Part III – Topics in Ergodic Theory (Ongoing course, rough)

Based on lectures by Dr. P. Varjú
Notes taken by Bhavik Mehta

Michaelmas 2018

Contents

1 Measure preserving systems

Lecture 1 Ergodic theory is all about measure preserving systems.

Definition (Measure preserving system)**.** A **measure preserving system** $(X, \mathcal{B}, \mu, T)$ with X a set, $\mathcal{B}$ a σ-algebra, μ a probability measure ($\mu(A) \geq 0 \ \forall A \in \mathcal{B}$ and $\mu(X) = 1$) and T is a measure preserving transformation. Recall a **measure preserving transformation** $T : X \to X$ is a measurable function such that $\mu(T^{-1}(A)) = \mu(A) \ \forall A \in \mathcal{B}$.

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

Example. For example, consider a circle rotation. We have $X = \mathbb{R}/\mathbb{Z}$, $\mathcal{B}$ is the Borel sets, μ the Lebesgue measure, and $T = R_\alpha$, with $x \mapsto x + \alpha$ and $\alpha \in \mathbb{R}/\mathbb{Z}$ is a parameter.

We also have the 'times 2' or 'doubling' map, with the same $X, \mathcal{B}, \mu$ and $T = T_2$, $x \mapsto 2 \cdot x$.

Proof that T_2 is measure preserving. First check for intervals: Let $I = (a, b)$, then $\mu(I) = b - a$. Also, $\mu(T_2^{-1} I) = \mu\left((\frac{a}{2}, \frac{b}{2}) \cup (\frac{a}{2} + \frac{1}{2}, \frac{b}{2} + \frac{1}{2})\right) = \frac{b}{2} - \frac{a}{2} + \frac{b}{2} - \frac{a}{2} = b - a$, as required.

Now, let $U \subset \mathbb{R}/\mathbb{Z}$ be open. Then $U = I_1 \sqcup I_2 \sqcup \cdots$ is a disjoint union of intervals:

$$\begin{aligned} \mu(T^{-1}U) &= \mu\left(\bigcup T^{-1} I_j\right) \\ &= \sum \mu(T^{-1} I_j) \\ &= \sum \mu(I_j) \\ &= \mu(U). \end{aligned}$$

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

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

Now let $A \in \mathcal{B}$ be arbitrary. Let $\epsilon > 0$. $\exists U$ open and $\exists K$ compact such that $K \subset A \subset U$ and $\mu(U \setminus K) < \epsilon$.

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

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

The two examples generalise to the Haar measure on a topological group and to endomorphisms respectively.

In ergodic theory, we study the long term behaviour of orbits.

Definition (Orbit)**.** The orbit of $x \in X$ is the sequence

$$x, Tx, T^2x, \ldots$$

Some questions we might ask are:

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

Example. Let $A = [0, \frac{1}{4}) \subset \mathbb{R}/\mathbb{Z}$. Then $T_2^n x \in A \iff$ the $n+1$th and $n+2$th 'binary digits' of x are 0:

For $x = 0.x_1 x_2 x_3 \dots_2$, $x \in A$ corresponds to x_1, x_2 both being 0 and the doubling map sends x to $T_2 x = x_2 x_3 \dots_2$, showing the required property.

For example, $x = \frac{1}{6} = 0.00101010\dots_2$ starts in A but never comes back to A, but 'most points' do return to A. Also, we have $\mu(\{\, x \in A \mid T_2^n x \,\}) = \frac{1}{16}$ for any $n \ge 2$.

Example (Markov shift)**.** Let $(P_1, P_2, \dots, P_n)$ be a probability vector. Let $A \in \mathbb{R}_{\ge 0}^{n \times n}$ be the 'matrix of transition probabilities'. Assume

$$A \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \begin{pmatrix} P_1 & P_2 & \dots & P_n \end{pmatrix} A = \begin{pmatrix} P_1 & P_2 & \dots & P_n \end{pmatrix}$$

Take $X = \{1, \dots, n\}^{\mathbb{Z}}$, $\mathcal{B} =$ Borel σ-algebra generated by the product topology of the discrete topology on $\{1, \dots, n\}$, $T = \sigma$ the shift map: $(\sigma x)_m = x_{m+1}$. Finally, set the measure

$$\mu(\{\, x \in X \mid x_m = i_0, x_{m+1} = i_1, \dots, x_{m+n} = i_n \,\}) = P_{i_0} a_{i_0 i_1} \cdots a_{i_{n-1} i_n}.$$

On the example sheet, we will confirm this extends to a measure and that it is invariant under T.

2 Furstenberg correspondence

An interesting problem from number theory and combinatorics is Szemerédi's theorem. In this course, we will give an ergodic theory proof of this, using Furstenberg's multiple recurrence theorem.

Theorem (Szemerédi's theorem)**.** Let $S \subset \mathbb{Z}$ of positive upper Banach density. That is,

$$\bar{d}(S) \coloneqq \limsup_{N,M:M-N\to\infty} \frac{1}{M-N}\big|S \cap [N, M-1]\big|$$

and $\bar{d}(S) > 0$. Then S contains arbitrarily long arithmetic progressions. That is, $\forall l, \exists a \in \mathbb{Z}, d \in \mathbb{Z}_{>0}$,

$$a, a+d, \ldots, a+(l-1)d \in S.$$

Theorem (Furstenberg, multiple recurrence)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $A \in \mathcal{B}$ such that $\mu(A) > 0$. Let $l \in \mathbb{Z}_{>0}$. Then

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

Lecture 2

To prove Furstenberg's theorem requires more preparation, and we will do it later in this course. However, we can now prove Szemerédi assuming Furstenberg multiple recurrence.

Proof of Szemerédi using Furstenberg. Aim to construct a measure preserving system in order to use Furstenberg multiple recurrence, using the given set S.

Take $X = \{0,1\}^{\mathbb{Z}}$, $\mathcal{B} =$ Borel σ-algebra and $\sigma =$ the shift map $(\sigma(x))_n = x_{n+1}$. Let $x^S \in X$ be defined by

$$x^S_n = \begin{cases} 1 & n \in S \\ 0 & n \notin S \end{cases}$$

(so x^S is effectively the indicator function of S). Also let $A \in \mathcal{B}$ be given by $A = \{\, x \in X \mid x_0 = 1 \,\}$. Observe then that

$$n \in S \iff x^S_n = 1 \iff (\sigma^n x^S)_0 = 1 \iff \sigma^n x^S \in A.$$

The equivalence $n \in S \iff \sigma^n x^S \in A$ allows us to relate the set S to the dynamics of σ.

Now, we begin to construct the invariant measure μ. Let $\{M_m\}$ and $\{N_m\}$ be sequences s.t. $M_m - N_m \to \infty$ and

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

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

$$\delta_x(B) = \begin{cases} 1 & x \in B \\ 0 & x \notin B. \end{cases}$$

Let μ be the weak limit of a subsequence of μ_m. Note that the μ could be different dependent on subsequence choice.

2.1 Aside on weak limits

We pause the proof to briefly recap material on weak limits.

Definition (Weak limit)**.** 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 **converges weakly** to μ if for any $f \in C(X)$, we have

$$\lim_{n\to\infty} \int_X f \, d\mu_n = \int_X f \, d\mu.$$

We write $\text{lim-w}_{n\to\infty} \mu_n = \mu$.

Theorem (Banach-Alaoglu, or Helly)**.** Let X be a compact metric space. Then $\mathcal{M}(X)$, the set of Borel probability measures on X, endowed with the topology of weak convergence, is compact and metrizable. That is, there is a weakly convergent subsequence in any sequence of Borel probability measures.

We are ready to return to the main proof: next we check that the system constructed is indeed a measure preserving system.

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

Proof sketch. Let $B \in \mathcal{B}$. Then

$$\begin{aligned}
\mu_m(B) &= \frac{1}{M_m - N_m} \left|\{\, n \in [N_m, M_m - 1] \mid \sigma^n x^S \in B \,\}\right| \\
\mu_m(\sigma^{-1}B) &= \frac{1}{M_m - N_m} \left|\{\, n \in [N_m, M_m - 1] \mid \sigma^n x^S \in \sigma^{-1}B \,\}\right| \\
&= \frac{1}{M_m - N_m} \left|\{\, n \in [N_m + 1, M_m] \mid \sigma^n x^S \in B \,\}\right|
\end{aligned}$$

So the difference is such that

$$\left|\mu_m(B) - \mu_m(\sigma^{-1}B)\right| \leq \frac{1}{M_m - N_m} \to 0$$

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

Remark. If B is a cylinder set, i.e. $\exists L \in \mathbb{Z}_{>0}$ and $\tilde{B} \subseteq \{0,1\}^{2L+1}$ such that

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

then B is both closed and open. Therefore χ_B, the characteristic function of B is continuous. Hence $\lim_{n\to\infty} \mu_m(B) = \mu(B)$, since $\mu_m(B) = \int \chi_B \, d\mu_m$ and $\mu(B) = \int \chi_B \, d\mu$.

Approximating any Borel set by such cylinder sets would help complete the proof, but we in fact can get this result on spaces where χ is not continuous on a nice set of sets. So we leave full proof till a more general theorem.

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

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

Then S contains an arithmetic progression of length l.

Proof. Without loss of generality, we can assume $\mu = \text{lim-w}\,\mu_m$ - if not, pass to a subsequence. Let $B = A \cap \sigma^{-n} A \cap \cdots \cap \sigma^{-n(l-1)}(A)$. Observe that B is a cylinder set. Then by the remark, $\mu(B) = \lim \mu_m(B)$, hence $\exists m$ such that $\mu_m(B) > 0$.

By definition of μ_m, $\exists k \in [N_m, M_m - 1]$ such that $\sigma^k x^S \in B$. Hence

$$\sigma^k x^S \in A, \sigma^k x^S \in \sigma^{-n}(A), \ldots, \sigma^k x^S \in \sigma^{-n(l-1)}(A).$$

Thus, $k, k+n, \ldots, k+n(l-1) \in S$. □

Returning to the overall proof, we note A is a cylinder set. Then $\mu_m(A) \to \mu(A)$, i.e.

$$\mu(A) = \underbrace{\lim_{m\to\infty} \frac{1}{M_m - N_m} \left|\{\, n \in [N_m, M_m - 1] : n \in S \,\}\right|}_{\bar{d}(S)} > 0$$

where the inequality comes from satisfying the conditions of Szemerédi, and Furstenberg's multiple recurrence finishes the argument. □

3 Ergodicity

Lecture 3 **Lemma.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then $\exists n \in \mathbb{Z}_{>0}$ such that $\mu(A \cap T^{-n}A) > 0$.

Proof. Suppose for contradiction that $\mu(A \cap T^{-n}A) = 0$ for all $n > 0$. Then

$$\mu(T^{-k}A \cap T^{-n}A) = \mu(A \cap T^{-(n-k)}A) = 0$$

for all $n > k \geq 0$. Thus the sets $A, T^{-1}A, \dots$ are 'almost pairwise disjoint'.

Then

$$\begin{aligned}
\mu(A \cup T^{-1}A \cup \cdots \cup T^{-n}A) = \mu(A)& \\
&+ \mu(T^{-1}A) - \underbrace{\mu(T^{-1}A \cap A)}_{=0} \\
&+ \mu(T^{-2}A) - \underbrace{\mu(T^{-2}A \cap (A \cup T^{-1}A))}_{=0} \\
&\vdots \\
&+ \mu(T^{-n}A) - \underbrace{\mu(T^{-n}A \cap (A \cup T^{-1}A \cup \cdots \cup T^{-(n-1)}A))}_{=0} \\
&= (n+1)\mu(A),
\end{aligned}$$

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

Theorem (Poincaré recurrence)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then almost every $x \in A$ returns to A infinitely often. That is:

$$\mu\left(A \setminus \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A\right) = 0.$$

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

Proof. Let A_0 be the set of points in A that never return to A. We first show $\mu(A_0) = 0$. Note that

$$\mu(A_0 \cap T^{-n}A_0) \leq \mu(A_0 \cap T^{-n}A) = \mu(\varnothing) = 0$$

for all $n > 0$. By the lemma, $\mu(A_0) = 0$.

Note that if

$$x \in A \setminus \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A,$$

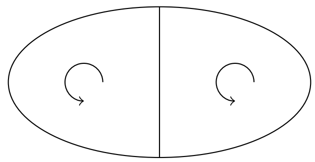
then there is a maximal $m \in \mathbb{Z}_{\geq 0}$ such that $T^m x \in A_0$. This means that

$$A \setminus \bigcap \bigcup T^{-n}A \subset \bigcup_{m=0}^{\infty} T^{-m}A_0$$

where the right hand side has measure 0. □

This effectively answers the first question we asked earlier - the orbit of almost every x visits A infinitely often if $\mu(A) > 0$.

