacer $S_n = [1-3^{-n-1}, 1-\frac{2}{3}3^{-n-1})$.

Now define T_{C_n} (for $n \geq 2$) be the map defined on $[0,1)$ by

$$T_{C_n}(x) = \begin{cases} x & \text{if } x \in [1-1/3^{n+1}, 1) \\ x - \frac{1}{2} - \frac{5}{6}\frac{1}{3^n} & \text{if } x \in S_n \\ x + \frac{1}{2} & \text{if } x \in I^{(n)}_{2 \cdot 3^{n-1}} \\ x - 1 + \frac{5}{3}\frac{1}{3^n} & \text{if } x \in I^{(n)}_{N(n)} \\ x + \cdots & \text{if otherwise} \end{cases}$$

(has to complete this... but for form is too complicated!!) That is, T_{C_n} is 'going up one level' where the levels are as displayed in the figure below.

Define T, the Chacón map as the limit of T_{C_n}.



(B) Construction of C_{n+1} from C_n

****

(1st November, Thursday)

(Will be showing that $([0,1), \mathcal{B}, \mu, T)$ is a measure preserving system in the example sheet 2.)

Lemma) Consider $([0,1), \mathcal{B}, \mu, T)$ where m is the Lebesgue measure and T is Chacón map. Then $\forall\, n, j$, we have

$$m(I^{(n)}_j \cap T^{N(n)} I^{(n)}_j) \geq \frac{1}{3} m(I^n_j)$$
$$m(I^{(n)}_j \cap T^{N(n)+1} I^{(n)}_j) \geq \frac{1}{3} m(I^n_j)$$

proof) Note

$$\begin{aligned} & I^{(n)}_j = I^{(n+1)}_j \cup I^{(n+1)}_{j+N(n)} \cup I^{n+1}_{j+2N(n)+1} \\ \text{and} \quad & T^{N(n)} I^{(n+1)}_j = I^{(n+1)}_{j+N(n)} \\ \text{and} \quad & T^{N(n)+1} I^{(n+1)}_{j+N(n)} = I^{(n+1)}_{j+2N(n)+1} \end{aligned}$$

Thus $I_j^{(n)} \cap T^{N(n)} I_j^{(n)} \supset I_{j+N(n)}^{(n+1)}$ and $I_j^{(n)} \cap T^{N(n)+1} I_j^{(n)} \supset I_{j+2N(n)+1}^{(n+1)}$.

(End of proof) □

Theorem) $([0,1), \mathcal{B}, \mu, T)$ is not mixing.

proof) Consider $A = I_1^{(2)}$ then $m(A) = 2/9$. For $n \geq 2$, A is the union of some intervals of the form $I_j^{(n)}$. Apply the lemma to each of these.

$$m(A \cap T^{N(n)} A) \geq \frac{1}{3} m(A) = \frac{3}{2} m(A)^2$$

this contradicts the definition of mixing.

(End of proof) □

Theorem) $([0,1), \mathcal{B}, \mu, T)$ is weak mixing.

proof) Let $f \in L^2$ such that $f \circ T = \lambda f$ a.e. We aim to show that f is constant a.e., then this implies that the system is weak mixing, by a theorem presented earlier. Note that $|\lambda| = 1$.

Fix $\epsilon > 0$ sufficiently small,

e.g. $\epsilon < 1/12$ and

$$m(x \in [0,1) : |f(x)| > 2) > 10\epsilon$$

This might require replacing f by a suitable constant multiple.

By Lousin's theorem, there is $h : [0,1] \to \mathbb{C}$ continuous such that $m(x \in [0,1) : f(x) \neq h(x)) < \epsilon$. Then if n is sufficiently large, then $|h(x_1) - h(x_2)| < \epsilon$ whenever $|x_1 - x_2| < \frac{2}{3^n}$, and in particular this holds when $x_1, x_2 \in I_j^{(n)}$ for some j. The intervals $I_j^{(n)}$ fill at least half of $[0,1)$. Therefore by pigeon hole principle, $\exists\ j$ s.t. $m(x \in I_j^{(n)} : f(x) \neq h(x)) < 2\epsilon m(I_j^{(n)})$. Let $z = f(x')$ for some $x' \in I_j^{(n)}$ such that $f(x') = h(x')$. Then

$$m(x \in I_j^{(n)} : |f(x) - \lambda^{N(n)} z| < \epsilon) \geq m\big(y \in I_j^{(n)} : T^{N(n)} y \in I_j^{(n)} \text{ and } f(y) = h(y)\big)$$

: denote $A = \{y \in I_j^{(n)} : T^{N(n)} y \in I_j^{(n)} \text{ and } f(y) = h(y)\}$, $B = \{x \in I_j^{(n)} : |f(x) - \lambda^{N(n)} z| < \epsilon\}$. If $y \in A$, then $x = T^{N(n)} y \in I_j^{(n)}$ and

$$f(x) = f(T^{(n)} y) = \lambda^{N(n)} f(y) = \lambda^{N(n)} h(y)$$

Since $|h(y) - z| < \epsilon$, $|f(x) - \lambda^{N(n)} z| < \epsilon$, so $x \in B$.

Hence

$$\begin{aligned} m(A) &\geq m(I_j^{(n)} \cap T^{N(n)} I_j^{(n)}) - m(y \in I_j^{(n)} : f(y) \neq h(y)) \\ &\geq \frac{1}{3} m(I_j^{(n)}) - 2\epsilon m(I_j^{(n)}) \end{aligned}$$

Since $\epsilon < 1/12$, has $\frac{1}{3} > 2\epsilon$, so $m(A) > 0$, i.e. $\exists\ x \in I_j^{(n)}$ such that

$$f(x) = h(x) \quad \text{and} \quad |f(x) - \lambda^{N(n)} z| < \epsilon$$

and therefore $|f(x) - z| < \epsilon$ (as $z = f(T^n(x')) = f(x')$ for some $x' \in I_j^{(n)}$), thus $|z - \lambda^{N(n)} z| < 2\epsilon$.

We will show in a minute that $|z| \geq 1$. If this is true, then

$$|1 - \lambda^{N(n)}| < 2\epsilon$$

and same argument using the second claim in the lemma gives

$$|1 - \lambda^{N(n)+1}| < 2\epsilon$$

Hence :

$$|\lambda^{N(n)} - \lambda^{N(n)+1}| < 4\epsilon \quad \Rightarrow \quad |1 - \lambda| < 4\epsilon$$

Taking $\epsilon \searrow 0<$ we get $\lambda = 1$.

We now have to show that $|z| \geq 1$.

Recall $m(x \in I_j^{(n)} : |f(x) - z| > \epsilon) < 2\epsilon m(I_j^{(n)})$. Using $T^{i-j} I_j^{(n)} = I_i^{(n)}$ we get

$$\begin{aligned} m(x \in I_j^{(n)} : |f(x) - \lambda^{i-j} z| \leq \epsilon) &\geq m(x \in [0,1) : |f(x)| > |z| + \epsilon) \\ &\leq \sum_{i=1}^{N(n)} 2\epsilon m(I_i^{(n)}) + m([0,1) \setminus \cup_i I_i^{(n)}) \leq 2\epsilon + \frac{2}{3^n} \end{aligned}$$

(3rd November, Saturday)

Recall :
$([0,1), \mathcal{B}, m, T)$ a Chacón map, $I_j^{(n)}$ intervals, $j = 1, \cdots, N(n)$. To show this is a weak mixing, we aimed to show $f \circ T = \lambda f$ a.e. implies f is constant a.e.

So far, we prove : $\forall \epsilon > 0$, $\forall n$ sufficiently large, $\exists z = z_{n,\epsilon}\mathbb{C}$ and $\exists j \in \{1, \cdots, N(n)\}$ such that

$$m(x \in I_j^{(n)} : |f(x) - z| > \epsilon) < 2\epsilon m(I_j^{(n)})$$

We also proved $\lambda = 1$, provided we can show $|z_{n,\epsilon} \geq 1|$ for any $\epsilon > 0$ and sufficiently large n. (Using $T^{j-i} I_j^{(n)} = I_i^{(n)}$ and $|f(T^{i-j}x)| = |f(x)|$)

Summing these up for $i = 1, \cdots, N(n)$, we get

$$m(x \in [0,1) : |f(x)| \leq |z| + \epsilon) \geq (1 - 2\epsilon) \sum_{i=1}^{N(n)} m(I_j^{(n)}) = (1 - 2\epsilon)(1 - 3^{-n})$$

If n is sufficiently large, then

$$m(x \in [0,1) : |f(x)| \leq 1 + \epsilon) \geq 1 - 10\epsilon$$

Comparing this with our ssumption on ϵ at the beginning of the proof, we get $|z| \geq 1$, so now we get $|z| = 1$. Thus $f(T^{i-j}x) = f(x)$.

The same argument using this instead of $|f(T^{i-j}x)| = |f(x)|$ gives

$$m(x \in [0,1) : |f(x) - z_{n,\epsilon}| \leq \epsilon) \geq (1 - 2\epsilon)(1 - 3^{-n}) \quad \cdots\cdots (\dagger)$$

Choose sequence $\epsilon_m \to 0$ and n_m
$rightarrow \infty$ such that z_{n_m, ϵ_m} converges to a complex number $\tilde{z}$(in fact, we have to first check that $z_{n,\epsilon}$ has a bounded subsequence), by Bozanno-Weierstrasss. If m is sufficiently large (depending on $\delta > 0$) then

$$m(x \in [0,1) : |f(x) - \tilde{z}| < \delta) > 1 - \delta$$

if we use $(\dagger)$ for n_m and ϵ_m.

10. Entropy

Bernoulli Shift : Let $(P_1, P_2, \cdots, P_d)$ be a probability vector. The $(P_1, \cdots, P_d)$-**Bernoulli shift** is the MPS

$$\big(\{1, \cdots, d\}^{\mathbb{Z}}, \mathcal{B}, \mu_{P_1,\cdots,P_d}, \sigma\big)$$

where $\mu_{P_1,\cdots,P_d}$ is the product measure with $(P_1, \cdots, P_d)$ on the coordinates and σ is the shift map.

Definition) The MPS $(X_1, \mathcal{B}_1, \mu_1, T_1)$ and $(X_2, \mathcal{B}_2, \mu_2, T_2)$ are called **isomorphic**, if there are maps

$$S_1 : X_1 \to X_2, \quad S_2 : X_2 \to X_1$$

such that $S_1 \circ S_2 = id_{X_2}$ a.e., $S_2 \circ S_1 = id_{X_1}$ a.e., $(S_1)_* \mu_1 = \mu_2$, $(S_2)_* \mu_2 = \mu_1$ and $T_1 \circ S_2 = T_1 \circ T_2$ a.e.

The Question : Are the $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ Bernoulli-shift isomorphic?

This question does not seem very difficult, but this had been unsolved for a long time. These two shifts have "the same" Koopman operators, and moreover Meshalkin proved that $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$-Bernoulli shifts are isomorphic. This problem was finally solved by Kolmogorov. He proved that these two systems are not isomorphic by attaching a quantity called *entropy*, which should be preserved by isomorphism, on each system and by showing they are not equal. Later, Ornstein showed that two Bernoulli shifts are isomorphic if and only if they have the same entropy. Actually, the introduction of notion of entropy in measure preserving systems was the starting point of ergodic theory being identified as an independent subject, so the importance of entropy in the field of ergodic theory cannot be overemphasize.

Let us do some actual mathematics now. We define the entropy as the measure of the amount how difficult it is to predict the system.

Definition) Let $(X, \mathcal{B}, \mu)$ be a probability space. A **countable measurable partition** is a collection of measurable sets $A_1, A_2, \cdots$ such that $A_i \cap A_j = \phi$ for all $i \neq j$ and $\bigcup A_i = X$. The sets A_i are called the atoms of partition.

The **join** or **coarsest common refinement** of two countable measurable partiotion ξ, η is

$$\xi \vee \eta = \{A \cap B : A \in \xi, B \in \eta\}$$

Define a function

$$H(p_1, \cdots, p_d) = -\sum_{j=1}^{d} p_j \log(p_j)$$

for all probability vector $(p_1, \cdots, p_d)$ with the convention $0 \cdot \log 0 = 0$. The **entropy** of a countable measurable partition ξ is

$$H_\mu(\xi) = H(\mu(A_1), \mu(A_2), \cdots)$$

where $\xi = \{A_1, A_2, \cdots\}$.

The **conditional entropy** of $\xi = \{A_1, A_2, \cdots\}$ relative to $\eta = \{B_1, B_2, \cdots\}$ is

$$H_\mu(\xi|\eta) = \sum_{n=1}^{\infty} \mu(B_n) \cdot H\Big(\frac{\mu(A_1 \cap B_n)}{\mu(B_n)}, \frac{\mu(A_2 \cap B_n)}{\mu(B_n)}, \cdots\Big)$$

This is the average average of entropy conditioned on each partition of

$\star$ Very Useful Interpretations :

Entropy of ξ provides the amount of information we can obtain from an experiment ξ. Conditional entropy of ξ relative to η provides the amount of information we can get from ξ given the information about experiment η. Entropy of join $\xi \vee \eta$ gives the amount of information we can get if we perform both experiments ξ and η.

Lemma)

(1) $H_\mu(\xi) \geq 0$.

(2) The value of $H_\mu(\xi)$ is maximal among partition ξ with k atoms if all atoms have the same measure $\frac{1}{k}$.

(3) $H_\mu(\{A_1, \cdots, A_k\}) = H_\mu(A_{\rho(1)}, \cdots, A_{\rho(k)})$ for all permutations $\rho \in \mathrm{Sym}(\{1, \cdots, k\})$.

(4) $H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\eta|\xi)$. This is called *chain rule.*

proof)

- (1) is trivial and (2) is going to be proved shortly using Jensen's inequality.
- (3) and (4) are going to be proved later, in more general setting.

Khinchin) Let $H_\cdot(\cdot) : P(X) \times \mathcal{B}$ be a function satisfying the conditions (1)-(4) of the lemma, where $P(X)$ is the set of Borel probability measures on $(X, \mathcal{B})$. Then $H_\cdot(\cdot)$ is uniquely determined by these properties, up to a multiplication of a scalar factor.

(6th November, Tuesday)

Definition) A function $[a, b] \to \mathbb{R} \cup \{\infty\}$ is **convex**, if $\forall x \in (a, b)$, $\exists \alpha_x \in \mathbb{R}$ such that

$$f(y) \geq f(x) + \alpha_x(y - x) \quad \forall y \in [a, b]$$

f is **strictly convex** if the equality occurs only for $x = y$.

Remark : If f is $C^2([a, b])$ and $f''(x) > 0$ for all $x \in (a, b)$, then f is strictly convex.

Jensen's inequality) Let $f : [a, b] \to \mathbb{R} \cup \{\infty\}$ be a convex function. Let $p_1, p_2, \cdots$ be a probability vector (possibly countably infinite). Let $x_1, x_2, \cdots \in [a, b]$. Then

$$f(p_1 x_1 + p_2 x_2 + \cdots) \leq \sum_i p_i f(x_i)$$

If f is strictly convex, then equality occurs *iff* those x_i for which $p_i > 0$ coincide.

Claim : Let $(X, \mathcal{B}, \mu)$ be a probability space. Let ξ be a measurable partition with k atoms. Then

$$H_\mu(\xi) \leq \log(k)$$

and equality occurs only if each atom of ξ has measures $\frac{1}{k}$.

proof) Apply Jensen's inequality to the function $x \mapsto x\log(x)$ with weights $p_i = \frac{1}{k}$ at the point $\mu(A_i)$, where A_i are the atoms of ξ. Note,

$$\begin{aligned} & \sum p_i \mu(A_i) = \frac{1}{k}\sum \mu(A_i) = \frac{1}{k} \\ \Rightarrow \quad & \frac{1}{k}\log(\frac{1}{k}) \leq \sum \frac{1}{k}\mu(A_i)\log(\mu(A_i)) \end{aligned}$$

so

$$\log k \geq \sum (-1)\mu(A_i)\log\mu(A_i) = H_\mu(\xi)$$

(End of proof) □

Definition) Let $(X, \mathcal{B}, \mu)$ be a probability space and let ξ be a countable measurable partition. The **information function** of ξ is

$$\begin{aligned} I_\mu(\xi) :& X \to \mathbb{R}\cup\{\infty\} \\ & x \mapsto -\log\mu\big([x]_\xi\big) \end{aligned}$$

where $[x]_\xi$ is the atom of ξ where x belongs.

If η is another partition, then the **conditional information function** of ξ relative to x is

$$I_\mu(\xi|\eta)(x) = -\log\frac{\mu\big([x]_{\xi\vee\eta}\big)}{\mu\big([x]_\eta\big)}$$

It is apparent that the information function is related to entropy. This is summarized in the following lemma.

Lemma) With notation as above,

$$\begin{aligned} H_\mu(\xi) &= \int I_\mu(\xi)d\mu \\ H_\mu(\xi|\eta) &= \int I_\mu(\xi|\eta)d\mu \end{aligned}$$

proof) The first equality is direct from the definition. For the second equality,

$$\begin{aligned} I_\mu(\xi|\eta)d\mu &= \sum_{A\in\xi, B\in\eta}\int_{A\cap B} I_\mu(\xi|\eta)d\mu = -\sum_{A\in\xi, B\in\eta}\mu(A\cap B)\log\Big(\frac{\mu(A\cap B)}{\mu(B)}\Big) \\ &= -\sum_{B\in\eta}\mu(B)\cdot\sum_{A\in\xi}\frac{\mu(A\cap B)}{\mu(B)}\log\Big(\frac{\mu(A\cap B)}{\mu(B)}\Big) \end{aligned}$$

(End of proof) □

One reason we use information function is that it is much easier to prove chain rule with information function.

Lemma) *(Chain rule)* Let $(X, \mathcal{B}, \mu)$ be a probability space and let ξ, η, λ be countable measurable partitions. Then

$$\begin{aligned} I_\mu(\xi\vee\eta|\lambda)(x) &= I_\mu(\xi|\lambda)(x) + I_\mu(\eta|\xi\vee\lambda)(x) \quad \forall x\in X \\ H_\mu(\xi\vee\eta|\lambda) &= H_\mu(\xi|\lambda) + H_\mu(\eta|\xi\vee\lambda) \end{aligned}$$

proof) For the first equality,

$$
\begin{aligned}
I_\mu(\xi \vee \eta | \lambda)(x) &= \log \frac{\mu([x]_\lambda)}{\mu([x]_{\xi\vee\eta\vee\lambda})} \\
I_\mu(\xi | \lambda)(x) &= \log \frac{\mu([x]_\lambda)}{\mu([x]_{\xi\vee\lambda})} \\
I_\mu(\eta | \lambda \vee \xi)(x) &= \log \frac{\mu([x]_{\xi\vee\lambda})}{\mu([x]_{\xi\vee\eta\vee\lambda})}
\end{aligned}
$$

and this proves the chain rule for information function.

The second equality follows from the first equality by integration the information function (as in the previous lemma).

(End of proof) □

The following inequality is very important in theory of mathematics of information.

Lemma) Let notation be as above. Then

$$H_\mu(\xi|\eta) \geq H_\mu(\xi|\eta \vee \lambda)$$

"The amount of information obtained from ξ given η is larger than information obtained from ξ given η and λ."

proof)

$$
\begin{aligned}
H_\mu(\xi|\eta\vee\lambda) &= \sum_{A\in\xi, B\in\eta, C\in\lambda} \mu(A\cap B\cap C) \log\left(\frac{\mu(B\cap C)}{\mu(A\cap B\cap C)}\right) \\
H_\mu(\xi|\eta) &= \sum_{A\in\xi, B\in\eta} \mu(A\cap B) \log\left(\frac{\mu(B)}{\mu(A\cap B)}\right)
\end{aligned}
$$

It is enough to show that for all fixed $A \in \xi$, $B \in \eta$, we have

$$\mu(A\cap B)\log\left(\frac{\mu(B)}{\mu(A\cap B)}\right) \geq \sum_{C\in\lambda} \mu(A\cap B\cap C)\log\left(\frac{\mu(B\cap C)}{\mu(A\cap B\cap C)}\right)$$

To see this, apply Jensen's inequality for $x \mapsto x\log x$ at points $\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}$ for $C \in \lambda$ with weights $\frac{\mu(B\cap C)}{\mu(B)}$. Write

$$\sum_{C\in\lambda} \frac{\mu(B\cap C)}{\mu(B)} \cdot \frac{\mu(A\cap B\cap C)}{\mu(B\cap C)} = \frac{1}{\mu(B)} \sum_{C\in\lambda} \mu(A\cap B\cap C) = \frac{\mu(A\cap B)}{\mu(B)}$$

and application of Jensen gives

$$\frac{\mu(A\cap B)}{\mu(B)} \cdot \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right) \leq \sum_{C\in\lambda} \frac{\mu(B\cap C)}{\mu(B)} \cdot \log\left(\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}\right)$$

and therefore

$$\mu(A\cap B) \cdot \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right) \leq \sum_{C\in\lambda} \mu(B\cap C) \cdot \log\left(\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}\right)$$

(End of proof) □

Corollary) $H_\mu(\xi) \le H_\mu(\xi \vee \eta) \le H_\mu(\xi) + H_\mu(\eta)$.

proof) Using the chain rule, obtain

$$H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\eta|\xi)$$

and form the previous lemma, has $H_\mu(\eta|\xi) \le H_\mu(\eta)$

(End of proof) □

(8th November, Thursday)

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let ξ, η be countable measurable partitions. Then :

$$\begin{aligned} I_\mu(T^{-1}\xi|T^{-1}\eta) = I_\mu(\xi|\eta)(Tx) \\ H_\mu(T^{-1}\xi|T^{-1}\eta) = H_\mu(\xi|\eta) \end{aligned}$$

where $T^{-1}\xi$ is the partition whose atoms are $T^{-1}([x]_\xi)$.

proof) Has

$$I_\mu(T^{-1}\xi|T^{-1}\eta)(x) = -\log\left(\frac{\mu([x]_{T^{-1}\xi\vee T^{-1}\eta})}{\mu([x])_{T^{-1}\eta}}\right)$$

Note

$$T^{-1}\xi \vee T^{-1}\eta = T^{-1}(\xi \vee \eta) \quad \text{and} \quad [x]_{T^{-1}\xi\vee T^{-1}\eta} = T^{-1}[Tx]_{\xi\vee\eta}$$

hence $\mu([x]_{T^{-1}\xi\vee T^{-1}\eta}) = \mu([Tx]_{\xi\vee\eta})$ by the measure preserving property. Similarly $\mu([x]_{T^{-1}\eta}) = \mu([Tx]_\eta)$. Then $I_\mu(T^{-1}\xi|T^{-1}\eta) = -\log\left(\frac{\mu([x]_{T^{-1}\xi\vee T^{-1}\eta})}{\mu([x])_{T^{-1}\eta}}\right) = I_\mu(\xi|\eta)(Tx)$

The statement on H_μ follows by integrating I_μ

(End of proof) □

Corollary) Writing $\xi_m^n = T^{-m}\xi \vee T^{-(m+1)}\xi \vee \cdots \vee T^{-n}\xi$, has

$$H_\mu(\xi_0^{n+m-1}) \le H_\mu(\xi_0^{n-1}) + H_\mu(\xi_0^{m-1})$$

proof) Note that $\xi_0^{n+m-1} = \xi_0^{n-1} \vee \xi_n^{n+m-1}$. So we have

$$\begin{aligned} H_\mu(\xi_0^{n+m-1}) &\le H_\mu(\xi_0^{n-1}) + H_\mu(\xi_n^{n+m-1}) = H_\mu(\xi_0^{n-1}) + H_\mu(T^{-n}\xi_0^{m-1}) \\ &= H_\mu(\xi_0^{n-1}) + H_\mu(\xi_0^{m-1}) \end{aligned}$$

where the last equality follows from the previous lemma.