But what if the point does not start in A? The main issue that can occur is that X splits into parts, which are preserved under T:



So, we define a notion of 'irreducible' for measure preserving systems.

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

If a measure preserving system is not ergodic, we have $A \in \mathcal{B}$ with $0 < \mu(A) < 1$ such that $T^{-1}A = A$, then we can restrict the measure preserving system to A. That is, we consider the measure preserving system $(A, \mathcal{B}_A, \mu_A, T|_A)$ where $\mathcal{B}_A = \{ B \in \mathcal{B} \mid B \subseteq A \}$ and $\mu_A(B) = \frac{\mu(B)}{\mu(A)}$ for all $B \in \mathcal{B}_A$.

Theorem. The following are equivalent for an measure preserving system $(X, \mathcal{B}, \mu, T)$.

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

(2) For all $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$\mu\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}A\right) = 1.$$

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

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

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

Proof. (1) $\Rightarrow$ (2). Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Let $B = \bigcap \bigcup T^{-n}A$. By Poincaré recurrence, $\mu(B) \ge \mu(A) > 0$. So if we show that $B = T^{-1}B$, then $\mu(B) = 1$ follows by ergodicity. But

$$x \in B \iff x \text{ visits } A \text{ infinitely often} \quad \iff Tx \text{ visits } A \text{ infinitely often} \quad \iff Tx \in B.$$

So $B = T^{-1}B$.

(2) $\Rightarrow$ (3). Let $A \in \mathcal{B}$ such that $\mu(A \triangle T^{-1}A) = 0$. If $\mu(A) = 0$, there is nothing to prove. Suppose $\mu(A) > 0$. Let $B = \bigcap \bigcup T^{-n}A$. By (2), we know that $\mu(B) = 1$.

We show $\mu(B \setminus A) = 0$, which completes the proof. Let $x \in B \setminus A$, then there is a first time m such that $T^m x \in A$, and $m > 0$. Hence $x \in T^{-m}A \setminus T^{-(m-1)}A$. Thus

$$B \setminus A \subseteq \bigcup_m T^{-m}A \setminus T^{-(m-1)}A,$$

and $\mu(T^{-m}A \setminus T^{-(m-1)}A) = \mu(T^{-1}A \setminus A) = 0$, so $\mu(B \setminus A) = 0$.

(3) $\Rightarrow$ (4). Let $f : X \to \mathbb{R}$ be a bounded measurable function such that $f = f \circ T$ almost everywhere. For all $t \in \mathbb{R}$, let $A_t = \{\, x \in X \mid f(x) \leq t \,\}$. Then $\mu(A_t \triangle T^{-1}A_t) = 0$. By (3), we have $\mu(A_t) = 0$ or 1 for all t.

If t is very small, then $\mu(A_t) = 0$. If t is very large, $\mu(A_t) = 1$. $t \mapsto \mu(A_t)$ is monotone, hence $\exists c \in \mathbb{R}$ such that $\mu(A_t) = 0$ for all $t < c$ and $\mu(A_t) = 1$ $\forall t > c$. Then $f(x) = c$ almost everywhere.

(4) $\Leftrightarrow$ (5) is left as an exercise.

(4) $\Rightarrow$ (1). Let $A \in \mathcal{B}$ with $A = T^{-1}A$. Then $\chi_A = \chi_A \circ T$ everywhere so χ_A is constant almost everywhere. $\square$

Example. The circle rotation map $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, R_\alpha)$ is ergodic iff α is irrational. Let $f : X \to \mathbb{R}$ be measurable. We can write $f(x) = \sum_{n\in\mathbb{Z}} a_n \exp(2\pi i n x)$.

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

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

4 Ergodic theorems

Lecture 4 In the previous section, we discussed recurrence, which is concerned with orbits visiting a particular set infinitely often. However, we have not addressed how frequent these visits are: the second question we asked earlier.

Theorem (Mean Ergodic Theorem, von Neumann)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system, and write

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

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

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

Theorem (Pointwise Ergodic Theorem, Birkhoff)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then for every $f \in L^1(X)$, there is a T-invariant function $f^* \in L^1(X)$ such that

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

When $f \in L^2(X)$, then the function f^* in the Pointwise Ergodic Theorem is $P_T f$. The Mean Ergodic Theorem also holds in L^p for $1 \leq p < \infty$, see the example sheet. We call the quantity $\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n$ the **ergodic average**.

Lemma. Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T : X \to X$ be a measurable transformation. Then T is measure preserving if and only if

$$\int_X f \circ T \, d\mu = \int_X f \, d\mu \tag{$*$}$$

for all $f \in L^1(X)$.

Proof. ($\Rightarrow$). Let $A \in \mathcal{B}$ and note that $x \in T^{-1}(A)$ is equivalent to $Tx \in A$. Hence we can write

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

so T is measure preserving.

($\Leftarrow$). As in ($\Rightarrow$), we can show ($*$) for characteristic functions. Then by linearity of integration, it also holds for simple functions. Now let $f \in L^1(X)$ be non-negative. Taking $f_1, f_2, \ldots$ an increasing sequence of simple functions with $f = \lim f_n$ almost everywhere. Then $f \circ T = \lim f_n \circ T$ almost everywhere (since T is measure preserving). Hence using the definition of integration and ($*$) for simple functions

$$\int f \circ T \, d\mu = \lim \int f_n \circ T \, d\mu = \lim \int f_n \, d\mu = \int f \, d\mu.$$

Finally, a general $f \in L^1(X)$ can be written as the difference of two non-negative functions, and conclude ($*$) by linearity. $\square$

Definition (Koopman operator)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. For $f : X \to \mathbb{C}$ a measurable function, define

$$U_T f = f \circ T$$

and call U_T the **Koopman operator**.

Lemma. The Koopman operator U_T is an isometry on the Hilbert space $L^2(X)$, i.e.

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

for all $f, g \in L^2(X)$.

Proof. Apply the previous lemma to $f \cdot \bar{g} \in L^1(X)$. □

Definition (Invertible)**.** A measure preserving system $(X, \mathcal{B}, \mu, T)$ is said to be **invertible** if there is a measure preserving map $S : X \to X$ such that $T \circ S = S \circ T = \mathrm{id}_X$ almost everywhere. If such a map exists, denote it as T^{-1}.

Example. The circle rotation is invertible, but the doubling map is not.

Lemma. If the measure preserving system $(X, \mathcal{B}, \mu, T)$ is invertible, then U_T is unitary on $L^2(X)$ and $U_T^* = U_{T^{-1}}$.

Proof. We use the earlier lemma to the function $f \cdot \overline{(g \circ T^{-1})}$:

$$\begin{aligned}
\langle f, U_{T^{-1}} g \rangle &= \int f \cdot \overline{g \circ T^{-1}} \, d\mu \\
&= \int (f \cdot \overline{(g \circ T^{-1})}) \circ T \, d\mu \\
&= \int f \circ T \cdot \bar{g} \, d\mu \\
&= \langle U_T f, g \rangle
\end{aligned}$$

This shows that $U_T^* = U_{T^{-1}}$. Clearly $U_T U_{T^{-1}} = U_{T^{-1}} U_T = \mathrm{id}_{L^2(X)}$, so U_T is unitary. □

The proofs we will give of the ergodic theorems rely on the idea that convergence is easy for certain special functions.

First, the ergodic averages of a T-invariant function $f \in I$ are equal to f, so they converge to f. If $f = U_T g - g$ for some g, then the ergodic averages become telescopic sums and only the boundary terms remain, that is:

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n (U_T g - g) = \frac{1}{N} (U_T^N g - g)$$

and the right hand side is easily seen to converge to 0. The following lemma shows that these two type of functions are enough to look at.

Lemma. Write

$$B \coloneqq \{ \, U_T g - g \mid g \in L^2(X) \, \} \, .$$

Then $I = B^\perp$.

It is important to note that the space B is not closed. So $L^2(X) = I \oplus \bar{B}$, but not every function in $L^2(X)$ is the sum of a T-invariant function and an element of B.

Proof. We can write

$$\begin{aligned} f \in B^\perp &\iff \langle f, U_T g - g\rangle = 0 \quad \forall g \in L^2(X) \\ &\iff \langle f, g\rangle = \langle f, U_T g\rangle = \langle U_T^* f, g\rangle \quad \forall g \in L^2(X) \\ &\iff f = U_T^* f. \end{aligned}$$

If the system was invertible, then we could finish the proof by applying $U_T = (U_T^*)^{-1}$ to both sides of the last equation.

But, we can prove that $f = U_T f \iff f = U_T^* f$ in the general case:

$$\begin{aligned} f = U_T f &\iff \langle f - U_T f, f - U_T f\rangle = 0 \\ &\iff \|f\|_2^2 + \|U_T f\|_2^2 - \langle f, U_T f\rangle - \langle U_T f, f\rangle = 0 \\ &\iff \|f\|_2^2 + \|U_T^*\|_2^2 - \langle U_T^* f, f\rangle - \langle f, U_T^* f\rangle + (\|U_T f\|_2^2 - \|U_T^* f\|_2^2) = 0 \\ &\iff \|f - U_T^* f\|_2^2 + (\|U_T f\|_2^2 - \|U_T^* f\|_2^2) = 0 \end{aligned}$$

Note that both terms in the left hand side of the final equation are non-negative, since $\|U_T f\|_2 = \|f\|_2$ since U_T is an isometry, and $\|U_T^* f\|_2 \le \|f\|_2$ as $\|U_T^*\| = \|U_T\| = 1$.

Thus,

$$f = U_T f \iff f = U_T^* f \text{ and } \|f\|_2 = \|U_T^* f\|_2.$$

The second statement of the right follows from the first one, hence $f = U_T f \iff f = U_T^* f$, as required. $\square$

Proof of Mean Ergodic Theorem. Let $f \in L^2(X)$ and fix $\epsilon > 0$. By the lemma, we can write $f = P_T f + U_T g - g + e$ for some $e, g \in L^2(X)$ such that $\|e\|_2 < \epsilon$. Thus

$$\limsup_{n\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f - P_T f \right\|_2 \le \limsup_{n\to\infty} \left\| \frac{1}{N} \left(U_T^N g - g\right) + \frac{1}{N} \sum_{n=0}^{N-1} U_T^n e \right\|_2 < \epsilon$$

and conclude by taking $\epsilon \to 0$. $\square$

5 Proof of pointwise ergodic theorem

Lecture 5 It is first useful to prove:

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

$$E_\alpha = \left\{ x \in X \;\middle|\; \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) > \alpha \right\}.$$

Then $\mu(E_\alpha) \le \alpha^{-1} \|f\|_1$.

First prove a useful proposition.

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

$$\begin{aligned} f_0 &= 0 \\ f_1 &= f \\ f_2 &= f \circ T + f \\ &\vdots \\ f_n &= f \circ T^{n-1} + \cdots + f \circ T + f \end{aligned}$$

and

$$F_N = \max_{n=0,\dots,N} f_n.$$

Then

$$\int_{E_N} f \, d\mu \ge 0 \quad \forall N.$$

where $E_N = \{ x \in X \mid F_N(x) > 0 \}$.

Proof. Suppose that $F_N(x) > 0$ for some x. Then $F_N(x) = f_n(x)$ for some $n \in \{1, \dots, N\}$. Then

$$\begin{aligned} F_N(x) = f_{n-1}(Tx) + f(x) &\le F_N(Tx) + f(x) \\ \implies f(x) &\ge F_N(x) - F_N(Tx) \\ \int_{E_N} f(x) \, d\mu &\ge \int_{E_N} (F_N(x) - F_N(Tx)) \, d\mu \end{aligned}$$

note if $F_N(x) \not< 0$, then $F_N(x) - F_N(Tx) \le 0$, so

$$\ge \int_X F_N(x) - F_N(Tx) \, d\mu = 0.$$

□

Proof of Maximal Ergodic Theorem. Define

$$\begin{aligned} E_{\alpha,M} &= \left\{ x \in X \;\middle|\; \max_{N=1,\dots,M} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) > \alpha \right\} \\ &= \left\{ x \in X \;\middle|\; \max_{N=1,\dots,M} \sum_{n=0}^{N-1} (f(T^n x) - \alpha) > 0 \right\}. \end{aligned}$$

We apply the proposition for the function $f-\alpha$. Then

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

Then

$$\int_{E_{\alpha,M}} f(x)\,d\mu \geq \alpha\mu(E_{\alpha,M})$$

and $\int_{E_{\alpha,M}} f(x)\,d\mu \leq \|f\|_1$. Note that $E_\alpha = \bigcup_M E_{\alpha,M}$, and this is an increasing union. $\square$

Proof of Pointwise Ergodic Theorem. Fix $\epsilon > 0$.