(End of proof) □

Lemma) *(Felate's lemma)* Let $(a_n) \subset \mathbb{R}$ be a subadditive sequence, that is

$$a_{n+m} \le a_n + a_m \quad \forall n, m$$

Then $\lim_{n\to\infty} a_n/n$ exists and equals $\inf_n a_n/n$.

proof sketch) Need to show that $\limsup_{n\to\infty} \frac{a_n}{n} \le \frac{a_{n_0}}{n_0}$ for all n_0. For each fixed n_0, we can write $n = j(n)n_0 + i(n)$, where $i(n) \in [0, n_0 - 1]$. Iterate sub-additivity to get $a_n \le j(n)a_{n_0} + a_{i(n)}$.

See the online note for the full proof.

Definition) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let ξ, η be countable measurable partitions such that $H_\mu(\xi) < \infty$. The **entropy of the MPS w.r.t.** ξ is :

$$h_\mu(\xi) = \lim_{n\to\infty} \frac{H_\mu(\xi_0^{n-1})}{n} = \inf_n \frac{H_\mu(\xi_0^{n-1})}{n}$$

whose existence of the limit is guaranteed by *Felate's lemma.*(in fact, $\frac{H_\mu(\xi_0^{n-1})}{n}$ is a monotone decreasing sequence - will show in the example sheet)

The **entropy of the MPS** is $h_\mu(T) = \sup_{\xi: H_\mu(\xi)<\infty} h_\mu(T|\xi)$.

$h_\mu(\xi)$ expresses how fast we can learn information from a particular experiment ξ, and $h_\mu(T)$ is the maximal information we can obtain from the system when an appropriate experiment is chosen.

The problem of this definition is that it is generally difficult to find out the supremum $\sup_{\xi:H_\mu(\xi)<\infty} h_\mu(T|\xi)$ - since this requires computing entropy w.r.t ξ for each ξ. The good news is that (at least for the Bernoulli shifts), if we can find a partition that satisfies a particular property(so called **2-sided generator**), then in fact the supremum is achieved by the partition.

Definition) Let $(X, \mathcal{B}, \mu, T)$ be an invertible MPS. Let $\xi \subset \mathcal{B}$ be a countable measurable partitions. We say that ξ is a **2-sided generator** if $\forall A \in \mathcal{B}$ and $\forall \epsilon > 0$, $\exists k \in \mathbb{Z}_{>0}$ such that $\exists A' \in \sigma(\xi_{-k}^k)$ and $\mu(A \triangle A') < \epsilon$.

Theorem) *(Kolmogorov-Sinai)* Let $(X, \mathcal{B}, \mu, T)$ be an *invertible* measure preserving system. Let ξ be a countable measurable partition with $H_\mu(\xi) < \infty$, which is a 2-sided generator. Then

$$h_\mu(T) = h_\mu(T, \xi)$$

We delay the proof of this theorem until next lecture. Instead, we start to compute something useful.

Example : Let $(\{1, 2, \cdots, k\}^{\mathbb{Z}}, \mathcal{B}, \mu, \sigma)$ be the $(p_1, \cdots, p_k)$-Bernoulli shift. Let $X = \{1, 2, \cdots, k\}^{\mathbb{Z}}$.

- **Claim :** The partition $\xi = \{\{x \in X : x_0 = j\} : j = 1, \cdots, k\}$ is a 2-sided generator.

  proof) The collection of sets

  $$\{A \in \mathcal{B} : \forall \epsilon,\ \exists k\ \exists A' \in \xi_{-k}^k \text{ with } \mu(A \triangle A') < \epsilon\} \subset \sigma(\xi) \subset \mathcal{B}$$

  is a σ-algebra, and it contains cylinder sets. Hence it is equal to $\mathcal{B}$, as $\mathcal{B}$ is generated by cylinder sets.

  (End of proof) $\square$

- **Claim :** With ξ defined as above, we have

  $$H_\mu(\xi|\xi_1^n) = H(p_1, p_2, \cdots, p_k) = -p_1 \log p_1 - \cdots - p_k \log p_k$$

  for all $n \in \mathbb{Z}_{\ge 0}$.

proof) Calculate the information function :

$$I_\mu(\xi|\xi_1^n)(x) = \log\Big(\frac{\mu([x]_{\xi_1^n})}{\mu([x]_{\xi_0^n})}\Big)$$

Note $[x]_{\xi_0^n} = \{y \in X : y_0 = x_0, \cdots, y_n = x_n\}$, so $\mu([x]_{\xi_0^n}) = p_{x_0} \cdots p_{x_n}$. Similarly, has $\mu([x]_{\xi_1^n}) = p_{x_1} \cdots p_{x_n}$, and

$$I_\mu(\xi|\xi_1^n)(x) = -\log p_{x_0}$$

therefore $H_\mu(\xi|\xi_1^n) = \sum_{j=1}^k p_j(-\log(p_j)) = H(p_1, \cdots, p_k)$.

(End of proof) $\square$

- Hence

$$\begin{aligned} H_\mu(\xi_1^{n-1}) &= H_\mu(\xi_{n-1}^{n-1}) + H_\mu(\xi_{n-2}^{n-2}|\xi_{n-1}^{n-1}) + H_\mu(\xi_{n-3}^{n-3}|\xi_{n-2}^{n-1}) + \cdots + H_\mu(\xi|\xi_1^{n-1}) \quad \text{(Chain rule)} \\ &= H_\mu(\xi) + H(\xi|\xi_1^1) + \cdots + H_\mu(\xi|\xi_1^{n-1}) \quad \text{(invariance, first lemma of the day)} \\ &= nH(p_1, \cdots, p_k) \end{aligned}$$

Divide by n and take the limit,

$$h_\mu(T) = h_\mu(T, \xi) = H(p_1, \cdots, p_k)$$

So the entropy of $(1/2, 1/2)$ shift is $\log 2$ and $(1/3, 1/3, 1/3)$ shift is $\log 3$ - which shows that two systems cannot be isomorphic.

(10th November, Saturday)

Theorem) *(Kolmogorov-Sinai)* Let $(X, \mathcal{B}, \mu, T)$ be an *invertible* measure preserving system. Let ξ be a countable measurable partition with $H_\mu(\xi) < \infty$, which is a 2-sided generator. Then

$$h_\mu(T) = h_\mu(T, \xi)$$

We will need three lemmas.

Lemma 1) Let $(X, \mathcal{B}, \mu, T)$ be an *invertible* measure preserving system. Let $\xi \subset B$ be a countable partition. Then

$$h_\mu(T, \xi_{-n}^n) = h_\mu(T, \xi)$$

Lemma 2) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi, \eta \subset \mathcal{B}$ be two countable partitions. Then

$$h_\mu(T, \eta) \leq h_\mu(T, \xi) + H_\mu(\eta|\xi)$$

Lemma 3) For any $\epsilon > 0$ and $k \in \mathbb{Z}_{>0}$, $\exists \delta > 0$ such that the following holds : let $(X, \mathcal{B}, \mu)$ be a probability space. Let $\xi \subset \mathcal{B}$ be a countable and $\eta \subset \mathcal{B}$ a finite partition. Suppose that η has k atoms and for each $A \in \eta$, $\exists B \in \sigma(\xi)$ such that $\mu(A \triangle B) < \delta$. Then

$$H_\mu(\eta|\xi) \leq \epsilon$$

Let us prove the theorem assuming the lemmas. We will come hack to the proof of the lemmas later.

proof of Theorem) We first show $h_\mu(T,\xi) = \sup_{\eta \text{ finite}} h_\mu(T,\eta)$. We need to show that for all finite partition $\eta \subset \mathcal{B}$, we have $h_\mu(T,\eta) \leq h_\mu(T,\xi)$. Fix $\epsilon > 0$. By **Lemma 3** and definition of 2-sided generator, $\exists n$ such that $H_\mu(\eta|\xi_{-n}^n) \leq \epsilon$. Then we have

$$\begin{aligned} h_\mu(T,\eta) &\leq h_\mu(T,\xi_{-n}^n) + H_\mu(\eta|\xi_{-n}^n) \leq h_\mu(T,\xi_{-n}^n) + \epsilon \quad \text{(Lemma 2)} \\ &= h_\mu(T,\xi) + \epsilon \quad \text{(Lemma 1)} \end{aligned}$$

We are done if we take $\epsilon \searrow 0$.

Now it is left to show that $\forall \epsilon > 0$, for every countable partition $\eta \subset \mathcal{B}$, $\exists \tilde{\eta} \subset \mathcal{B}$ finite such that $h_\mu(T,\eta) \leq h_\mu(T,\tilde{\eta}) + \epsilon$. By **Lemma 2,** it is enough to show $H_\mu(\eta|\tilde{\eta}) \leq \epsilon$. When $\eta = \{A_1, A_2, \cdots\}$, let $\tilde{\eta} = (A_1, A_2, \cdots, A_n, \bigcup_{j>n} A_j)$ for sufficiently large n . Then

$$H_\mu(\eta) = H_\mu(\eta \vee \tilde{\eta}) = H_\mu(\tilde{\eta}) + H_\mu(\eta|\tilde{\eta})$$

Thus

$$\begin{aligned} H_\mu(\eta|\tilde{\eta}) = H_\mu(\eta) - H_\mu(\tilde{\eta}) &= \sum_{j>n} (-1)\mu(A_j) \log \mu(A_j) + \mu\Big(\cup_{j=n+1}^\infty A_j\Big) \log \mu\Big(\bigcup_{j=n+1}^\infty A_j\Big) \\ &\leq \sum_{j>n} (-1)\mu(A_j) \log \mu(A_j) \leq \epsilon \end{aligned}$$

if n is sufficiently large.

(End of proof) $\square$

Now we prove the lemmas.

proof of Lemma 1)

$$\begin{aligned} h_\mu(T,\xi_{-m}^m) &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_{-m}^{n+m-1}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n+2m-1}) \\ &= \lim_{n\to\infty} \frac{1}{n+2m-1} H_\mu(\xi_0^{n+2m-1}) = h_\mu(T,\xi) \end{aligned}$$

(End of proof) $\square$

proof of Lemma 2)

$$
\begin{aligned}
h_\mu(T,\eta) &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\eta_0^{n-1}) \\
&\le \lim_{n\to\infty} \frac{1}{n} H_\mu((\xi\vee\eta)_0^{n-1}) \\
&= \lim_{n\to\infty} \frac{1}{n} \Big(H_\mu(\xi_0^{n-1}) + H_\mu(\eta_0^{n-1}|\xi_0^{n-1}) \Big) \\
&= h_\mu(T,\xi) + \lim_{n\to\infty} \frac{1}{n} \Big(\sum_{j=0}^{n-1} H_\mu(\eta_j^1|\xi_0^{n-1}\vee\eta_{j+1}^{n-1}) \Big) \quad \text{(Chain rule)} \\
&= h_\mu(T,\xi) + \lim_{n\to\infty} \frac{1}{n} \Big(\sum_{j=0}^{n-1} H_\mu(\eta_j^j|\xi_j^j) \Big) \quad \text{(throwing away kwown information increases entropy} \\
&= h_\mu(T,\xi) + H_\mu(\eta|\xi)
\end{aligned}
$$

where the last equality follows because $H_\mu(\eta_j^j|\xi_j^j) = H_\mu(\eta|\xi)$.

(End of proof) $\square$

proof of Lemma 3) Write $A_1, \cdots, A_k$ for the atoms of η and for each $i \in \{1, \cdots, k\}$. Let $B_i \in \sigma(\xi)$ such that $\mu(A_i \triangle B_i) < \delta$. We consider the partition λ, which has the following $k+1$ atoms

: $C_0 = \bigcup_{i=1}^k (A_i \cap (B_i \backslash \bigcup_{j\neq i} B_j))$ and $C_i = A_i \backslash C_0$ for $i = 1, \cdots, k$.

They have some useful properties :

- C_0 is big, i.e. $\mu(C_0) \ge \sum_{i=1}^k (\mu(A_i) - k\delta) = 1 - k^2\delta$.
- If $x \in C_0$, then $x \in A_i \Leftrightarrow x \in B_i$.

Also note,

$$
\begin{aligned}
H_\mu(\lambda) &= -\mu(C_0)\log(\mu(C_0)) - \sum_{i=1}^k \mu(C_i)\log\mu(C_i) \\
&\le -\mu(C_0)\log(\mu(C_0)) - \sum_{i=1}^n \mu(C_i)\log\Big(\frac{\sum_{i=1}^k \mu(C_i)}{k}\Big) \quad \text{(Jensen to } x \mapsto x\log x)
\end{aligned}
$$

If δ is sufficiently small, then $\mu(C_0)$ can be made as close to 1 as we want and $\sum_{i=1}^k \mu(C_i)$ is as small as we want. Then $H_\mu(\lambda) < \epsilon$ if δ is sufficiently small. Now, we may write

$$
\begin{aligned}
H_\mu(\eta|\xi) &\le H_\mu(\eta\vee\lambda|\xi) \le H_\mu(\lambda|\xi) + H_\mu(\eta|\xi\vee\lambda) \\
&\le H_\mu + H_\mu(\eta|\xi\vee\lambda) < \epsilon + H_\mu(\eta|\xi\vee\lambda)
\end{aligned}
$$

so it is enough to show that

$$H_\mu(\eta|\xi\vee\lambda) = 0$$

However, this trivially holds because $\sigma(\eta) \subset \sigma(\xi\vee\lambda)$, i.e. $[x]_{\xi\vee\lambda} \subset [x]_\eta$.

: formally, this can be deduced from the fact

$$
\begin{aligned}
A_i &= C_i \cup (C_0 \cap A_i) = C_i \cup (A_i \cap (B_i \backslash \cup_{j\neq i} B_j)) \\
&= C_i \cup (C_0 \cap (B_i \backslash \cup_{j\neq i} B_j)) \in \sigma(\xi\vee\lambda)
\end{aligned}
$$

(End of proof) □

(13th November, Tuesday)

11. The Shannon McMillan Breiman theorem

Theorem) *(Shannon-McMillan-Breiman)* Let $(X, \mathcal{B}, \mu, T)$ be an ergodic MPS. Let $\xi \subset \mathcal{B}$ be a countable measurable partition. with $H_\mu(\xi) < \infty$. Then

$$\frac{1}{n} I(\xi_0^{n-1}) \xrightarrow{n\to\infty} h_\mu(T, \xi)$$

pointwise μ-a.e. and in L^1.

Recall : $I(\xi_0^{n-1})(x) = -\log \mu([x]_{\xi_0^{n-1}})$, so get $\mu([x]_{\xi_0^{n-1}}) \approx \exp(-nh_\mu(T,\xi))$ approximately.

A lecture and a half would be devoted in proving this theorem.

> **Idea :** Using the chain rule and then invariance,
>
> $$\begin{aligned} \frac{1}{n} I_\mu(\xi_0^{n-1})(x) &= \frac{1}{n} \sum_{j=0}^{n-1} I_\mu(\xi_j^j | \xi_{j+1}^{n-1})(x) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} I_\mu(\xi | \xi_1^{n-j-1})(T^j x) \end{aligned}$$
>
> so this looks as if we can apply pointwise ergodic theorem. However, we still the notion of conditional entropy to develop this idea.

Conditional expectation

Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. The for all $f \in L^1(X, \mathcal{B}, \mu)$, $\exists f^* \in L^1(X, \mathcal{B}, \mu)$ such that

(1) f^* is $\mathcal{A}$-measurable.

(2) $\int_A f d\mu = \int_A f^* d\mu$ for all $A \in \mathcal{A}$.

If f_1^* and f_2^* both satisfy (1) and (2) in the role of f^*, then $f_1^* = f_2^*$ μ-a.e. The function f^* is called the **condition expectation** of f relative to $\mathcal{A}$ and it is denoted by $\mathbb{E}[f|\mathcal{A}]$.

Remarks :

- Writing $V_\mathcal{A}$ for the closed subspace of $L^2(X, \mathcal{B}, \mu)$ consisting of $\mathcal{A}$-measurable functions, $\mathbb{E}[f|\mathcal{A}]$ is the orthogonal projection of f to $V_\mathcal{A}$ provided $f \in L^2$.
- If $\mathcal{A}$ is generated by a countable partition ξ, then

$$\mathbb{E}[f|\mathcal{A}](x) = \mu([x]_\xi)^{-1} \int_{[x]_\xi} f d\mu$$

Theorem) Let $(X, \mathcal{B}, \mu)$ be the probability space, let $f, f_1, f_2 \in L^1(X, \mathcal{B}, \mu)$, $\mathcal{A}, \mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{B}$ be sub-σ-algebra. Then

(1) $\mathbb{E}[f_1 + f_2|\mathcal{A}] = \mathbb{E}[f_1|\mathcal{A}] + \mathbb{E}f_2|\mathcal{A}$.

(2) If f_1 is $\mathcal{A}$-measurable, then $\mathbb{E}[f_1 f|\mathcal{A}] = f_1\mathbb{E}[f|\mathcal{A}]$.

(3) If $\mathcal{A}_1 \subset \mathcal{A}_2$ then $\mathbb{E}[f|\mathcal{A}_1] = \mathbb{E}[\mathbb{E}[\mathcal{A}|\mathcal{A}_2]|\mathcal{A}_1]$.

Theorem) *(Martingale theorems)* Let $(X, \mathcal{B}, \mu)$ be a probability space, let $f \in L^1$ and let $\mathcal{A}, \mathcal{A}_1, \mathcal{A}_2, \cdots \subset \mathcal{B}$ be sub-σ-algebras. Assume that *either* (1) $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$ and $\mathcal{A} = \sigma(\mathcal{A}_j : j \geq 1)$ *or* (2) $\mathcal{A}_1 \supset \mathcal{A}_2 \supset \cdots$ and $\mathcal{A} = \cap \mathcal{A}_j$. Then

$$\lim_{n\to\infty} \mathbb{E}[f|\mathcal{A}_n] = \mathbb{E}[f|\mathcal{A}]$$

both pointwise μ-a.e. and in L^1.

If you have not seen these theorems but too lazy to go over the proofs, you can see these easily for $f \in L^2$ case.

Conditional information and entropy and entropy relative to σ-algebra

Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\eta \subset \mathcal{B}$ be a countable partition let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then the **conditional information function of η relative to $\mathcal{A}$** is

$$I_\mu(\eta|\mathcal{A})(x) = \sum_{A\in\eta} -\chi_A \cdot \log\Big(\mathbb{E}[\chi_A|\mathcal{A}]\Big)$$

Example : Suppose that $\mathcal{A}$ is generated by a countable partition $\xi \subset \mathcal{B}$. Then

$$\begin{aligned}
I_\mu(\eta|\mathcal{A})(x) &= \sum_{A\in\eta} -\chi_A \log \mathbb{E}[\chi_A|\mathcal{A}](x) \\
&= -\log \mathbb{E}[\chi_{[x]_\eta}|\mathcal{A}](x) \\
&= -\log \frac{1}{\mu([x]_\xi)} \int_{[x]_\xi} \chi_{[x]_\eta} d\mu \\
&= -\log \frac{\mu([x]_\eta \cap [x]_\xi)}{\mu([x]_\xi)} = -\log \frac{[x]_{\eta\vee\xi}}{\mu([x]_\xi)}
\end{aligned}$$

Definition) The **conditional entropy of η relative to $\mathcal{A}$** is defined as $H_\mu(\eta|\mathcal{A}) = \int I_\mu(\eta|\mathcal{A})d\mu$.

Theorem) *(Maximal inequality)* Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $\xi_1, \xi_2, \cdots$ be countable measurable partitions such that $\sigma(\xi_1) \subset \sigma(\xi_2) \subset \cdots$. Let $\mathcal{A} = \sigma(\xi_j : j \geq 1)$. Let $\eta \subset \mathcal{B}$ be another countable partition. Then

$$I_\mu(\eta|\xi_n) \xrightarrow{n\to\infty} I_\mu(\eta|\mathcal{A}) \quad \text{pointwise a.e. and in } L^1$$

Moreover, I^* defined by $I^* = \sup_{n\in\mathbb{Z}_{>0}} I_\mu(\eta|\xi_n)$ is in $L^1(X, \mathcal{B}, \mu)$.

The theorem gives a useful tool to deal with $I_\mu(\eta|\mathcal{A})$ when $\mathcal{A}$ can be approximated by countable partitions.

(15th November, Thursday)

(??? Why are we using the different version of Maximal inequality?)