Using the density of $L^2 \cap L^1$ in L^1 (since simple functions are dense), we have $\exists f_\epsilon \in L^2$, $e_{\epsilon,1} \in L^1$ such that $f = f_\epsilon + e_{\epsilon_1}$, with $\|e_{\epsilon,1}\|_1 < \epsilon$.

Next, recall the subspace I from the Mean Ergodic Theorem, and the result that $L^2(X) = I \oplus \bar{B}$. Thus we can say $\exists g_\epsilon \in L^2$, $e_{\epsilon,2} \in L^1$ such that $f_\epsilon = P_T f_\epsilon + g_\epsilon \circ T - g_\epsilon + e_{\epsilon,2}$ and $\|e_{\epsilon,2}\|_1 \leq \|e_{\epsilon,2}\|_2 < \epsilon$ (L^1 norm is bounded by the L^2 norm in probability spaces).

Finally, $\exists h_\epsilon \in L^\infty$, $e_{\epsilon,3} \in L^1$ such that $g_\epsilon = h_\epsilon + e_{\epsilon,3}$ and $\|e_{\epsilon,3}\|_1 < \epsilon$ since L^∞ is dense in L^1, and g_ϵ is in L^1.

Thus, $f = P_T f_\epsilon + h_\epsilon \circ T - h_\epsilon + e_\epsilon$, where $e_\epsilon \in L^1$ with $\|e_\epsilon\|_1 < 4\epsilon$.

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

Let

$$E_{\epsilon,\alpha} = \left\{ x \in X \;\middle|\; \limsup_{N\to\infty} \left| \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) - P_T f_\epsilon(x) \right| > \alpha \right\}.$$

Applying the Maximal Ergodic Theorem for the function e_ϵ:

$$\mu(E_{\epsilon,\alpha}) \leq \alpha^{-1}\|e_\epsilon\|_1 \leq \frac{4\epsilon}{\alpha}.$$

Let F be the set of points x such that $\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$ does not converge at x. Then $F \subset \bigcup F_\alpha$ where

$$F_\alpha = \left\{ x \in X \;\middle|\; \limsup_{N_1,N_2\to\infty} \left| \frac{1}{N_1}\sum_{n=0}^{N_1-1} f(T^n x) - \frac{1}{N_2}\sum_{n=0}^{N_2-1} f(T^n x) \right| > 2\alpha \right\}.$$

Notice $F_\alpha \subset E_{\epsilon,\alpha}$ for all $\epsilon > 0$.

$$\mu(F_\alpha) \leq \mu(E_{\epsilon,\alpha}) \leq \frac{4\epsilon}{\alpha}.$$

Therefore $\mu(F_\alpha) = 0$.

We can take a countable sequence of α's and conclude $\mu(F) = 0$. We proved that $\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to f^*(x)$ for some function f^*.

By Fatou's lemma, $f^* \in L^1$. It remains to prove $f^*(x) = f^*(Tx)$ almost everywhere.

For almost every x,

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

Then $f^*(x) - f^*(Tx) = \lim \frac{1}{N} f(x) = 0$. $\square$

6 Unique ergodicity

Lecture 6 **Definition** (Normal)**.** A number $x \in [0,1)$ is called **normal** in base K, if for every $b_1, b_2, \dots, b_M \in \{0, \dots, K\}$, we have

$$\frac{1}{N} |\{ n \in \{0, \dots, N-1\} \mid x_{n+1} = b_1, \dots, x_{n+M} = b_M \}| \to \frac{1}{K^M}$$

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

Theorem. Almost every number (with respect to Lebesgue measure) is normal in every base $k \ge 2$.

Proof. Consider the measure preserving system $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, m, T_K)$ where $T_K(x) = K \cdot x$ and m refers to Lebesgue measure. On the example sheet, we show this is an ergodic measure preserving system. Now fix M and $b_1, \dots, b_M$ as in the definition and consider the set

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

Note: $T^n x \in A \iff x_{n+1} = b_1, \dots, x_{n+M} = b_M$. 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 $K, M, b_1, \dots, b_m$, the theorem follows. □

The Pointwise Ergodic Theorem as applied here shows that a property holds for almost every number, but gives no information about any specific numbers. Unique ergodicity attempts to ask when does the Pointwise Ergodic Theorem hold for *all* points.

If there is more than one T-invariant measure for a given transformation T, then we can apply the Pointwise Ergodic Theorem for both measures, which can give different limits for the ergodic averages. (This is not a contradiction, because the set where convergence holds in one case may be null with respect to the other measure). However, this suggests that *everywhere* convergence of ergodic averages prohibits the existence of multiple invariant measures.

Definition (Topological dynamical system)**.** 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.

Definition (Unique ergodicity)**.** We say that a topological dynamical system (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 following are equivalent:

(1) (X, T) is uniquely ergodic

(2) For every $f \in C(X)$, there is $c_f \in \mathbb{C}$ such that

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

uniformly on X.

(3) There is a dense $A \subset C(X)$ such that for each $f \in A$ there is $c_f \in \mathbb{C}$ such that

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

for all $x \in X$, not necessarily uniformly.

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 Borel measure $\mathcal{M}(X)$ on X to bounded linear functionals on $C(X)$.

Corollary. Let μ_1, μ_2 be Borel measures on a compact metric space. Then $\mu_1 = \mu_2$ iff

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

Definition (Push-forward measure)**.** Let (X, T) be a topological dynamical system, let μ be a Borel measure. The **push-forward** of μ via T is the measure

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

Lemma. Let X, T, μ be as above. Then

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

for every bounded measurable f.

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

$$\begin{aligned} \int \chi_A \, dT_*\mu &= T_*\mu(A) \\ &= \mu(T^{-1}A) \\ &= \int \chi_{T^{-1}A} \, d\mu \\ &= \int \chi_A \circ T \, d\mu. \end{aligned}$$

Using the standard argument, this can be extended to all bounded measurable f. □

Lemma. μ is T-invariant iff

$$\int f \, d\mu = \int f \circ T \, d\mu \quad \forall f \in C(X). \tag{$*$}$$

Proof. We have already seen that μ being T-invariant $\Rightarrow$ ($*$). For the other direction: Suppose that ($*$) holds. Then

$$\int f \, dT_*\mu = \int f \circ T \, d\mu = \int f \, d\mu \quad \forall f \in C(X).$$

By the corollary, we have $\mu = T_*\mu$. □

Theorem. Let (X,T) be a topological dynamical system. Let ν_j be a sequence of Borel probability measures on X. Let $\{N_j\} \subset \mathbb{Z}_{\geq 0}$ be a sequence such that $N_j \to \infty$. 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 $f \in C(X)$. Without loss of generality, assume

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

(if not, pass to a subsequence). Now,

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

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

$$\left| \int f \, d\mu - \int f \circ T \, d\mu \right| \leq \limsup_{j\to\infty} \frac{\|f\|_\infty + \|f \circ T\|_\infty}{N_j} = \limsup_{j\to\infty} \frac{\|f\|_\infty + \|f\|_\infty}{N_j} \to 0. \qquad \square$$

Lecture 7 Now, return to prove the equivalent conditions for unique ergodicity stated earlier in this section.

Proof. (1) $\Rightarrow$ (2). Suppose that (2) fails with $c_f = \int f \, d\mu$, where μ is the unique invariant measure. Then

$$\exists \epsilon > 0, \ \exists x_1, x_2, \dots X, \ \exists N_j \in \mathbb{Z} \text{ such that}$$
$$\left| \frac{1}{N_j} \sum_{n=0}^{N_j-1} f(T^n x_j) - \int f \, d\mu \right| > \epsilon. \qquad (*)$$

We may suppose that

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

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

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

for some probability measure ν. By the theorem earlier, 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 ($*$), $|a - \int f\,d\mu| > \epsilon$, so $\int f d\mu \neq \int f\,d\nu$, hence $\mu \neq \nu$, a contradiction.

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

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

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

Clearly (2) $\Rightarrow$ (3), closing the loop. □

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

$$\begin{aligned}\int \exp(2\pi i n x)\,d\mu &= \int \exp(2\pi i n R_\alpha(x))\,d\mu\\ &= \int \exp(2\pi i n(x+\alpha))\,d\mu\\ &= \exp(2\pi i n\alpha)\exp(2\pi i n x)\,d\mu.\end{aligned}$$

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

$$\int \exp(2\pi i n x)\,d\mu = 0 \quad \forall u \neq 0 \qquad \int 1\,d\mu = 1. \tag{$**$}$$

Let f be a trigonometric polynomial, that is a finite linear combination of the functions $\exp(2\pi inx)$ when $n \in \mathbb{Z}$. Then (∗∗) implies

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

Now use the fact that trigonometric polynomials are dense in $C(X)$: Stone-Weierstrass theorem.

Definition (Equidistributed)**.** A sequence $x_1, x_2, \ldots \in [0, 1)$ 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 $0 \le a < b < 1$. Then $x_1, x_2, \ldots$ is equidistributed if and only if

$$\frac{1}{N} |\{ n \in [0, N-1] \mid x_n \in [a, b] \}| \to b - a.$$

Corollary. $\{n\alpha + x\}$ is equidistributed for all α irrational $\forall x \in [0, 1]$.

Proof. This follows from the unique ergodicity of the circle rotation. □

7 Equidistribution of polynomials

Definition (Generic)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Suppose that X is a compact metric space, and $T : X \to X$ is continuous. Then $x \in X$ is called **generic** with respect to μ if the following holds

$$\frac{1}{N}\sum_{n=0}^{N-1} T_*^n \delta_x \xrightarrow{\text{weak}} \mu.$$

Equivalently,

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int f \, d\mu \quad \forall f \in C(X). \tag{$\dagger$}$$

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

Proof. By the Pointwise Ergodic Theorem, $\forall f \in C(X)$, $\exists X_f$ with $\mu(X_f) = 1$ such that ($\dagger$) holds. Observe that every point in $\bigcap_{f \in A} X_f$, where $A \subset C(X)$ is dense and countable is μ-generic. ☐

Lecture 8 **Theorem** (Furstenberg's skew-product theorem)**.** Let (X, T) be a uniquely ergodic topological dynamical system. Denote by μ the invariant measure. Write $S : X \times \mathbb{R}/\mathbb{Z} \to X \times \mathbb{R}/\mathbb{Z}$ defined by

$$S(x, y) = (Tx, y + c(x))$$

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

This result holds with any compact metrizable topological group in place of $\mathbb{R}/\mathbb{Z}$, called the skew-product construction.

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

$$\begin{aligned}
\iint f(S(x,y)) \, d\mu(x) \, dy &= \int_X \int_0^1 f(Tx, y + c(x)) \, dy \, d\mu(x) \\
&= \int_X \int_{-c(x)}^{1-c(x)} f(Tx, y) \, dy \, d\mu(x) \\
&= \int_{\mathbb{R}/\mathbb{Z}} \int_X f(Tx, y) \, d\mu(x) \, dy \\
&= \iint f(x, y) \, d\mu(x) \, dy.
\end{aligned}$$

So $\mu \times m$ is indeed S-invariant. Now we assume that $\mu \times m$ is S-ergodic. We show that $(X \times \mathbb{R}/\mathbb{Z}, S)$ is uniquely ergodic. Let E be the set of $\mu \times m$-generic points. We have seen that ergodicity implies $\mu \times m(E) = 1$.

Claim: If $(x, y) \in E$, then $(x, t + y) \in E$ for all $t \in \mathbb{R}/\mathbb{Z}$.

Proof: Observe $S \circ U_t = U_t \circ S$, where $U_t(x, y) = (x, t + y)$. Indeed:

$$\begin{aligned}
S \circ U_t(x, y) = S(x, t + y) &= (Tx, t + y + c(x)) \\
&= U_t(Tx, y + c(x)) = U_t \circ S(x, y).
\end{aligned}$$

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

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

using that $(x,y) \in E$

$$\to \iint f \circ U_t \, dm \, d\mu$$
$$= \iint f \, dm \, d\mu$$

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. Note $\mu(A) = \mu \times m(E) = 1$. Let ν be an S-invariant measure on $X \times \mathbb{R}/\mathbb{Z}$.