Theorem) *(Maximal inequality)* Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$ be a sequence of sub-σ-algebras of $\mathcal{B}$, and let $\eta \subset \mathcal{B}$ be a countable partition with $H_\mu(\eta) > \infty$. Then :

$$I_\mu(\eta|\mathcal{A}_n) \to I_\mu(\eta|\mathcal{A}) \quad \text{as } n \to \infty$$

pointwise μ-a.e. and in L^1, where $\mathcal{A} = \sigma(\cup_n \mathcal{A}_n)$.
Moreover, $I^*(x) = \sup_{n \in \mathbb{Z}_{>0}} I_\mu(\eta|\mathcal{A}_n)(x) \in L^1$.

proof for η finite) By definition and martingale convergence,

$$I_\mu(\eta|\mathcal{A}_n)(x) = -\log \mathbb{E}[\chi_{[x]_\eta}|\mathcal{A}_n](x) \xrightarrow{n\to\infty} -\log \mathbb{E}[\chi_{[x]_\eta}|\mathcal{A}](x) = I_\mu(\eta|\mathcal{A})(x)$$

This proves pointwise convergence. By dominated convergence, L^1 convergence follows if assuming $I^* \in L^1$.

Now it is enough to show $I^* \in L^1$

: fix $A \in \eta$, fix $\alpha > 0$. For $x \in X_1$, let $n(x)$ be defined by

$$n(x) = \min\{n : \log(\mathbb{E}[\chi_A|\mathcal{A}_n](x)) > \alpha\}$$

If the above does not hold for any n, the we write $n(x) = \infty$. Define $B_n = \{x \in X : n(x) = n\}$. Note that B_n is $\mathcal{A}_n$-measurable as we may write

$$B_n = \{x \in X : \mathbb{E}[\chi_A|\mathcal{A}_n](x) < \exp(-\alpha) \text{ but } \mathbb{E}[\chi_A|\mathcal{A}(x) \geq \exp(-\alpha)\ \forall j < n]\}$$

Then

$$\mu(B_n)\exp(-\alpha) \geq \int_{B_n} \mathbb{E}[\chi_A|\mathcal{A}_n] d\mu = \int_{B_n} \chi_A d\mu = \mu(B_n \cap A)$$

Define $A^* = \{x \in A : I^*(x) > \alpha\} \subset A \cap (\cup_n B_n)$. Then since B_n's are disjoint,

$$\mu(A^*) \leq \sum_{n=1}^{\infty} \mu(A \cap B_n) \leq \exp(-\alpha) \sum_{n=1}^{\infty} \mu(B_n) \leq \exp(-\alpha)$$

and summing these over all elements of η, we obtain

$$\mu\big(\{x \in X : I^*(x) > \alpha\}\big) \leq |\eta| \exp(-\alpha)$$

Then

$$\int I^*(x) d\mu \leq \sum_{n=1}^{\infty} \mu(x \in X : I^*(x) > n-1) \cdot n \leq \sum_{n=1}^{\infty} |\eta| \exp(-(n-1)) \cdot n < \infty$$

(End of proof) □

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi \subset \mathcal{B}$ be a countable partition with $H_\mu(\xi) < \infty$. Then

$$h_\mu(T, \xi) = \lim_{n\to\infty} H_\mu(\xi|\xi_1^n)$$

proof) By chain rule and invariance of T,

$$\frac{1}{n}H_\mu(\xi_0^{n-1}) = \frac{1}{n}\sum_{j=0}^{n-1} H_\mu(\xi|\xi_1^j)$$

hence

$$h_\mu(T,\xi) = C - \lim_{n\to\infty} H_\mu(\xi|\xi_1^n)$$

Observe that $H_\mu(\xi|\xi_1^n)$ is a monotone decreasing sequence, hence it converges, so in fact the Cesàro limit implies strong limit.

(End of proof) □

We are now ready to prove Shannon-McMillan-Breiman theorem.

Theorem) Let $(X,\mathcal{B},\mu,T)$ be an ergodic MPS. Let $\xi \subset \mathcal{B}$ be a countable partition with $H_\mu(\xi) < \infty$. Then

$$\frac{1}{N}I_\mu(\xi_0^{N-1})(x) \to h_\mu(\xi,T)$$

pointwise a.e. and in L^1.

proof)

$$\begin{aligned}\frac{1}{N}I_\mu(\xi_0^{N-1})(x) =& \frac{1}{N}\sum_{n=0}^{N-1} I_\mu(\xi|\xi_1^{N-n-1})(T^n x)\\ &+ \frac{1}{N}\sum_{n=0}^{N-1} I_\mu(\xi|\mathcal{B}(\xi_1^\infty))(T^n x)\\ &+ \frac{1}{N}\sum_{n=0}^{N-1}\Big(I_\mu(\xi|\xi_1^{N-n-1})(T^n x) - I_\mu(\xi|\mathcal{B}(\xi_1^\infty)(T^n x)\Big)\end{aligned}$$

where $\mathcal{B}(\xi_1^\infty)$ is the σ-algebra generated by $\bigcup_{n=1}^\infty \xi_1^n$.

By the pointwise ergodic theorem, has

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1} I_\mu(\xi|\mathcal{B}(\xi_1^\infty))(T^n x) \to& \int I_\mu(\xi|\mathcal{B}(\xi_1^\infty))d\mu\\ &= \lim_{n\to\infty}\int I_\mu(\xi|\xi_1^n)d\mu\\ &= \lim_{n\to\infty} H(\xi|\xi_1^n) = h_\mu(T,\xi)\end{aligned}$$

Define

$$I_K^*(x) = \sup_{k\geq K}\Big|I_\mu(\xi|\xi_1^k)(x) - I_\mu(\xi|\mathcal{B}(\xi_1^\infty))(x)\Big|$$

By the maximal inequality, $I_K^* \in L^1$ for all x and $I_K^*(x) \to 0$ for μ-a.e. x. I_K^* are pointwise monotone decreasing, so we have

$$\int I_K^* d\mu \to 0 \quad \text{as } K\to\infty$$

Now again by pointwise ergodic theorem,

$$\Big|\frac{1}{N}\sum_{n=0}^{N-1}\Big(I_\mu(\xi|\xi_1^{N-n-1})(T^n x) - I_\mu(\xi|\mathcal{B}(\xi_1^\infty)(T^n x)\Big)\Big|$$
$$\le \frac{1}{N}\sum_{n=0}^{N-K-1} I_K^*(T^n x) + \frac{1}{N}\sum_{n=N-K}^{N-1} I_0^*(T^n x) \xrightarrow{N\to\infty} \int I_K^* d\mu$$

as

$$\frac{1}{N}\sum_{n=N-K}^{N-1} I_0^*(T^n x) = \frac{1}{N}\sum_{n=0}^{N-1} I_0^*(T^n x) - \frac{N-K}{N}\frac{1}{N-K}\sum_{n=0}^{N-K-1} I_0^*(T^n x) \to \int I_0^* d\mu - \int I_0^* d\mu = 0$$

Hence,

$$\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\Big(I_\mu(\xi|\xi_1^{N-n-1})(T^n x) - I_\mu(\xi|\mathcal{B}(\xi_1^\infty)(T^n x)\Big)\Big| \le \int I_K^* d\mu$$

Since K was arbitrary, and $\int I_K^* d\mu \xrightarrow{K\to\infty} = 0$, so

$$\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\Big(I_\mu(\xi|\xi_1^{N-n-1})(T^n x) - I_\mu(\xi|\mathcal{B}(\xi_1^\infty)(T^n x)\Big)\Big| = 0$$

so we have pointwise convergence.

Moreover, if we observe carefully, we have L^1 convergence at each line of the proof, so we also have the L^1 convergence.

(End of proof) □

(17th November, Saturday)

12. Mixing and Entropy

Definition) An MPS $(X, \mathcal{B}, \mu, T)$ is called **K(olmogorov)-mixing** if $\forall \epsilon > 0$, $A \in \mathcal{B}$ and $\xi \subset \mathcal{B}$ finite partition, $\exists N \in \mathbb{Z}_{>0}$ such that : $\forall B \in \sigma(\xi_N^\infty)$ we have

$$|\mu(A \cap B) - \mu(A)\mu(B)| < \epsilon$$

Invertible MPS's that are K-mixing are called **K-automorphism**.(this is more common notion in literatures)

From now on in the course, we are only using finite partitions. Some results also holds for countable partitions, but we do not need such stronger results.

Remark : This implies mixing

: Let $A, C \in \mathcal{B}$ and consider $\xi = \{C, X\backslash C\}$. Then for N sufficiently large, we can take $B = T^{-N}(C)$ in the definition.

Definition) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi \subset \mathcal{B}$ be a finite partition. Then the **tail σ-algebra of** ξ is

$$\tau(\xi) = \bigcap_{n=1}^{\infty} \sigma(\xi_n^\infty)$$

Theorem) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. The following are equivalent.

(1) $(X, \mathcal{B}, \mu, T)$ is a K-mixing.

(2) $\tau(\xi)$ is *trivial* for all finite partition $\xi \subset \mathcal{B}$. That is, $\forall A \in \tau(\xi)$, we have $\mu(A) \in \{0, 1\}$.

(3) $(X, \mathcal{B}, \mu, T)$ is of *totally positive entropy.* That is, $h_\mu(T, \xi) > 0$ for all partition $\xi \subset \mathcal{B}$ finite such that $H_\mu(\xi) > 0$.

-Note : Bernoulli shift is a K-mixing and so (2) implies the Kolmogorov 0-1 law.

proof of (1) $\Leftrightarrow$ (2))

(1) $\Rightarrow$ (2) Suppose (1) holds and let $A \in \tau(\xi)$. Then $A \in \sigma(\xi_n^\infty)$ for all n, by the definition of tail σ-algebra. For any $\epsilon > 0$, by K-mixing property,

$$|\mu(A \cap A) - \mu(A)\mu(A)| < \epsilon$$

so $\mu(A) = \mu(A)^2$. Hence $\mu(A) \in \{0, 1\}$.

(2) $\Rightarrow$ (1) Suppose that (2) holds. Let A, ξ, ϵ be as in the definition of K-mixing. Let N be sufficiently large and $B \in \mathcal{B}(\xi_N^\infty)$. Then

$$\mu(A \cap B) = \int_B \chi_A d\mu = \int_B \mathbb{E}[\chi_A | \sigma(\xi_N^\infty)] d\mu$$

By (backward) martingale convergence, if N is sufficiently large,

$$\left|\mu(A \cap B) - \int_B \mathbb{E}[\chi_A | \tau(\xi)] d\mu\right| < \epsilon$$

Claim : $\mathbb{E}[\chi_A | \tau(\xi)](x) = \mu(A)$ for μ-a.e. x.

: suppose the contrary. Letting $C_\epsilon = \{x \in X : \mathbb{E}[\chi_A | \tau()\xi] > \mu(A) + \epsilon\}$. Then $\exists \epsilon > 0$ such that $\mu(C_\epsilon) > 0$. Note $C \in \tau(\xi)$ so $\mu(C_\epsilon) = 1$. But then

$$\mu(A) = \int_C \mathbb{E}[\chi_A | \tau(\xi)] d\mu > \mu(A) + \epsilon$$

which is a contradiction.

From the claim, we see that $\int_B \mathbb{E}[\chi_A | \tau(\xi)] d\mu = \mu(A)\mu(B)$, and hence the desired result.

For the rest of the theorem, we need additional results.