Next, aim to show $\nu(E) = 1$. 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 $B \subset X$ Borel.

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

So indeed $P_*\nu = \mu$. Then $P_*\nu(A) = \mu(A) = 1$, and $\nu(E) = \nu(A \times \mathbb{R}/\mathbb{Z}) = 1$.

Finally, we show that

$$\int f \, d\nu = \iint f \, d\mu \, dm \quad \forall f \in C(X \times \mathbb{R}/\mathbb{Z}),$$

and this proves $\nu = \mu \times m$. If $(x,y) \in E$, then

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

Since $\nu(E) = 1$, $(*)$ holds ν-almost everywhere. By dominated convergence,

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

So $\int f \, d\nu = \iint f \, d\mu \, dm$ indeed. □

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

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

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

Proof. By induction on d. $d = 1$ case is the circle rotation that has already been 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 (where m is Lebesgue measure).

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

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\mathbf{n}\in\mathbb{Z}^d} a_{\mathbf{n}} \exp(2\pi i \mathbf{n}\cdot x) \\ f(S\mathbf{x}) &= \sum_{\mathbf{n}\in\mathbb{Z}^d} a_{\mathbf{n}} \exp(2\pi i(n_1(x_1+\alpha)+n_2(x_2+x_1)+\cdots+n_d(x_d+x_{d-1}))) \\ &= \sum_{\mathbf{n}\in\mathbb{Z}^d} \exp(2\pi i n_1\alpha)\, a_{\mathbf{n}} \exp\Big(2\pi i((n_1+n_2)x_1+(n_2+n_3)x_2+\cdots \\ &\qquad\qquad\qquad\qquad + (n_{d-1}+n_d)x_{d-1}+n_d x_d)\Big) \\ &= \sum_{\mathbf{n}\in\mathbb{Z}^d} e^{2\pi i n_1\alpha} a_{\mathbf{n}} e^{2\pi i \hat{S}(\mathbf{n})\cdot\mathbf{x}}, \end{aligned}$$

where

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

Suppose $f = f\circ S$ almost everywhere. Then

$$a_{\hat{S}(\mathbf{n})} = \exp(2\pi i\alpha n_1) a_{\mathbf{n}},$$

in particular $|a_{\hat{S}(\mathbf{n})}| = |a_{\mathbf{n}}|$.

Suppose that $a_{\mathbf{m}} \neq 0$ for some $\mathbf{m}\in\mathbb{Z}^d$. We aim to show $\mathbf{m} = \mathbf{0}$, which implies that f is constant. By Parseval's formula,

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

This means that there are at most finite $\mathbf{n}$'s such that $|a_{\mathbf{m}}| = |a_{\mathbf{n}}|$. In particular, the orbit $\mathbf{m}, \hat{S}(\mathbf{m}), \hat{S}^2(\mathbf{m}), \dots$ must be periodic. Note

$$\left(\hat{S}^j(m)\right)_{d-1} = m_{d-1} + j m_d.$$

Thus $m_d = 0$. A similar argument gives $m_j = 0\ \forall j = 2, 3, \dots, d$. Then $\hat{S}^j(\mathbf{m}) = \mathbf{m}$. We have

$$a_{\mathbf{m}} = \exp(2\pi i\alpha m_1) a_{\mathbf{m}}.$$

Since $a_{\mathbf{m}} \neq 0$, we must have

$$\exp(2\pi i\alpha m_1) = 1.$$

As α is irrational, $m_1 = 0$. □

Lecture 9 **Theorem** (Weyl)**.** Let $P(x) = a_d x^d + \cdots + a_1 x + a_0$ be a polynomial such that a_j is irrational for at least one $j \neq 0$. Then the sequence $(\{P(n)\})_{n\ge\mathbb{Z}_{>0}}$ is equidistributed in $[0,1)$.

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

$$S^n \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} \binom{n}{0}x_1 + \binom{n}{1}\alpha \\ \binom{n}{0}x_2 + \binom{n}{1}x_1 + \binom{n}{2}\alpha \\ \vdots \\ \binom{n}{0}x_d + \binom{n}{1}x_{d-1} + \cdots + \binom{n}{d-1}x_1 + \binom{n}{d}\alpha \end{pmatrix}$$

This can be proved by induction on n. Consider the polynomials

$$q_j = \frac{t(t-1)\cdots(t-j+1)}{j!}.$$

The polynomials $q_0, q_1, \ldots, q_d$ form a basis in the vector space of polynomials of degree $\leq d$. In particular there are $x_1, x_2, \ldots, x_d, \alpha \in \mathbb{R}$ such that

$$p(t) = \alpha q_d(t) + x_1 q_{d-1}(t) + \ldots + x_d q_0(t), \tag{$*$}$$

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

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

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

Now we have

$$\frac{1}{N}\sum_{n=0}^{N-1} f(p(n)) = \frac{1}{N}\sum_{n=0}^{N-1} g(S^n(x_1, \ldots, x_d)) \tag{$**$}$$

where $x_1, \ldots, x_d$ are as in ($*$) and α in the definition of S is also coming from ($*$).

By unique ergodicity,

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

This proves equidistribution.

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

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

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

8 Mixing properties

Definition (Mixing)**.** An measure preserving system $(X, \mathcal{B}, \mu, T)$ is called **mixing** if

$$\forall A, B \in \mathcal{B},\ \forall \epsilon > 0,\ \exists N \text{ such that}$$
$$|\mu(A \cap T^{-n} B) - \mu(A)\mu(B)| < \epsilon \quad \forall n > N$$

Definition (Mixing on k sets)**.** A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is called **mixing!on k setsmixing on k sets** if $\forall A_0, A_1, \ldots, A_{k-1} \in \mathcal{B}$ and $\forall \epsilon > 0$, $\exists N$ such that

$$|\mu(A_0 \cap T^{-n_1} A_1 \cap \cdots \cap T^{-n_{k-1}} A_{k-1}) - \mu(A_0)\mu(A_1) \cdots \mu(A_{k-1})| < \epsilon$$

$\forall n_1, \ldots, n_{k-1}$ if $n_1 > N, n_2 - n_1 > N, \ldots, n_{k-1} - n_{k-2} > N$.

Open problem: Is there a measure preserving system that is mixing on 2 sets but not on 3 sets?

Definition (Density)**.** 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 a sequence of complex numbers a_n **converges in density** to $a \in \mathbb{C}$ if: $\{\, n \mid |a_n - a| < \epsilon \,\}$ has full density $\forall \epsilon > 0$. In notation, write $\text{D-lim}\, a_n = a$.

Definition (Cesàro convergence)**.** 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 this $\text{C-lim}\, a_n = a$.

Definition. A measure preserving system $(X, \mathcal{B}, \mu, T)$ is **weak mixing** if $\forall A, B \in \mathcal{B}$, we have

$$\underset{n \to \infty}{\text{D-lim}}\, \mu(A \cap T^{-n} B) = \mu(A)\mu(B).$$

Lecture 10 **Lemma.** Let $(a_n) \subset \mathbb{R}$ be a bounded sequence, let $a \in \mathbb{R}$. Then the following are equivalent.

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

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

(3) $\text{C-lim}_{n \to \infty} (a_n - a)^2 = 0$.

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

Proof.

- (1) $\Rightarrow$ (2). Fix $\epsilon > 0$. Let $M = \sup |a_n|$. Suppose that N is large enough such that

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

We estimate

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

M is constant, ϵ is arbitrary, so

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

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

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

By (2), we get $\frac{\epsilon}{N}\,|\{\, n\in[1,N] \mid |a_n - a| > \epsilon \,\}| \to 0$. This proves (1).

- The argument (1) $\Leftrightarrow$ (3) is the same.
- (1),(2) $\Rightarrow$ (4).

$$\left|\frac{1}{N}\sum_{n=1}^{N}(a_n - a)\right| \leq \frac{1}{N}\sum_{n=1}^{N}|a_n - a| \to 0$$

by (2). This implies C-lim $a_n = a$.
By the definition of D-lim:

$$\text{D-lim}\, a_n = a \implies \text{D-lim}\, a_n^2 = a^2$$

So using (1) and the implication (1) $\Rightarrow$ (3), C-lim $|a_n^2 - a^2| = 0$ hence C-lim $a_n^2 = a^2$.

- (4) $\Rightarrow$ (3).

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}(a_n - a)^2 &= \frac{1}{N}\sum_{n=1}^{N}a_n^2 + \frac{1}{N}\sum_{n=1}^{N}a^2 - 2\frac{1}{N}\sum_{n=1}^{N}a a_n\\ &\to a^2 + a^2 - 2a\cdot a = 0.\end{aligned}$$

□

Theorem. Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system. The following are equivalent

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

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

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

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

(5) U_T has no non-constant eigenfunctions, 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 a.e.

We will come back to prove this in a moment.

Lemma. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\mathcal{S} \subset \mathcal{B}$ be a semi-algebra (π-system) that generates $\mathcal{B}$. Then $(X, \mathcal{B}, \mu, T)$ is weak mixing iff

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

and $(X, \mathcal{B}, \mu, T)$ is ergodic iff

$$\operatorname*{C-lim}_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

Proof. See example sheet 2, question 2. □

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

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{n=1}^{N} [\mu \times \nu((T \times S)^{-n}(B_1 \times C_1) \cap B_2 \times C_2) - \mu \times \nu(B_1 \times C_1)\mu \times \nu(B_2 \times C_2)] \right| \\
=&\left| \frac{1}{N} \sum_{n=1}^{N} [\mu(T^{-n}B_1 \cap B_2)\nu(S^{-n}C_1 \cap C_2) - \mu(B_1)\mu(B_2)\nu(C_1)\nu(C_2)] \right| \\
\le&\frac{1}{N} \left| \sum_{n=1}^{N} \mu(T^{-n}B_1 \cap B_2)\nu(S^{-n}C_1 \cap C_2) - \mu(B_1)\mu(B_2)\nu(S^{-n}C_1 \cap C_2) \right| \\
&\qquad + \frac{1}{N} \left| \sum_{n=1}^{N} \left[\mu(B_1)\mu(B_2)\nu(S^{-n}C_1 \cap C_2) - \mu(B_1)\mu(B_2)\nu(C_1)\nu(C_2)\right] \right| \\
\le&\frac{1}{N} \sum_{n=1}^{N} |\mu(T^{-n}B_1 \cap B_2) - \mu(B_1)\mu(B_2)| \\
&\qquad + \frac{1}{N}\mu(B_1)\mu(B_2) \left| \sum_{n=1}^{N} \left[\nu(S^{-n}C_1 \cap C_2) - \nu(C_1)\nu(C_2)\right] \right| \to 0.
\end{aligned}$$

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

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

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

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

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

Then

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

This proves (1).

(1) $\Leftrightarrow$ (4) is the same arguments.

Lecture 11 (3) $\Rightarrow$ (5). Suppose that $f \circ T = \lambda \cdot f$ μ-a.e. Consider the function $\tilde{f}(x,y) = f(x) \cdot \overline{f(y)}$. We will show that $\tilde{f}$ is $T \times T$-invariant. If f is not constant, then $\tilde{f}$ is also not constant, so $(X \times X, \dots)$ is not ergodic.

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

Since U_T is an isometry,

$$\langle f, f \rangle = \langle U_T f, U_T f \rangle = \langle \lambda f, \lambda f \rangle = |\lambda|^2 \langle f, f \rangle.$$

So $|\lambda| = 1$, and $\tilde{f} \circ T \times T = \tilde{f}$ indeed.

Showing condition (5) implies weak mixing is more difficult, and is non-examinable. Two approaches are given in questions 7 and 8 of example sheet 2. $\square$

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

$$\underbrace{\frac{1}{N} \sum_{n=1}^{N} U_T^n f_1 \cdot U_T^{2n} f_2 \cdots U_T^{kn} f_k}_{g_N} \xrightarrow{L^2} \underbrace{\int f_1 \, d\mu \cdots \int f_k \, d\mu}_{\gamma}.$$

Corollary. Let notation be as above, and let $f_0 \in L^\infty$. Then

$$\frac{1}{N} \sum_{n=1}^{N} f_0 U_T^n f_1 \cdot U_T^{2n} f_2 \cdots U_T^{kn} f_k \longrightarrow \int f_0 \, d\mu \int f_1 \, d\mu \cdots \int f_k \, d\mu.$$

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

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

which gives Furstenberg's multiple recurrence for weak mixing systems.

Note the theorem says $g_n \to \gamma$ in L^2, and the corollary says $\langle f_0, \overline{g_n} \rangle \to \langle f_0, \overline{\gamma} \rangle$, so is certainly weaker.

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

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

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

$$\left\| \frac{1}{N} \sum_{n=1}^{N} u_n \right\| \to 0.$$

Proof idea.

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

□

Proof. Fix $\epsilon > 0$. Let H be such that

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

Write

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=1}^N u_n - \frac{1}{NH}\sum_{n=1}^N\sum_{h=1}^H u_{n+h}\right\| &\le \frac{1}{N}\left[\sum_{n=1}^H \|u_n\| + \sum_{n=N+1}^{N+H}\|u_n\|\right] \\ &< \epsilon\end{aligned}$$

if N is large enough. **unfinished** □

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

$$\operatorname*{D-lim}_{n\to\infty}\langle U_T^n f, g\rangle = \int f\,d\mu \cdot \int g\,d\mu.$$

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

Both these lemmas are proved on the example sheet.

Lecture 12 *Proof of theorem.* Induct on k. If $k = 1$, then this reduces to the Mean Ergodic Theorem.

Suppose that $k > 1$, and the theorem and corollary hold for $k-1$. We first consider the special case $\int f_k\,d\mu = 0$, and apply van der Corput's lemma for

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

It remains to check the conditions of van der Corput, so we compute s_k.

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

Apply the corollary for $k-1$ and $f_j U_T^{jk}\overline{f_j}$ in the role of f_{j-1}. Then we get:

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

$$\limsup_{N\to\infty}\frac{1}{N}\left|\sum_{k=1}^N\langle u_k, u_{n+k}\rangle\right| \le \|f_1\|_\infty^2 \cdots \|f_{k-1}\|_\infty^2 \cdot \left|\int f_k U_T^{kh}\overline{f_k}\,d\mu\right|$$

By an earlier lemma, (X, β, μ, T^k) is weak mixing. By the other lemma, we then have:

$$\text{D-lim} \int f_k U_T^{kh} \overline{f_k} \, d\mu = \int f_k \overline{f_k} \, d\mu = 0.$$

Therefore $\text{D-lim}\, s_k = 0$. Then van der Corput's lemma applies.

This completes the proof of the special case. General case: We can write $f_k = \int f_k \, d\mu + f'_k$, where $\int f'_k \, d\mu = 0$. Then

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{kn} f_k \\ &= \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{kn} f'_k}_{\to 0} + \int f_k \, d\mu \cdot \underbrace{\frac{1}{N} \sum_{k=0}^{N-1} U_T^n f_1 \cdots U_T^{(k-1)n} f_{k-1}}_{\to \int f_1 \, d\mu \cdots \int f_{k-1} \, d\mu} . \qquad \square \end{aligned}$$

8.1 Cutting and stacking

For each $n \in \mathbb{N}$, let $I_1^{(n)}, \ldots, I_{N(n)}^{(n)}$ 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

$$\mu\left([0,1) \setminus \bigcup_{j=1}^{N(n)} I_j^{(n)}\right) \to 0 \quad \text{as } n \to \infty.$$

(3)

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

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

Lemma. Given $I_j^{(n)}$ satisfying the above conditions, there is a unique (up to a set 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,

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

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

For almost every point x, eventually $x \in I_j^{(n)}$ with $j \neq N(n)$, then we know $I_j^{(n)} \to I_{j+1}^{(n)}$ is a translation, giving Tx.

Definition (Chacón map)**.** Let $I_1^{(1)} = [0, \frac{2}{3})$, $N(1) = 1$. Suppose that $I_j^{(n)}$ are defined for some n. Cut $I_j^{(n)}$ into 3 equal pieces. Define

- $I_j^{(n+1)}$: the piece on the left hand side
- $I_{j+N(n)}^{(n+1)}$: the piece in the middle
- $I_{j+2N(n)+1}^{(n+1)}$: the piece on the right hand side.
- $N(n+1) = 3N(n) + 1$.
- Finally, define $I_{2N(n)+1}^{(n+1)}$: cut off an interval of the same length as the other $I_j^{(n+1)}$'s from the remaining part of $[0, 1) \setminus \bigcup_{j=1}^{N(n)} I_j^{(n)}$, and let this be $I_{2N(n)+1}^{(n+1)}$.

Lecture 13 **Lemma.** Consider $([0, 1), \mathcal{B}, m, T)$ where m is the Lebesgue measure and T is the Chacón map. Then $\forall n, j$:

$$m\left(I_j^{(n)} \cap T^{N(n)} I_j^{(n)}\right) \geq \frac{1}{3} m\left(I_j^{(n)}\right)$$
$$m\left(I_j^{(n)} \cap T^{N(n)+1} I_j^{(n)}\right) \geq \frac{1}{3} m\left(I_j^{(n)}\right)$$

Proof. Note
$$I_j^{(n)} = I_j^{(n+1)} \cup I_{j+N(n)}^{(n+1)} \cup I_{j+2N(n)+1}^{(n+1)}$$
and $T^{N(n)} I_j^{(n+1)} = I_{j+N(n)}^{(n+1)}$ and $T^{N(n)+1} I_{j+N(n)}^{(n+1)} = I_{j+2N(n)+1}^{(n+1)}$. Thus $I_j^{(n)} \cap T^{N(n)} I_j^{(n)} \supseteq I_{j+N(n)}^{(n+1)}$ and $I_j^{(n)} \cap T^{N(n)+1} I_j^{(n)} \supseteq I_{j+2N(n)+1}^{(n+1)}$. □

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

Proof. Consider $A = I_1^{(2)}$ (recall $I_1^{(1)} = [0, \frac{2}{3})$), $m(A) = \frac{2}{9}$. 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. □

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

Proof. Let $f \in L^2$ such that $\lambda \circ T = \lambda f$ a.e. We aim to show that f is constant a.e. Note that $|\lambda| = 1$. Fix $\epsilon > 0$ sufficiently small: $\epsilon < \frac{1}{12}$, also

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

This might require replacing f by a suitable constant multiple.

By Luzin'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 $|x_1-x_2|<\frac{2}{3^n}$, 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 $\exists j$ such that $m(x \in I_j^{(n)} | f(x) \neq h(x)) < 2\epsilon m(I_j^{(n)})$. Let $z = f(x)$ for some $x \in I_j^{(n)}$ such that $f(x) = h(x)$.

$$\begin{aligned} m(\underbrace{x \in I_j : |f(x) - \lambda^{N(n)} z| < \epsilon}_{B}) &\geq m(\underbrace{y \in I_j^{(n)} : T^{N(n)} y \in I_j^{(n)} \text{and} f(y) = h(y)}_{A}) \\ &\geq m(I_j^{(n)} \cap T^{N(n)} I_j^{(n)}) - m(y \in I_j^{(n)} : f(y) \neq h(y)) \\ &\geq \frac{1}{3} m(I_j^{(n)}) - 2\epsilon m(I_j^{(n)}). \end{aligned}$$

(If $y \in A$, then $x = T^{N(n)} y \in I_j^{(n)}$ and $f(T^{N(n)}y) = \lambda^{N(n)} f(y) = \lambda^{N(n)} h(y)$, and the LHS is exactly $f(x)$. Since $|h(y) - z| < \epsilon$, $|f(x) - \lambda^{N(n)} z| < \epsilon$. So $x \in B$.)

We will show in a minute that $|z| \geq 1$. Then $|1 - \lambda^{N(n)}| < 2\epsilon$. Same argument using the second claim in the lemma gives

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

Hence $|\lambda^{N(n)} - \lambda^{N(n)+1}| < 4\epsilon$, and $|1-\lambda| < 4\epsilon$. Taking $\epsilon \searrow 0$, we get $\lambda = 1$.

Lecture 14 Recall:

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

Using $T^{i-j} I_j^{(n)} = I_i^{(n)}$, and

$$|f(T^{i-j}x)| = |f(x)| \tag{$*$}$$

we get

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

Summing for $i = 1, \dots, N(n)$, we get

$$\begin{aligned} m(x \in [0,1)) &\geq (1-2\epsilon) \sum_{i=1}^{N(n)} m(I_i^{(n)}) \\ &= (1-2\epsilon)(1 - \frac{1}{3^n}). \end{aligned}$$

If n is sufficiently large, then

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

comparing this with our assumption on ϵ at the beginning of the proof, we get $|z| \geq 1$, so $\lambda = 1$.

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

$$m(x \in [0,1) | f(x) - z_{n,\epsilon} \leq \epsilon) \geq (1-2\epsilon)(1 - \frac{1}{3^n}) \tag{$**$}$$

Choose sequences $\epsilon_m \to 0$, $n_m \to \infty$ such that z_{n_m, e_m} converges to a complex number $\tilde{z}$. (Possible by Bolzano-Weierstrass). 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 ($**$) for n_m and ϵ_m. $\square$

9 Entropy

Definition (Bernoulli shift)**.** Let $(p_1, p_2, \dots, p_d)$ be a probability vector. The $(p_1, \dots, p_d)$-**Bernoulli shift** is the measure preserving system

$$(\{1, \dots, d\}^{\mathbb{Z}}, \mathcal{B}, \mu_{p_1,\dots,p_d}, \sigma)$$

where σ is the shift map and $\mu_{p_1,\dots,p_d}$ is the product measure with $(p_1, \dots, p_d)$ on the coordinates.

This is just a special case of the Markov shift.

Definition (Isomorphic)**.** Measure preserving systems $(X_1, \mathcal{B}_1, \mu_1, T_1)$ and $(X_2, \mathcal{B}_2, \mu_2, T_2)$ are called **isomorphic** if there are maps

$$S_1 : X_1 \longrightarrow X_2$$
$$S_2 : X_2 \longrightarrow X_1$$

such that $S_1 \circ S_2 = \mathrm{id}_{X_2}$ and $S_2 \circ S_1 = \mathrm{id}_{X_1}$, the pushforwards $S_{1*}\mu_1 = \mu_2$, $S_{2*}\mu_2 = \mu_1$, and $T_1 \circ S_2 = S_1 \circ T_2$ almost everywhere.

Question: Are the $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ Bernoulli shifts isomorphic? This question was open for a long time, but was solved by Kolmogorov when he defined the notion of entropy.

- These two shifts have 'the same' Koopman operators, so attempts using operator theory are unlikely to succeed.
- After Kolmogorov's proof, Meshalkin proved $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ Bernoulli shifts are isomorphic. At first glance, these two shifts do not appear more different than the earlier two shifts, but these have the same entropy.
- Much later, Ornstein proved two systems have the same entropy if and only if they are isomorphic.

Roughly, the entropy measures how difficult it is to predict the next element of the orbit if you know a long segment.

Definition (Partition)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space. A countable measurable **partition** is a collection of measurable sets $A_1, A_2, \dots$ such that $A_i \cap A_j = \varnothing$ for all $i \neq j$ and $\bigcup A_i = X$. The sets A_i are called the **atoms** of the partition. The **join** or coarsest common refinement of two countable measurable partitions ξ, η is

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

Definition (Entropy)**.** Define

$$H(p_1, \dots, p_d) = -p_1 \log p_1 - \cdots - p_d \log p_d$$

for all probability vectors $(p_1, \dots, p_d)$ with the convention $0 \log 0 = 0$ (and extended to countable vectors also). The **entropy** of a countable measurable partition ξ is

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

where $\xi = \{ A_1, A_2, \dots \}$

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

$$H_\mu(\xi \mid \eta) = \sum_{n=1}^{\infty} \mu(B_n) \cdot H\left(\frac{\mu(A_1 \cap B_n)}{\mu(B_n)}, \frac{\mu(A_2 \cap B_n)}{\mu(B_n)}, \dots\right).$$

Lemma.

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

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

(3)
$$H_\mu(\{A_1, A_2, \dots, A_k\}) = H_\mu(A_{\rho_1}, \dots, A_{\rho_k})$$
for all permutations $\rho : \{1, \dots, k\} \to \{1, \dots, k\}$.

(4) $H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\eta \mid \xi)$, sometimes called the chain rule.

Lecture 15 **Definition** (Convex)**.** 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 equality occurs only for $x = y$.