Lemma) Let $\xi, \eta \subset \mathcal{B}$ be a finite partition, let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Suppose that $\exists$ a sequence of finite partitions $\lambda_1, \lambda_2, \cdots$ such that $\mathcal{A}$ is the smallest σ-algebra that contains the atoms of τ_n for all n. Then

(1) $I_\mu(T^{-1}\xi|T^{-1}\mathcal{A})(x) = I_\mu(\xi|\mathcal{A})(Tx)$.

(2) $H_\mu(T^{-1}\xi|T^{-1}\mathcal{A}) = H_\mu(\xi|\mathcal{A})$.

(3) $I_\mu(\xi \vee \eta|\mathcal{A}) = I_\mu(\xi|\mathcal{A}) + I_\mu(\eta|\xi \vee \mathcal{A})$.

(4) $H_\mu(\xi \vee \eta|\mathcal{A}) = H_\mu(\xi|\mathcal{A}) + H_\mu(\eta|\xi \vee \mathcal{A})$.

(5) $H_\mu(\xi|[A] \vee \eta) \leq H_\mu(\xi|\mathcal{A})$.

The assumption of existence of such $\lambda_1, \lambda_2 \cdots$ is not essential, but makes the proof much easier. However, the assumption is not very serious restriction. The proof is to be done on the example sheet. - if you like measure theory, try proving the general version.

Proposition) Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $\xi \subset \mathcal{B}$ be a finite partition. Let $\mathcal{A}_1 \supset \mathcal{A}_2 \supset \cdots$ be sub-σ-algebras of $\mathcal{B}$ and let $\mathcal{A} = \cap_{n=1}^\infty \mathcal{A}_n$. Then

$$I_\mu(\xi|\mathcal{A}_n) \to I_\mu(\xi|\mathcal{A}) \quad \mu\text{-a.e. and in } L^1$$

proof) We prove the stronger result

$$H_\mu(\xi|\mathcal{A}_n) \to H_\mu(\xi|\mathcal{A}) \quad \mu\text{-a.e. and in } L^1$$

This is a result of backward-martingale convergence and dominated convergence : we clearly have pointwise convergence by continuity of logarithm function. Also, each $H_\mu(\xi|\mathcal{A}_n)$ is dominated by $\log(|xi|)$, so we may apply dominated convergence to obtain L^1-convergence.

(End of proof) □

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS and let $\xi \subset \mathcal{B}$ be a finite partition. Then

$$h_\mu(T, \xi) = H_\mu(\xi|\sigma(\xi_1^\infty))$$

proof) We have already seen that

$$h_\mu(T, \xi) = \lim_{n\to\infty} H_\mu(\xi|\xi_1^n)$$

and this equals $H_\mu(\xi|\sigma(\xi_1^\infty))$ by the statement of *maximal inequality.*

(End of proof) □

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi \subset \mathcal{B}$ be a finite partition such that $h_\mu(T, \xi) = 0$. Then $H_\mu(\xi|\tau(\xi)) = 0$.

proof) From the previous lemma, we already know $H_\mu(\xi|\sigma(\xi_1^\infty)) = 0$. We show that $H_\mu(\xi|\sigma(\xi_n^\infty)) = 0$ for all n and then we are done by the previous proposition. To this end, we write

$$\begin{aligned}
H_\mu(\xi|\sigma(\xi_n^\infty)) &\leq H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) \\
&= \sum_{j=0}^{n-1} H_\mu(\xi_j^j|\sigma(\xi_{j+1}^\infty)) \quad \text{(chain rule)} \\
&= nH_\mu(\xi|\sigma(\xi_1^\infty)) = 0 \quad \text{(invariance)}
\end{aligned}$$

(End of proof) □

proof of (2) $\Rightarrow$ (3) in Theorem) We assume $\tau(\xi)$ is trivial, $h_\mu(T,\xi) = 0$ and we show $H_\mu(\xi) = 0$. By the lemma, we have $H_\mu(\xi|\tau(\xi)) = 0$. We have already seen that condition (2) implies (as in proof of (2) $\Rightarrow$ (1) of the theorem)

$$\mathbb{E}[\chi_A|\tau(\xi)](x) = \mu(A) \quad \mu\text{-a.e.} \quad \forall A \in \mathcal{B}$$

so

$$\begin{aligned} 0 = H_\mu(\xi|\tau(\xi)) = \int I_\mu(\xi|\tau(\xi))d\mu &= -\int \log \mathbb{E}\big[\chi_{[x]_\xi}|\tau(\xi)\big](x)d\mu(x) \\ &= -\int \log \mu([x]_\xi)d\mu(x) \\ &= -\sum_{A\in\xi} \log(\mu(A))\mu(A) = H_\mu(\xi) \end{aligned}$$

(20th November, Tuesday)

Left to prove : If $h_\mu(T,\xi) > 0$ for all $\xi \subset \mathcal{B}$ finite with $H_\mu(\xi) > 0$ then $\tau(\xi)$ is trivial $\forall \xi \subset \mathcal{B}$ finite.

Proposition) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi, \eta \subset \mathcal{B}$ be two finite partitions. Then

$$h_\mu(T,\xi) = H_\mu(\xi|\sigma(\xi_1^\infty) \vee \tau(\eta))$$

Note : This tells us that if the entropy is positive, then the tail-σ-algebra cannot be too big.

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi \subset \mathcal{B}$ be a finite partition. Then

$$h_\mu(T,\xi) = \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) \quad \forall n \in \mathbb{Z}_{>0}$$

proof) By chain rule and invariance :

$$\begin{aligned} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) &= \sum_{j=0}^{n-1} H_\mu(\xi_j^j|\sigma(\xi_j^\infty)) \\ &= nH_\mu(\xi|\sigma(\xi_1^\infty)) = nh_\mu(T,\xi) \end{aligned}$$

as claimed.

(End of proof) $\square$

The next lemma is bit harder, and where all the magic happens.

Lemma) Let $(X, \mathcal{B}, \mu, T)$ be an MPS. Let $\xi, \eta \subset \mathcal{B}$ be finite measurable partitions. Then

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty))$$

That is, "$\sigma(\eta_n^\infty)$ does not add to much information if n is sufficiently large".

proof) Observe that $\forall n \in \mathbb{Z}_{>0}$, has

$$\begin{aligned} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) &\le H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) \\ &= nh_\mu(T,\xi) \end{aligned}$$

by the last lemma so we already have the inequality in one direction.

Now we do the following : we choose two sequence a_n and b_n such that $a_n \le b_n$ for all n and

$$\lim_n \frac{1}{n}\Big[H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) + a_n\Big] = \lim_n \frac{1}{n}\Big[H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) + b_n\Big] \quad \cdots\cdots\cdots (\diamond)$$

Suppose to the contrary that

$$\liminf_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) < h_\mu(T,\xi) = \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty))$$

Then $a_n \le b_n$ and this contradicts our assumption $(\diamond)$ on a_n, b_n. This means that :

$$\begin{aligned} h_\mu(T,\xi) &\le \liminf_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) \\ &\le \limsup_{n\to\infty} \frac{1}{n} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) \\ &\le h_\mu(T,\xi) \end{aligned}$$

so this prove the lemma, given the existence of a_n, b_n with the required property $(\diamond)$.

For the existence of such $(a_n), (b_n)$, we take

$$\begin{aligned} a_n &= H_\mu(\eta_0^{n-1}|\sigma(\xi_0^\infty) \vee \sigma(\eta_n^\infty)) \\ b_n &= H_\mu(\eta_0^{n-1}|\sigma(\xi_0^\infty)) \end{aligned}$$

Clearly, $a_n \le b_n$ and indeed

$$\begin{aligned} \textbf{(A)} &= H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) + a_n = H_\mu(\xi_0^{n-1} \vee \eta_0^{n-1}|\sigma(\xi_n^\infty) \vee \sigma(\eta_n^\infty)) \\ \textbf{(B)} &= H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty)) + b_n = H_\mu(\xi_0^{n-1} \vee \eta_0^{n-1}|\sigma(\xi_n^\infty)) \end{aligned}$$

By the last lemma,

$$\frac{\textbf{(A)}}{n} = h_\mu(T, \xi \vee \eta) \quad \forall n$$

and therefore

$$h_\mu(T,\xi\vee\eta) = \frac{\textbf{(A)}}{n} \le \frac{\textbf{(B)}}{n} \le \frac{H_\mu(\xi_0^{n-1} \vee \eta_0^{n-1})}{n} \xrightarrow{n\to\infty} h_\mu(T,\xi\vee\eta)$$

(End of proof) $\square$

proof of the proposition) Using the chain rule and invariance, we have

$$\begin{aligned} H_\mu(\xi_0^{n-1}|\sigma(\xi_n^\infty \vee \sigma(\eta_n^\infty))) &= \sum_{j=0}^{n-1} H_\mu(\xi_j^j|\sigma(\xi_{j+1}^\infty) \vee \sigma(\eta_n^\infty)) \\ &= \sum_{j=0}^{n-1} H_\mu(\xi|\sigma(\xi_1^\infty) \vee \sigma(\eta_{n-j}^\infty)) \end{aligned}$$

Therefore, it follows from the previous lemma that

$$h_\mu(\xi, T) = C - \lim_{n\to\infty} H_\mu(\xi|\sigma(\xi_1^\infty) \vee \sigma(\eta_n^\infty))$$

But by the the previous lemma, we already know

$$\lim_{n\to\infty} H_\mu(\xi|\sigma(\xi_1^\infty) \vee \sigma(\eta_n^\infty)) = H_\mu(\xi|\sigma(\xi_1^\infty) \vee \tau(\eta))$$

and therefore the claim follows.

(End of proof) □

We finally conclude the main theorem of the chapter.

proof of Theorem (3) ⇒ (2)) Suppose that $h_\mu(T, \xi) > 0$ for all $\xi \subset \mathcal{B}$ finite with $H_\mu(\xi) > 0$.

Suppose to the contrary that $\tau(\xi)$ is non-trivial for some $\xi \subset \mathcal{B}$ finite. Then $\exists A \subset \tau(\xi)$ s.t. $0 < \mu(A) < 1$. Define $\eta = \{A, X\backslash A\}$. Apply the proposition with the roles of ξ and η interchanged,

$$h_\mu(T, \eta) = H_\mu(\eta|\sigma(\eta_1^\infty) \vee \tau(\xi))$$

If we prove $h_\mu(T, \eta) = 0$, this gives a contradiction, as $H_\mu(\eta) > 0$.

: In fact, we will show $I_\mu(\eta|\sigma(\eta_1^\infty) \vee \tau(\xi))(x) = 0$ for a.e. x. Indeed, since $\eta \in \tau(\xi)$, has

$$\begin{aligned} I_\mu(\eta|\sigma(\eta_1^\infty) \vee \tau(\xi)) &= -\log \mathbb{E}[\chi_{[x]_\eta}|\sigma(\eta_1^\infty \vee \tau(\xi))] \\ &= -\log \chi_{[x]_\eta} = 0 \end{aligned}$$

(End of proof) □

13. Rudolph's Theorem

Theorem) *(Rudolph)* Let μ be a probability measure on $\mathbb{R}/\mathbb{Z}$ that is invariant and *ergodic* with respect to the joint action of $T_2 : x \mapsto 2x$ and $T_3 : x \mapsto 3x$. That is to say, if $A \subset \mathcal{B}$ is such that $T_2^{-1}A = A$ and $T_3^{-1}A = A$ then $\mu(A) \in \{0, 1\}$.

Suppose that $h_\mu(T_2) > 0$ or $h_\mu(T_3) > 0$. Then μ is the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$.

If we drop the assumption of positive entropy, then we might expect the measure to be Lebesgue measure or a measure supported on some set of rational numbers (this is a conjecture).

Theorem) *(Host)* Let μ be an *ergodic* T-invariant measure on $\mathbb{R}/\mathbb{Z}$ with $h_\mu(T_3) > 0$. Then μ-a.e. $x \in \mathbb{R}/\mathbb{Z}$ is **normal** in base 2, i.e. the sequence $T_2^n x$ is equi-distributed in $\mathbb{R}/\mathbb{Z}$.

Moreover, if μ is not necessarily ergodic, but the other assumptions holds, then

$$\mu(x \in \mathbb{R}/\mathbb{Z} : x \text{ is normal in base2}) \geq \frac{h_\mu(T_3)}{\log 3}$$

Notice this result does not depend on the measure.

(22nd November, Thursday)

Theorem) *(Host)* Let μ be an *ergodic* T-invariant measure on $\mathbb{R}/\mathbb{Z}$ with $h_\mu(T_3) > 0$. Then μ-a.e. $x \in \mathbb{R}/\mathbb{Z}$ is **normal** in base 2, i.e. the sequence $T_2^n x$ is equi-distributed in $\mathbb{R}/\mathbb{Z}$.

Moreover, if μ is not necessarily ergodic, but the other assumptions holds, then

$$\mu(x \in \mathbb{R}/\mathbb{Z} : x \text{ is normal in base2}) \geq \frac{h_\mu(T_3)}{\log 3}$$

Lemma) We have $\mathrm{ord}_{3^k} 2 = 2 \cdot 3^{k-1}$, where $\mathrm{ord}_{3^k}(2)$ is the order of 2 modulo 3^k. That is, the smallest integer n such that $3^k | 2^n - 1$.

proof) By induction on k, we prove that $\mathrm{ord}_{3^k} = 2 \cdot 3^{k-1}$ and $3^{k+1} \not| \, 2^{2\cdot 3^{k-1}} - 1$

: Easy to check for $k = 0, 1$. Assume that $k \geq 2$ and the claim holds for $k-1$. Then $3^k | 2^{\mathrm{ord}_{3^k}(2)} - 1$, so $3^{k-1} | 2^{\mathrm{ord}_{3^k}(2)} - 1$ hence $\mathrm{ord}_{3^k}(2) = a \cdot 2 \cdot 3^{k-2}$ for some $a \in \mathbb{Z}_{>0}$. We also know $2^{2\cdot 3^{k-2}} = b \cdot 3^{k-1} + 1$ for some $b \in \mathbb{Z}_{>0}$ and $3 \not| \, b$. Then by the binomial theorem we have

$$\begin{aligned} 2^{2\cdot 2\cdot 3^{k-2}} &= b^2 \cdot 3^{2(k-1)} + 2b3^{k-1} + 1 \\ 2^{3\cdot 2\cdot 3^{k-2}} &= b^3 \cdot 3^{3(k-1)} + 3b^2 \cdot 3^{2(k-1)} + 3 \cdot b \cdot 3^{k-1} \end{aligned}$$

From this we see that $3^k \not| \, 2^{2\cdot 2\cdot 3^{k-2}} - 1$ but $3^k | 2^{3\cdot 2\cdot 3^{k-2}} - 1$ and $3^{k+1} \not| \, 2^{3\cdot 2\cdot 3^{k-2}} - 1$.

(End of proof) $\square$

Observations : look at the T_2 orbit of $a/3^k$ where $a \in \mathbb{Z}$ and $3 \not| \, a$. This orbit is

$$\{\frac{b}{3^k} : b \in [0, \cdots, 3^k - 1], 3 \not| \, b\}$$

Also, $T_2^n(x + \frac{a}{3^k}) = T_2^n(x) + T_2^n(\frac{a}{3^k})$.

Another idea : If $h_\mu(T_3) > 0$, then μ will not "concentrate" on a few element of the set

$$\{x + \frac{a}{3^k} : a \in [0, \cdots, 3^k - 1], 3 \not| \, a\}$$

We need to show that for μ-a.e. x and for all $f \in C(\mathbb{R}/\mathbb{Z})$ we have

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_2^n x) \to \int f(x)dx$$

In fact, it is enough to prove this for functions of the form

$$f(x) = \exp(2\pi i m x), \quad m \in \mathbb{Z}$$

because linear combinations of trigonometric functions is dense in the set of continuous functions.

Notation : for $N \in \mathbb{Z}_{>0}$, let

$$P_{N,m}(x) = \frac{1}{N}\sum_{n=0}^{N-1} \exp(2\pi i m 2^n x)$$

Now it is enough to show : for all fixed $m \neq 0$ and for μ-a.e. x, we have $P_{N,m}(x) \to 0$.

Lemma) Fix $m \neq 0$ and $x \in \mathbb{R}/\mathbb{Z}$. Let α be the largest exponent such that $3^\alpha | m$. Let $N < 2 \cdot 3^{k-\alpha-1}$ for some $k \in \mathbb{Z}_{>0}$. Then

$$\sum_{a=0}^{3^k-1} \left|P_{N,m}(x + \frac{a}{3^k})\right|^2 = \frac{3^k}{N}$$

proof)

$$\begin{aligned}
\left|P_{N,m}(x)\right|^2 &= \left|\frac{1}{N}\sum_{n=0}^{N-1} \exp(2\pi i m 2^n x)\right|^2 \\
&= \frac{1}{N^2}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} \exp(2\pi i m 2^{n_1} x)\overline{\exp(2\pi i m 2^{n_2} x)} \\
&= \frac{1}{N^2}\sum_{n_1=0}^{N-1} \exp(2\pi i m (2^{n_1} - 2^{n_2}) x)
\end{aligned}$$

Note that for each $b \in \mathbb{Z}_{>0}$, :

$$\sum_{a=0}^{3^k-1} \exp(2\pi i b(x + \frac{a}{3^k})) = \exp(2\pi i b x)\sum_{a=0}^{3^k-1} \exp(2\pi i b a/3^k) \quad \cdots\cdots\cdots (\star)$$

If $3^k \nmid b$, then $(\star) = 0$. If $3^k | b$, then $(\star) = 3^k \exp(2\pi i b x)$.

When does it happen that $3^k | m \cdot (2^{n_1} - 2^{n_2})$?

Suppose $n_1 > n_2$. Then $m \cdot (2^{n_1} - 2^{n_2}) = m \cdot 2^{n_1}(2^{n_1 - n_2} - 1)$ Since $n_1 - n_2 \leq N < 2 \cdot 3^{k-\alpha-1}$, we have $3^{k-\alpha} \nmid 2^{n_1-n_2} - 1$. So, $3^k \nmid m \cdot (2^{n_1} - 2^{n_2})$ (recall α was chosen to be the largest s.t. $3^\alpha | m$). Same applies if $n_2 > n_1$. We get that $3^k | m(2^{n_1} - 2^{n_2})$ *iff* $n_1 = n_2$.

Therefore

$$|P_{N,m}(x)|^2 = \frac{1}{N^2}\sum_{n=0}^{N-1} 3^k = \frac{3^k}{N}$$

(End of proof) $\square$

(24th November, Saturday)

Proposition) Let μ be an ergodic T_3-invariant probability measure on $\mathbb{R}/\mathbb{Z}$. Let $k \in \mathbb{Z}_{>0}$. Write

$$\mu_k(A) = \sum_{a=0}^{3^k-1} \mu(A + \frac{a}{3^k}) \quad \forall A \in \mathcal{B}$$

Let $\xi = \{[0, \frac{1}{3}), [\frac{1}{3}, \frac{2}{3}), [\frac{2}{3}, 1)\}$. Let $\epsilon > 0$. Then

$$\mu_k([x]_{\xi_0^K}) \geq \exp((h_\mu(T_3) - \epsilon)k)\mu([x]_{\xi_0^K})$$

holds for μ-a.e. x, provided k is sufficiently large depending on ϵ, x and K is sufficiently large depending on ϵ, x, k.
"μ does not concentrate too much on a particular atom."

Remark : We may think of $[x]_{\xi_0^K}$ as

$$[x]_{\xi_0^K} = \{y \in \mathbb{R}/\mathbb{Z} = [0,1) : y = 0.y_1y_2\cdots_{(3)}, y_1 = x_2, \cdots, y_{K+1} = x_{K+1}\}$$

where $x = 0.x_1x_2\cdots_{(3)}$. This is an interval of length $3^{-(K+1)}$ containing x. Hence, whenever $K \geq k$,

$$\mu_k([x]_{\xi_0^K}) = \mu\Big(\bigcup_{a=0}^{3^k-1} [x]_{\xi_0^K} + \frac{a}{3^k}\Big)$$

Lemma) In the above setting, we have

$$\frac{1}{k} I_\mu(\xi_0^{k-1} | \sigma(\xi_k^\infty))(x) \to h_\mu(T_3) \quad \mu\text{-a.e.}$$

proof) Similar to Shannon-Mcmillan-Breiman, but much simpler. Using chain rule and invariance, one obtains, for μ-a.e. x,

$$\begin{aligned}
\frac{1}{k} I_\mu(\xi_0^{k-1} | \sigma(\xi_k^\infty))(x) =& \frac{1}{k}\sum_{j=0}^{k-1} I_\mu(\xi_j^j | \sigma(\xi_{j+1}^\infty))(x) \\
=& \frac{1}{k}\sum_{j=0}^{k-1} I_\mu(\xi | \sigma(\xi_1^\infty))(T^j x) \\
\xrightarrow{\text{by P.E.T}}& \int I_\mu(\xi|\sigma(\xi_1^\infty))d\mu = H_\mu(\xi|\sigma(\xi_1^\infty)) = h_\mu(T_3, \xi)
\end{aligned}$$

Proof will be complete with the following lemma.

Lemma) The partition ξ as in the setting is a 1-sided generator for $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, T_3)$. That is, $\forall A \in \mathcal{B}$, $\forall \epsilon > 0$, $\exists u \in \mathbb{Z}_{>0}$ and $B \in \sigma(\xi_0^n)$ such that $\mu(A \triangle B) < \epsilon$.

In particular, ξ is a 2-sided generator, and therefore $h_\mu(T, \xi) = h_\mu(T)$.

proof) Consider

$$\mathcal{C} = \{A \in \mathcal{B} : \forall \epsilon > 0, \exists n \in \mathbb{Z}_{>0}, B \in \sigma(\xi_0^n) \text{ s.t. } \mu(A \triangle B) < \epsilon\}$$

This is a σ-algebra, so it is enough to show that it contains all open sets.

Let $U \subset \mathbb{R}/\mathbb{Z}$ be open. We claim that

$$U = \bigcup_{x \in U, n \in \mathbb{Z}_{>0}, [x]_{\xi_0^n} \subset U} [x]_{\xi_0^n}$$

Note that this is a countable union and $[x]_{\xi_0^n} \in \mathcal{C}$ for all x, n. To see the claim, it is enough to show that $[x]_{\xi_0^n} \subset U$ for any $x \in U$ for n sufficiently large. But if n is sufficiently large, then $[x]_{\xi_0^n}$ is contained in an arbitrarily small neighbourhood of x.