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

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

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

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

Proof. The proof of Jensen's inequality is not part of this course, but is still worth reading. □

Proof of lemma. (1) is immediate by definition. For (2), let $(X, \mathcal{B}, \mu)$ be a probability space, let ξ be a measurable partition with k atoms. Apply Jensen's inequality to the function $x \mapsto x \log x$ with weights $p_i = \frac{1}{k}$ and the points $\mu(A_i)$ where A_i are the atoms of ξ. Note

$$\begin{aligned} \sum p_i \mu(A_i) &= \frac{1}{k} \sum \mu(A_i) = \frac{1}{k} \\ \frac{1}{k} \log \frac{1}{k} &\leq \sum \frac{1}{k} \mu(A_i) \log \mu(A_i) \\ \log k &\geq \sum -\mu(A_i) \log \mu(A_i) = H_\mu(\xi). \end{aligned}$$

(3) is immediate from definition, and we will prove (4) in greater generality below. □

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

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

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

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

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

Lemma. With notation as above,

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

Proof. The first is a special case of the second, where $\eta = \{X\}$, so we only prove the second.

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

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

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

Proof. Note the second immediately follows from the first by the previous lemma.

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

Lemma. Let notation be as above. Then

$$H_\mu(\xi \mid \eta) \geq H_\mu(\xi \mid \eta \vee \lambda).$$

Proof.

$$H_\mu(\xi \mid \eta \vee \lambda) = \sum_{A\in\xi, B\in\eta, C\in\lambda} \mu(A\cap B\cap C)\log\frac{\mu(B\cap C)}{\mu(A\cap B\cap C)}$$
$$H_\mu(\xi \mid \eta) = \sum_{A\in\xi, B\in\eta} \mu(A\cap B)\log\frac{\mu(B)}{\mu(A\cap B)}.$$

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

$$\mu(A\cap B)\log\frac{\mu(B)}{\mu(A\cap B)} \ge \sum_{C\in\lambda}\mu(A\cap B\cap C)\log\frac{\mu(B\cap C)}{\mu(A\cap B\cap C)}.$$

Apply Jensen's inequality for $x \mapsto x\log x$ at the 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

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

thus

$$\frac{\mu(A\cap B)}{\mu(B)}\cdot\log\frac{\mu(A\cap B)}{\mu(B)} \le \sum_{C\in\lambda}\frac{\mu(B\cap C)}{\mu(B)}\cdot\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}\cdot\log\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}$$
$$\mu(A\cap B)\cdot\log\frac{\mu(A\cap B)}{\mu(B)} \le \sum_{C\in\lambda}\mu(A\cap B\cap C)\cdot\log\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}.$$ □

Corollary.

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

Proof. Using the chain rule:

$$H_\mu(\xi\vee\eta) = H_\mu(\xi) + H_\mu(\eta \mid \xi).$$ □

Lecture 16 **Lemma.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let ξ, η be countable measurable partitions. Then

$$I_\mu(T^{-1}\xi \mid T^{-1}\eta)(x) = I_\mu(\xi \mid \eta)(Tx)$$
$$H_\mu(T^{-1}\xi \mid T^{-1}\eta) = H_\mu(\xi \mid \eta)$$

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

Proof.

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

Note

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

$\mu([x]_{T^{-1}\xi\vee T^{-1}\eta}) = \mu([Tx]_{\xi\vee\eta})$ by the measure preserving property. Similarly,

$$\mu([x]_{T^{-1}\eta}) = \mu([Tx]_\eta).$$

Then

$$\begin{aligned} I_\mu(T^{-1}\xi \mid T^{-1}\eta)(x) &= -\log \frac{\mu([Tx]_{\xi\vee\eta})}{\mu([Tx]_\eta)} \\ &= I_\mu(\xi \mid \eta)(Tx). \end{aligned}$$

□

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

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

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

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

□

Lemma (Fekete's lemma)**.** Let $(a_n) \subset \mathbb{R}$ be a subadditive sequence, i.e.

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

Then $\lim_{n\to\infty} \frac{a_n}{n} = \inf \frac{a_n}{n}$.

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 $0 \le i(n) < n_0$. Iterate subadditivity: $a_n \le j(n)a_{n_0} + a_{i(n)}$. □

Definition (Entropy of a system)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let ξ be a countable measurable partition such that $H_\mu(\xi) < \infty$. The **entropy** of the measure preserving system with respect to ξ is

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

The **entropy** of the measure preserving system is

$$h_\mu(T) = \sup \{\, h_\mu(T, \xi) \mid H_\mu(\xi) < \infty \,\}$$

Definition (2-sided generator)**.** Let $(X, \mathcal{B}, \mu, T)$ be an invertible measure preserving system. Let $\xi \subset \mathcal{B}$ be a countable measurable partition. 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 $\exists A' \in \mathcal{B}(\xi_{-k}^k)$ (the σ-algebra generated by ξ_{-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 return to the proof later; for now we calculate the entropy of Bernoulli shift.

Let $(\{1, 2, \ldots, k\}^{\mathbb{Z}}, \mathcal{B}, \mu, \sigma)$ be the $(p_1, \ldots, p_k)$-Bernoulli shift. Claim: The partition $\xi = \{\, \{\, x \in X \mid x_0 = j \,\} \mid j = 1, \ldots, k \,\}$ is a 2-sided generator.

Proof. The collection of sets

$$\{ A \in \mathcal{B} \mid \forall \epsilon \ \exists k \ \exists A' \in \xi_{-k}^{k} \text{ with } \mu(A \triangle A') < \epsilon \}$$

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

Claim. With ξ defined as before, we have

$$\begin{aligned} H_\mu(\xi \mid \xi_1^n) &= H(p_1, p_2, \ldots, p_k) \\ &= -p_1 \log p_1 - \cdots - p_k \log p_k \end{aligned}$$

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

Proof. Calculate the information function:

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

Note $[x]_{\xi_0^n} = \{ y \in X \mid y_0 = x_0, y_1 = x_1, \ldots, y_n = x_n \}$. $\mu([x]_{\xi_0^n}) = p_{x_0} \cdots p_{x_n}$. Similarly $\mu([x]_{\xi_1^n}) = p_{x_1} \cdots p_{x_n}$

$$\begin{aligned} I_\mu(\xi \mid \xi_1^n)(x) &= -\log p_{x_0} \\ H_\mu(\xi \mid \xi_1^n)(x) &= \sum_{j=1}^{k} p_j \cdot (-\log p_j) = H(p_1, \ldots, p_k). \end{aligned}$$

$$\begin{aligned} H_\mu(\xi_0^{n-1}) &= H_\mu(\xi_{n-1}^{n-1}) + H(\xi_{n-2}^{n-2} \mid \xi_{n-1}^{n-1}) + H(\xi_{n-3}^{n-3} \mid \xi_{n-2}^{n-1}) + \cdots + H_\mu(\xi \mid \xi_1^{n-1}) \\ &= H_\mu(\xi) + H_\mu(\xi \mid \xi_1^1) + H_\mu(\xi \mid \xi_1^2) + \cdots + H_\mu(\xi \mid \xi_1^{n-1}) \\ &= nH(p_1, \ldots, p_k). \end{aligned}$$

Divide by n and take the limit:

$$h_\mu(T) = h_\mu(T, \xi) = H(p_1, \ldots, p_k).$$ $\square$

Thus, the entropy of $(\frac{1}{2}, \frac{1}{2})$ shift is $\log 2$, and entropy of $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ shift is $\log 3$.

Lecture 17 **Lemma 1.** Let $(X, \mathcal{B}, \mu, T)$ be an invertible measure preserving system. Let $\xi \subset \mathcal{B}$ be a countable measurable partition. Then

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

Lemma 2. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\xi, \eta \subset \mathcal{B}$ be two countable measurable partitions. Then:

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

Lemma 3. For all $\epsilon > 0$, $k \in \mathbb{Z}_{>0}$, there is $\delta > 0$ 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 $A \in \eta$, $\exists B \in \mathcal{B}(\xi)$ such that $\mu(A \triangle B) < \delta$. Then $H_\mu(\eta \mid \xi) \leq \epsilon$.

Proof of Kolmogorov-Sinai theorem. We first show $h_\mu(T,\xi) = \sup_{\eta \text{ finite}} h_\mu(T,\eta)$. We need to show that for all finite partitions $\eta \subset \mathcal{B}$, we have $h_\mu(T,\eta) \le h_\mu(T,\xi)$. Fix $\epsilon > 0$. By Lemma 3, and definition of 2-sided generator, $\exists n$ such that $H_\mu(\eta \mid \xi_{-n}^n) \le \epsilon$. Then we have

$$h_\mu(T,\eta) \le h_\mu(T,\xi_{-n}^n) + \epsilon$$

by Lemma 2. e are done if ϵ sufficiently small, by Lemma 1.

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 $h_\mu(T,\eta) \le h_\mu(T,\tilde{\eta}) + \epsilon$. By Lemma 2, enough to show $H_\mu(\eta \mid \tilde{\eta}) \le \epsilon$. Let $\tilde{\eta} = (A_1, A_2, \dots, A_n, \bigcup_{j>n} A_j)$ for sufficiently large n, when $\eta = \{A_1, A_2, \dots\}$. Then

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

Thus

$$\begin{aligned} H_\mu(\eta \mid \tilde{\eta}) &= \sum_{j>n} -\mu(A_j)\log\mu(A_j) + \mu\left(\bigcup_{j=n+1}^{\infty} A_j\right) \cdot \log\mu\left(\bigcup_{j=n+1}^{\infty} A_j\right) \\ &\le \sum_{j>n} -\mu(A_j)\log\mu(A_j) \le \epsilon \end{aligned}$$

if n is sufficiently large. □

Proof of Lemma 1.

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

Proof of Lemma 2.

$$\begin{aligned} h_\mu(T,\eta) &= \lim \frac{1}{n} H_\mu(\eta_0^{n-1}) \le \lim \frac{1}{n} H_\mu((\xi \vee \eta)_0^{n-1}) \\ &= \lim \frac{1}{n}\left(H_\mu(\xi_0^{n-1}) + H_\mu(\eta_0^{n-1} \mid \xi_0^{n-1})\right) \\ &= h_\mu(T,\xi) + \lim \frac{1}{n}\left(\sum_{j=0}^{n-1} H_\mu\left(\eta'_j \mid \xi_0^{n-1} \vee \eta_{j+1}^{n-1}\right)\right) \\ &\le h_\mu(T,\xi) + \lim \frac{1}{n}\left(\sum_{j=0}^{n-1} \underbrace{H_\mu(\eta_j^j \mid \xi_j^j)}_{H_\mu(\eta\mid\xi)}\right) = h_\mu(T,\xi) + H_\mu(\eta \mid \xi). \end{aligned}$$ □

Proof of Lemma 3. Write $A_1, \dots, A_k$ for the atoms of η. And for each $i \in \{1,\dots,k\}$, let $B_i \in \mathcal{B}(\xi)$ such that $\mu(A_i \triangle B_i) < \delta$. We consider the partition λ which has the following

$k+1$ atoms.

$$C_0 = \bigcup_{i=1}^{k} \left(A_i \cap \left(B_i \setminus \bigcup_{j \neq i} B_j \right) \right)$$
$$C_i = A_i \setminus C_0 \text{ for } i = 1, \ldots, k.$$

C_0 has two crucial properties we note here:

– C_0 is big:

$$\begin{aligned} \mu(C_0) &\geq \sum (\mu(A_i) - k\delta) \\ &= 1 - k^2\delta \end{aligned}$$

– If $x \in C_0$, then $x \in A_i \iff x \in B_i$.

Note

$$\begin{aligned} H_\mu(\lambda) &= -\mu(C_0) \log \mu(C_0) - \sum_{i=1}^{k} \mu(C_i) \log \mu(C_i) \\ &\leq -\mu(C_0) \log \mu(C_0) - \left(\sum_{i=1}^{k} \mu(C_i) \right) \log \left(\frac{\sum_{i=1}^{k} \mu(C_i)}{k} \right). \end{aligned}$$

If δ is sufficiently small, then $\mu(C_0)$ is as close to 1 as we want, and $\sum_{i=1}^{k} \mu(C_i)$ is as close to 0 as we want. Then $H_\mu(\lambda) < \epsilon$ if δ is sufficiently small. Now it is enough to show that $H_\mu(\eta \mid \xi \vee \lambda) = 0$. Then,

$$\begin{aligned} H_\mu(\eta \mid \xi) &\leq H_\mu(\eta \vee \lambda \mid \xi) \\ &\leq \underbrace{H_\mu(\lambda \mid \xi)}_{\leq H_\mu(\lambda) < \epsilon} + \underbrace{H_\mu(\eta \mid \xi \vee \lambda)}_{=0} \end{aligned}$$

It is now enough to show
$$I_\mu(\eta \mid \xi \vee \lambda)(x) = 0 \quad \forall x.$$

To this end, it is enough to show

$$[x]_{\xi \vee \lambda} \subset [x]_\eta \quad \forall x.$$

We consider two cases.