(End of proof) □

proof of Proposition) By the lemma, if k is sufficienlty large depending on x, ϵ, then

$$I_\mu(\xi_0^{k-1}|\sigma(\xi_k^\infty)) \geq k(h_\mu(T_3) - \epsilon/2)$$

If K is sufficiently large, then

$$I_\mu(\xi_0^{k-1}|\xi_k^K)(x) \geq k(h_\mu(T_3) - \epsilon)$$

Observe : $[x]_{\xi_k^K} = \bigcup_{a\in 0,\cdots,3^k-1} \left([x]_{\xi_0^K} + \frac{a}{3^k}\right)$. Therefore $\mu_k([x]_{\xi_0^K}) = \mu([x]_{\xi_k^K})$. So

$$I_\mu(\xi_0^{k-1}|\xi_k^K)(x) = \log\left(\frac{\mu([x]_{\xi_k^K})}{\mu([x]_{\xi_0^K})}\right) = \log\left(\frac{\mu_k([x]_{\xi_0^K})}{\mu([x]_{\xi_0^K})}\right)$$

(End of proof) □

Corollary) Let μ be a T_3-invariant ergodic probability measure on $\mathbb{R}/\mathbb{Z}$ and μ_k be defined as in the previous proposition. Then we can find a family $(A_k)_{k\in\mathbb{Z}_{>0}} \subset \mathcal{B}$ such that

(1) μ-a.e. x is contained in A_k for k sufficiently large.

(2) $\mu_k(U) \geq \exp(kh_\mu(T_3)/2)\mu(A_k \cap U)$ holds for all open sets $U \subset \mathbb{R}/\mathbb{Z}$.

proof) Let A_k be the set of points x such that

$$\mu_k([x]_{\xi_0^K}) \geq \exp(kh_\mu(T_3)/2)\mu([x]_{\xi_0^K}) \quad \cdots\cdots\cdots (\divideontimes)$$

for all sufficiently large K. Then the first point (1) follows from the proposition.

To prove point (2), fix $k \in \mathbb{Z}_{>0}$ and fix $U \subset \mathbb{R}/\mathbb{Z}$ open. For $K \in \mathbb{Z}_{>k}$ write $\mathcal{A}_K$ for collection of atoms $[x]_{\xi_0^K}$ such that $[x]_{\xi_0^K} \subset U$ and $(\divideontimes)$ holds for this atom. Since U is open and by the definition of A_k, if $x \in A_k \cap U$ and K is sufficiently large, then $[x]_{\xi_0^K} \in \mathcal{A}_K$. Let $B_K = \bigcup_{A\in\mathcal{A}_K} A$. Then

$$\mu_k(B_K) \geq \exp(kh_\mu(T_3)/2)\mu(B_K)$$

follows by summing $(\divideontimes)$, and $\mu_k(B_K) \leq \mu_k(U)$. Now consider $\bigcap_{K\geq L} B_K$ for $L \in \mathbb{Z}_{>k}$. These sets increase as L grows, and their union is $A_k \cap U$. (Recall that $\forall x \in A_k$, we have $[x]_{\xi_0^K} \in \mathcal{A}_K$ for K sufficiently large. Therefore $x \in B_K$ if K is sufficiently large so $x \in \bigcap_{K\geq L} B_K$ if L is sufficiently large.) Moreover, since we have $\bigcap_{K\geq L} B_K \subset B_L$ for each L, the desired result follows.

(End of proof) □

(27th November, Tuesday)

(Last example class is 24th January, Thursday)

Theorem) *(Host)* Let μ be a T_3-invariant ergodic probability measure on $\mathbb{R}/\mathbb{Z}$ with $h_\mu(T_3) > 0$. Then μ-a.e. x is normal in base 2.

Without assuming ergodicity, we have

$$\mu(x \in \mathbb{R}/\mathbb{Z},\ x\text{is normal in base 2}) \geq \frac{h_\mu(T_3)}{\log 3}$$

We only prove the first part of the theorem, but use the second part of the theorem to deduce Rudolf's result.

Recall: $P_{N,m}(x) = \frac{1}{N}\sum_{n=0}^{N-1}\exp(2\pi i m 2^n x)$. We already reduced Host's theorem (in the ergodic case) to showing $P_{N,m}(x) \to 0$ for μ-a.e. x for all $m \neq 0$.

We also have prove the following results.

Lemma 1) Let α be the largest exponent such that $3^\alpha | m$ and let $N < 2 \cdot 3^{k-1-\alpha}$. Then

$$\sum_{a=0}^{3^k-1} \left| P_{N,m}(x + \frac{a}{3^k}) \right|^2 = \frac{3^k}{N}$$

Corollary) Let μ be a T_3-invariant, ergodic probability measure on $\mathbb{R}/\mathbb{Z}$. For $k \in \mathbb{Z}_{>0}$, we write $\mu_k(A) = \sum_{a=0}^{3^k-1}\mu(A + \frac{a}{3^k})$ for $A \in \mathcal{B}$. Then for all k, $A_k \in \mathcal{B}$ such that the following holds :

(1) For μ-a.e. x, we have $x \in A_k$ for all k sufficiently large.

(2) $\forall U \subset \mathbb{R}/\mathbb{Z}$ open, we have

$$\mu(U \cap A_k) \le \exp(-kh_\mu(T_3)/2)\mu_k(U)$$

Borel-Cantelli lemma would also be useful in our proof.

Lemma) *(Borel-Cantelli)* Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $B_1, B_2, \cdots \in \mathcal{B}$. Suppose

$$\sum_{n=1}^{\infty} \mu(B_n) < \infty$$

The for μ-a.e. x, $x \in B_n$ holds for at most finitely many n's.

To check $P_{N,m}(x) \to 0$, it is enough to check a cleverly chosen subsequence of $P_{N,m}(x)$ converges, since $P_{N,m}$ is given as an average, and hence cannot oscillate too fast. The following version of this claim would be useful in the proof of our theorem.

Lemma 2) $\forall m \neq 0$, and $c \in \mathbb{Z}_{>0}$, we have

$$P_{N,m}(x) \xrightarrow{N\to\infty} 0 \quad \Leftrightarrow \quad P_{L^c,m}(x) \xrightarrow{L\to\infty} 0$$

proof) See online notes.

We put these results together to prove the first part of Host's theorem.

proof of Host's theorem) By **Lemma 2**, it is enough to show that $P_{L^c(m)}(x) \to 0$ on $L \to \infty$ for some c (to be chosen later) and all $m \neq 0$. Fix $m \neq 0$ and $\epsilon > 0$ and let α be as in the statement of **Lemma 1**. For $N \in \mathbb{Z}_{>0}$, we set $k = k(N)$ to be an integer such that $2 \cdot 3^{k-2-\alpha} \le N < 2 \cdot 3^{k-1-\alpha}$. By **Lemma 1**, we have

$$\sum_{a=0}^{3^k-1} |P_{N,m}(x + \frac{a}{3^k})|^2 = \frac{3^k}{N} \le \frac{N}{2}\frac{3^{\alpha+2}}{N} = \frac{3^{\alpha+2}}{2}$$

By the definition of μ_k, we have

$$\int |P_{N,m}(x)|^2 d\mu_k(x) = \int \sum_{a=0}^{3^k-1} |P_{N,m}(x+\frac{a}{3^k})|^2 d\mu \leq \frac{3^{\alpha+2}}{2}$$

Using this estimate in Markov's inequality,

$$\mu_k(x : \left|P_{N,m}(x)\right| > \epsilon) \leq \frac{3^{\alpha+2}}{2\epsilon^2}$$

By the **Corollary**, with A_k set as in the **Corollary**, we have

$$\mu(x : \left|P_{N,m}(x)\right| > \epsilon, x \in A_k) \leq \exp(-kh_\mu(T_3)/2) \cdot \frac{3^{\alpha+2}}{2\epsilon^2}$$

Let $B_\epsilon = \{x : |P_{L^c,m}(x)| > \epsilon \text{ for infinitely many } L\text{'s}\}$. Our goal is equivalent to showing that $\mu(B_\epsilon) = 0$ for each $\epsilon > 0$.

Write $D_{L,\epsilon} = \{x : x \in A_{k(L^c)}, |P_{L^c,m}(x)| > \epsilon\}$. Note that if $x \in B_\epsilon$ then $x \in D_{L,\epsilon}$ for infinitely many L's and sufficiently large $k(L^c)$. So if we invoke *Borel-Cantelli lemma*, it is enough to show $\sum_{L=1}^{\infty} \mu(D_{L,\epsilon}) < \infty$. Recall we have

$$\mu(D_{L,\epsilon}) \leq \exp(-k(L^c)h_\mu(T_3)/2) \cdot \frac{3^{\alpha+2}}{2\epsilon^2}$$

We may choose c sufficiently large so that, then

$$k(L^c) > 4\log(L)/h_\mu(T_3)$$

Therefore,

$$\sum \mu(D_{L,\epsilon}) \leq \sum L^{-2} \frac{3^{\alpha+2}}{2\epsilon^2} < +\infty$$

(End of proof) □

We have not proved the second part of Host's theorem, but we will use it to prove Rudolph's theorem.

Theorem) *(Rudolph)* Let μ be a probability measure on $\mathbb{R}/\mathbb{Z}$ that is invariant and *ergodic* with respect to the joint action of $T_2 : x \mapsto 2x$ and $T_3 : x \mapsto 3x$. Suppose that $h_\mu(T_3) > 0$. Then μ is the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$.

proof) Write $A = \{x : \text{is normal in base } 2\}$. By *Host's theorem*, has $\mu(A) > 0$. First observe that $x \in A$ *iff* $T_2 x \in A$, so $A = T_2^{-1}A$.

Claim : if $x \in A$, then $T_3 x \in A$.

proof of claim) This claim is a consequence of the fact that T_2 and T_3 commute. If $x \in A$, then $\forall f \in C(X)$, by definition of being normal, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_2^n x) \to \int f(x)dx \quad \cdots\cdots\cdots (\heartsuit)$$

Now we can write

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_2^n T_3 x) = \frac{1}{N}\sum_{n=0}^{N-1} f(T_3 T_2^n x)$$
$$\xrightarrow{(\heartsuit)} \int f\circ T_3(x)dx = \int f(x)dx$$

The claim implies $A \subset T_3^{-1}A$. Since $\mu(A) = \mu(T_3^{-1}A)$, we have $\mu(A\triangle T_3^{-1}A) = 0$. Take

$$B = \bigcup_{k=0}^{\infty} T_3^{-k}A$$

then we have $T_2^{-1}B = B$, $T_3^{-1}B = B$ and $\mu(B\backslash A) = 0$. Using ergodicity assumption, we have $\mu(B) = 1$, and hence $\mu(A) = 1$. Therefore $(\heartsuit)$ holds for μ-a.e. x, i.e.

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_2^n x) \to \int f(x)dx \quad \forall f \in C(X) \text{ and for } \mu\text{-a.e. } x$$

Now by dominated convergence, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\int f(T_2^n x)d\mu = \int f(x)dx$$

But since μ was T_2-invariant, we have $\int f(T_2^n x)d\mu = \int f(x)d\mu$ and so

$$\int f d\mu = \int f(x)dx$$

for each continuous f. This proves μ is the Lebesgue measure.

(End of proof) $\square$