1. $x \in C_0$. Let $A_i = [x]_\eta$. This means that $x \in A_i$. Because $x \in C_0$, we must have $x \in B_i$. Now $C_0, B_i \in \mathcal{B}(\xi \vee \lambda)$. Then $[x]_{\xi \vee \lambda} \subset C_0 \cap B_i \subset A_i$.

2. $x \in C_i$ for some I. Then $[x]_\eta = A_i$.

$$x \in C_i \Rightarrow [x]_{\eta \vee \lambda} \subset C_i \subset A_i.$$

□

Lecture 18 **Theorem** (Shannon-McMillan-Brieman)**.** Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Let $\xi \subset \mathcal{B}$ be a countable measurable partition with $H_\mu(\xi) < \infty$. Then

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

pointwise μ-almost everywhere and in L^1.

Recall

$$I(\xi_0^{N-1})(x) = -\log \mu([x]_{\xi_0^{N-1}}),$$

so this theorem is loosely saying

$$\exp(-N h_\mu(T, \xi)) \approx \mu([x]\xi_0^{N-1}).$$

Idea: using the chain rule and then invariance:

$$\begin{aligned} \frac{1}{N} I_\mu(\xi_0^{N-1})(x) &= \frac{1}{N} \sum_{n=0}^{N-1} I_\mu(\xi_n^n \mid \xi_{n+1}^{N-1})(x) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} I_\mu(\xi \mid \xi_1^{N-n-1})(T^n x). \end{aligned}$$

Definition. Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then 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) $\int_A f \, d\mu = \int_A f^* \, d\mu \quad \forall A \in \mathcal{A}$.

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

Remark. Writing $V_\mathcal{A}$ for the closed subspace of $L^2(X, \mathcal{B}, \mu)$ consisting of $\mathcal{A}$-measurable functions, $\mathbb{E}[f \mid \mathcal{A}]$ 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

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

Theorem. Let $(X, \mathcal{B}, \mu)$ be a probability space, let

$$f, f_1, f_2 \in L^1(X, \mathcal{B}, \mu), \quad \mathcal{A}, \mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{B} \text{ be sub-}\sigma\text{-algebras.}$$

Then

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

(2) If f_1 is $\mathcal{A}$-measurable, then

$$\mathbb{E}[f_1 f_2 \mid \mathcal{A}] = f_1 \mathbb{E}[f_2 \mid \mathcal{A}]$$

(3) If $\mathcal{A}_1 \subseteq \mathcal{A}_2$ then

$$\mathbb{E}[f \mid \mathcal{A}_1] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{A}_2] \mid \mathcal{A}_1].$$

Theorem (Martingale theorems)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space, let $f \in L^1$ and let $\mathcal{A}, \mathcal{A}_1, \mathcal{A}_2, \ldots \subset \mathcal{B}$ be sub-σ-algebras. Assume that either

(1) $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \cdots$ and $\mathcal{A} = \mathcal{B}(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots)$

or

(2) $\mathcal{A}_1 \supset \mathcal{A}_2 \supset \mathcal{A}_3 \supset \cdots$ and $\mathcal{A} = \bigcap \mathcal{A}_n$

Then

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

pointwise μ-almost everywhere and in L^1.

We now aim to define conditional information and entropy relative to σ-algebras.

Definition (Conditional information relative to sub-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

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

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

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

Definition.

$$H_\mu(\eta \mid \mathcal{A}) \coloneqq \int I_\mu(\eta \mid \mathcal{A}) \, d\mu.$$

Lecture 19 **Theorem** (Maximal inequality)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space, and let

$$\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots$$

be a sequence of sub-σ-algebras of $\mathcal{B}$ and let $\eta \subset \mathcal{B}$ be another countable partition with $H_\mu(\eta) < \infty$. Then

$$I_\mu(\eta \mid \mathcal{A}_n) \to I_\mu(\eta \mid \mathcal{A})$$

as $n \to \infty$ pointwise and in L^1, where $\mathcal{A} = \mathcal{B}(\bigcup \mathcal{A}_n)$. Moreover,

$$I^*(x) \coloneqq \sup_{n \in \mathbb{Z}_{>0}} I_\mu(\eta \mid \mathcal{A}_n) \in L^1(X, \mathcal{B}, \mu). \tag{$*$}$$

Proof. We do only the case where η is a finite partition, the countable case is dealt with on the example sheet. By definition and martingale convergence:

$$\begin{aligned} I_\mu(\eta \mid \mathcal{A}_n)(x) &= -\log \mathbb{E}[\chi_{[x]_\eta} \mid \mathcal{A}_n](x) \\ &\xrightarrow{n\to\infty} -\log \mathbb{E}[\chi_{[x]_\eta} \mid \mathcal{A}](x) \\ &= I_\mu(\eta \mid \mathcal{A})(x). \end{aligned}$$

This proves pointwise convergence.

By dominated convergence, L^1 convergence follows by ($*$). Enough to show ($*$). Fix $A \in \eta$. Fix $\alpha > 0$. For $x \in X$, let $n(x)$ be the smallest positive integer n such that

$$-\log(\mathbb{E}[\chi_A \mid \mathcal{A}_n](x)) > \alpha.$$

If the above does not hold for any n, then we write $n(x) = \infty$. We write:

$$B_n = \{ x \in X : n(x) = n \}.$$

Note that B_n is $\mathcal{A}_n$-measurable. Indeed

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

Then:

$$\int_{B_n} \mathbb{E}[\chi_A \mid \mathcal{A}_n]\, d\mu = \int_{B_n} \chi_A = \mu(B_n \cap A)$$

by definition of conditional expectation, but

$$\mu(B_n) \cdot \exp(-\alpha) \ge \int_{B_n} \mathbb{E}[\chi_A \mid \mathcal{A}_n]\, d\mu.$$

Define

$$A^* = \{ x \in A \mid I^*(x) > \alpha \} \subset A \cap \left(\bigcup B_n \right)$$

so

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

$$\mu(\{ x \in X \mid I^*(x) > \alpha \}) \le |\eta| \exp(-\alpha).$$

Now,

$$\begin{aligned} \int I^*(x)\, d\mu &\le \sum_{n=1}^\infty \mu(x \in X \mid I^*(x) > n-1) \cdot n \\ &\le \sum_{n=1}^\infty |\eta| \exp(-(n-1))n < \infty. \end{aligned}$$

□

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

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

Proof.

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

$$h_\mu(T,\xi) = \underset{n\to\infty}{\text{C-lim}}\, H_\mu(\xi \mid \xi_1^n)$$

Observe that $H_\mu(\xi \mid \xi_1^n)$ is a monotone decreasing sequence, hence it converges. Then the Cesaro limit is equal to the ordinary limit, so done. □

Proof of Shannon-McMillan-Brieman.

$$\begin{aligned}
\frac{1}{N} I_\mu(\xi_0^{N-1})(x) &= \frac{1}{N} \sum_{n=0}^{N-1} I_\mu(\xi_n^n \mid \xi_{n+1}^{N-1})(x) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} I_\mu(\xi \mid \xi_1^{N-n-1})(T^n x) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} I_\mu(\xi \mid \mathcal{B}(\xi_1^\infty))(T^n x) \\
&\quad + \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} \left(I_\mu(\xi \mid \xi_1^{N-n-1})(T^n x) - I_\mu(\xi \mid \mathcal{B}(\xi_1^\infty))(T^n x) \right)}_{(**)}
\end{aligned}$$

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

By the Pointwise Ergodic Theorem:

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

Write

$$I_K^*(x) = \sup_{k\ge K} \left| I_\mu\left(\xi \mid \xi_1^k\right)(x) - I_\mu\left(\xi \mid \mathcal{B}(\xi_1^\infty)\right)(x) \right|.$$

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

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

Now,

$$|(**)| \le \underbrace{\frac{1}{N} \sum_{n=0}^{N-K-1} I_K^*(T^n x)}_{\to \int I_K^*\, d\mu} + \frac{1}{N} \sum_{n=N-K}^{N} I_0^*(T^n x).$$

The right hand term is

$$\underbrace{\frac{1}{N}\sum_{n=0}^{N_1} I_0^*(T^n x)}_{\to \int I_0^* \, d\mu} - \underbrace{\frac{N-K}{N}}_{\to 1} \underbrace{\frac{1}{N-K}\sum_{n=0}^{N-K-1} I_0^*(T^n x)}_{\to \int I_0^* \, d\mu}$$

which converges to 0. Thus $\limsup_{N\to\infty} |(**)| \le \int I_K^* \, d\mu$. K is arbitrary, and

$$\int I_K^* \, d\mu \xrightarrow{K\to\infty} 0,$$

so $\limsup_{N\to\infty} |(**)| = 0$. □

Lecture 20 Missing content Left to prove: If $h_\mu(T,\xi) > 0 \ \forall \xi \in \mathcal{B}$ finite with $H_\mu(\xi) > 0$ then $\mathcal{T}(\xi)$ is trivial $\forall \xi \subseteq \mathcal{B}$ finite.

Proposition. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\xi, \eta \subset \mathcal{B}$ be two finite partitions. Then

$$h_\mu(T,\xi) = H_\mu(\xi \mid \mathcal{B}(\xi_1^\infty) \vee \mathcal{T}(\eta)).$$

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

$$h_\mu(T,\xi) = \frac{1}{n} H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty))$$

Proof. By chain rule and invariance:

$$\begin{aligned} H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty)) &= \sum_{j=0}^{n-1} H_\mu(\xi_j^j \mid \mathcal{B}(\xi_{j+1}^\infty)) \\ &= n H_\mu(\xi \mid \mathcal{B}(\xi_1^\infty)) \\ &= n h_\mu(T,\xi). \end{aligned}$$ □

Lemma. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\eta, \xi \subset \mathcal{B}$ be finite measurable partitions. Then

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

Proof. Observe that $\forall n \in \mathbb{Z}_{>0}$,

$$\begin{aligned} H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty)) &\le H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty)) \\ &= n h_\mu(T,\xi). \end{aligned}$$

We do the following. We choose two sequences a_n and b_n such that $a_n \le b_n \ \forall n$ and

$$\lim \frac{1}{n}[H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty)) + a_n] = \lim \frac{1}{n}[H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty)) + b_n]. \tag{1}$$

Suppose to the contrary that

$$\liminf \frac{1}{n}[H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty))] < h_\mu(T,\xi) = H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty)).$$

Then $a_n \leq b_n$ and this contradicts (1). This means that

$$\begin{aligned} h_\mu(T,\xi) &\leq \liminf \frac{1}{n}[H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty))] \\ &\leq \limsup \frac{1}{n}[H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty))] \\ &\leq h_\mu(T,\xi). \end{aligned}$$

This proves the lemma.

We take

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

Clearly $a_n \leq b_n$ indeed.

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

By the earlier lemma,

$$\frac{H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty)) + a_n}{n} = h_\mu(T, \xi \vee \eta)$$

so we have

$$h_\mu(T,\xi\vee\eta) = \frac{H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty) \vee \mathcal{B}(\eta_n^\infty)) + a_n}{n} = \frac{H_\mu(\xi_0^{n-1} \mid \mathcal{B}(\xi_n^\infty)) + b_n}{n} \leq \frac{H_\mu(\xi_0^{n-1} \vee \eta_0^{n-1})}{n} \to h_\mu(T,\xi\vee\eta).$$

□

Suppose that $h_\mu(T,\xi) > 0$ $\forall \xi \subset \mathcal{B}$ finite with $H_\mu(\xi) > 0$. Suppose to the contrary that $\mathcal{T}(\xi)$ is non-trivial for some $\xi \subset \mathcal{B}$ finite. Then $\exists A \subset \mathcal{T}(\xi)$ such that $0 < \mu(A) < 1$. Define $\eta = \{A, X \setminus A\}$. Apply the proposition with $\xi \leftrightarrow \eta$:

$$h_\mu(T,\eta) = H_\mu(\eta \mid \mathcal{B}(\eta_1^\infty) \vee \mathcal{T}(\xi)).$$

We prove that $h_\mu(T,\eta) = 0$, which is contradiction, as $H_\mu(\eta) > 0$.

In fact we will show

$$I_\mu(\eta \mid \mathcal{B}(\eta_1^\infty) \vee \mathcal{T}(\xi))(x) = 0$$

for a.e. x.

$$\begin{aligned} I_\mu(\eta \mid \mathcal{B}(\eta_1^\infty) \vee \mathcal{T}(\xi)) &= -\log \mathbb{E}[\xi_{[x]_\eta} \mid \mathcal{B}(\eta_1^\infty) \vee \mathcal{T}(\xi)] \\ &= -\log \xi_{[x]_\eta} = 0. \end{aligned}$$

10 Rudolph's Theorem

Theorem. 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 $A \subset \mathcal{B}$ such that $T_2^{-1}A = A$ and $T_3^{-1}A = A$, then $\mu(A) = 0$ or 1. Suppose that $h_\mu(T_2)$ or $h_\mu(T_3) > 0$. Then μ is the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$.

Theorem (Host)**.** Let μ be an ergodic T_3-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 equidistributed in $\mathbb{R}/\mathbb{Z}$. Moreover, if μ is not necessarily ergodic, but the other assumptions hold, then

$$\mu(x \in \mathbb{R}/\mathbb{Z} : x \text{ is normal in base 2}) \geq \frac{h_\mu(T_3)}{\log 3}.$$

Lecture 22

Lemma. We have $\mathrm{ord}_{3^k}(2) = 2 \cdot 3^{k-1}$, where $\mathrm{ord}_{3^k}(2)$ is the order of 2 mod 3^k, that is, the smallest integer n such that $3^k \mid 2^n - 1$.

Proof. By induction on k, we prove that $\mathrm{ord}_{3^k}(2) = 2 \cdot 3^{k-1}$ and $3^{k+1} \nmid 2^{2\cdot 3^{k-1}} - 1$ Easy to check for $k = 0, 1$. Assume that $k \geq 2$ and the claim holds for $k-1$. $3^k \mid 2^{\mathrm{ord}_{3^k}(2)} - 1$, so $3^{k-1} \mid 2^{\mathrm{ord}_{3^k}(2)} - 1$, hence $\mathrm{ord}_{3^k}(2) = a \cdot 2 \cdot 3^{k-2}$ for some $a \in \mathbb{Z}_{>0}$. We also know

$$2^{2\cdot 3^{k-1}} = b \cdot 3^{k-1} + 1$$

for some $b \in \mathbb{Z}_{>0}$ and $3 \nmid b$. Then by the binomial theorem, we have:

$$\begin{aligned} 2^{2\cdot 2\cdot 3^{k-2}} &= b^2 \cdot 3^{2(k-1)} + 2b \cdot 3^{k-1} + 1 \\ 2^{3\cdot 2\cdot 3^{k-2}} &= b^3 \cdot 3^{(k-1)} + 3b^2 \cdot 3^{2(k-1)} + 3b \cdot 3^{k-1} + 1 \end{aligned}$$

From this we see that $3^k \nmid 2^{2\cdot 2\cdot 3^{k-2}} - 1$ but $3^k \mid 2^{3\cdot 2\cdot 3^{k-2}} - 1$ and $3^{k+1} \nmid 2^{3\cdot 2\cdot 3^{k-2}} - 1$. □

Remark. Look at the T_2 orbit of $\frac{a}{3^k}$ where $a \in \mathbb{Z}$ and $3 \nmid a$. This orbit is

$$\{\, \frac{b}{3^k} \mid b \in [0, \ldots, 3^k - 1], 3 \nmid b \,\}.$$

Also $T_2^n(x + \frac{a}{3^k}) = T_2^n(x) + T_2^n(\frac{a}{3^k})$.

Another idea: If $h_\mu(T_3) > 0$, the μ 'will not concentrate' on a few elements of the set

$$\{\, x + \frac{a}{3^k} : a \in [0, \ldots, 3^k - 1], 3 \nmid a \,\}$$

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: $f(x) = \exp(2\pi i m x)$ for $m \in \mathbb{Z}$ (by density and linearity).

Notation. For $N \in \mathbb{Z}_{>0}$, $m \in \mathbb{Z}$, 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 $3^\alpha \mid m$. 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.

$$\begin{aligned} |P_{N,m}(x)|^2 &= \left| \frac{1}{N} \sum_{n=0}^{N-1} \exp 2\pi i m 2^n x \right|^2 \\ &= \frac{1}{N^2} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \exp(2\pi i m 2^{n_1} x) \cdot \overline{\exp(2\pi i m 2^{n_2} x)} \\ &= \frac{1}{N^2} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \exp(2\pi i m (2^{n_1} - 2^{n_2}) x). \end{aligned}$$

Note that for some $b \in \mathbb{Z}$:

$$\sum_{a=0}^{3^k-1} \exp(2\pi i b (x + \frac{a}{3^k})) = \exp(2\pi i b x) \cdot \sum_{a=0}^{3^k-1} \exp(2\pi i b \frac{a}{3^k}) \tag{$*$}$$

If $3^k \nmid b$, then $(*)$ is 0. If $3^k \mid b$, then $(*) = 3^k \exp(2\pi i b x)$. When does it happen that $3^k \mid m \cdot (2^{n_1} - 2^{n_2})$. Suppose $n_1 > n_2$. Then $m \cdot (2^{n_1} - 2^{n_2}) = m \cdot 2^{n_2}(2^{n_1 - n_2} - 1)$. Since $n_1 - n_2 \le N < 2 \cdot 3^{k-\alpha-1}$, we have $3^{k-\alpha} \nmid 2^{n_1-n_2} - 1$ so $3^k \nmid m \cdot (2^{n_1} - 2^{n_2})$. Same applies if $n_2 > n_1$. We get that $3^k \mid m(2^{n_1} - 2^{n_2}) \iff n_1 = n_2$. Therefore,

$$|P_{N,m}(x)|^2 = \frac{1}{N^2} \sum_{n=0}^{N-1} 3^k = \frac{3^k}{N}.$$

$\square$

Lecture 23

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

$$\mu_k([x]_{\xi_0^K}) \ge \exp((h_\mu(T_3) - \epsilon)k)\mu([x]_{\xi_0^K})$$

holds for μ-almost every x provided k is sufficiently large depending on ϵ, x and K is sufficiently large depending on ϵ, x, k.

Remark.

$$[x]_{\xi_0^K} = \{\, y \in \mathbb{R}/\mathbb{Z} = [0,1) \mid y = 0.y_1 y_2 y_3 \ldots_{(3)}, y_1 = x_1, \ldots, y_{K+1} = x_{K+1} \,\}$$

where $x = 0.x_1x_2x_3\ldots_{(3)}$. This is an interval of length $3^{-(K+1)}$ containing x.

$$\mu_k([x]_{\xi_0^K}) = \mu(\bigcup_{a=0}^{3^k-1}[x]_{\xi_0^K} + \frac{a}{3^k}).$$

Lemma. In the above setting, we have

$$\frac{1}{k}I_\mu(\xi_0^{k-1}|\mathcal{B}(\xi_k^\infty))(x) \to h_\mu(T_3)$$

μ-almost everywhere.

Proof. Using chain rule and invariance,

$$\begin{aligned}
\frac{1}{k}I_\mu(\xi_0^{k-1} \mid \mathcal{B}(\xi_k^\infty))(x) &= \frac{1}{k}\sum_{j=0}^{k-1} I_\mu(\xi_j^j \mid \mathcal{B}(\xi_{j+1}^\infty))(x) \\
&= \frac{1}{k}\sum_{j=0}^{k-1} I_\mu(\xi \mid \mathcal{B}(\xi_1^\infty))(T^j x) \\
\text{by pointwise ergodic theorem} \quad &\to \int I_\mu(\xi \mid \mathcal{B}(\xi_1^\infty))\, d\mu \\
&= H_\mu(\xi \mid \mathcal{B}(\xi_1^\infty)) = h_\mu(T_3,\xi).
\end{aligned}$$

The proof will be complete with the following.

Lemma. The partition ξ is a 1-sided generator for $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, T_3)$. That is, $\forall A \in \mathcal{B}$ and $\epsilon > 0$, $\exists n \in \mathbb{Z}_{>0}$ and $B \in \mathcal{B}(\xi_0^k)$ such that $\mu(A \triangle B) < \epsilon$.

Proof. Consider:

$$\mathcal{C} = \{\, A \in \mathcal{B} \mid \forall \epsilon > 0, \exists n \in \mathbb{Z}_{>0}, B \in \mathcal{B}(\xi_0^n) \text{ such that } \mu(A \triangle B) < \epsilon \,\}.$$

This is a σ-algebra, so it is enough to show that it contains all open sets. Let $U \subseteq \mathbb{R}/\mathbb{Z}$ be open. Claim that

$$U = \bigcup_{x \in U, n \in \mathbb{Z}_{>0} \mid [x]_{\xi_0^n} \subset U} [x]_{\xi_0^n}.$$

Note that this is a countable union and $[x]_{\xi_0^n} \in \mathcal{C}\ \forall x \forall n$. □

Corollary. Let μ be a T_3-invariant ergodic probability measure on $\mathbb{R}/\mathbb{Z}$. Then for all $k \in \mathbb{Z}_{>0}$,

$$\exists A_k \in \mathcal{B}$$

such that

(1) μ-almost every x is contained in A_k for k sufficiently large.

(2)
$$\mu_k(U) \geq \exp(k h_\mu(T_3)/2)\mu(A_k \cap U)$$
holds for all open sets $U \subset \mathbb{R}/\mathbb{Z}$.

Proof of proposition. By the lemma, if k is sufficiently large depending on x, ϵ, then

$$I_\mu(\xi_0^{k-1} \mid \mathcal{B}(\xi_k^\infty))(x) \geq k(h_\mu(T_3) - \epsilon/2).$$

If K sufficiently large, then

$$I_\mu(\xi_0^{k-1} \mid \xi_k^K)(x) \geq k(h_\mu(T_3) - \epsilon)$$

by SMB, and the LHS is

$$\log \frac{\mu([x]_{\xi_k^K})}{\mu([x]_{\xi_0^K})}.$$

Observe

$$[x]_{\xi_k^K} = \bigcup_{a=0,\dots,3^k-1} ([x]_{\xi_0^K} + \frac{a}{3^k})$$

Thus $\mu_k([x]_{\xi_0^K}) = \mu([x]_{\xi_k^K})$, and so we have

$$\log \frac{\mu_k([x]_{\xi_0^K})}{\mu([x]_{\xi_0^K})} \geq k(h_\mu(T_3) - \epsilon).$$

□

Proof of corollary. Let A_k be the set of points x such that

$$\mu_k([x]_{\xi_0^K}) \geq \exp(k \cdot \mu_k(T_3)/2)\mu([x]_{\xi_0^K}) \tag{$*$}$$

for all sufficiently large K. Claim (1) follows from the proposition.

To prove (2), fix $k \in \mathbb{Z}_{>0}$ and fix $U \subseteq \mathbb{R}/\mathbb{Z}$ open. For $K \in \mathbb{Z}_{>k}$, write $\mathcal{A}_K$ for the collection of atoms $[x]_{\xi_0^K}$ such that $[x]_{\xi_0^K} \subset U$ and (($*$)) 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 $B_K = \bigcup_{A \in \mathcal{A}_K} A$.

$$\mu_K(B_K) \geq \exp(k \cdot h_\mu(T_3)/2)\mu(B_K)$$

follows by summing (($*$)). Also, $\mu_k(B_K) \leq \mu_k(U)$. Now consider $\bigcup_{K \geq L} B_K$ for $L \in \mathbb{Z}_{>k}$. These sets increase as L grows, and their union is $A_k \cap U$: Recall that $\forall x \in A_k \cap U$, 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.

Since $\bigcap_{K \geq L} B_K \subset B_L$, we have

$$\mu_k(U) \geq \cdots \geq \mu(\bigcap_{K \geq L} B_K).$$

□

Lecture 24 We have reduced Host's theorem (in the ergodic case) to showing $P_{N,m}(x) \to 0$ for μ-a.e. x for all $m \neq 0$.

Recall

Lemma (Borel-Cantelli)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $B_1, B_2, \ldots, \in \mathcal{B}$. Suppose

$$\sum_{n=1}^{\infty} \mu(B_n) < \infty.$$

Then for μ-almost every x, $x \in B_n$ holds for at most finitely many n.

Lemma. For all $m \neq 0$ and $c \in \mathbb{Z}_{>0}$,

$$P_{N,m}(x) \xrightarrow{N\to\infty} 0 \iff P_{L^c,m}(x) \xrightarrow{L\to\infty} 0.$$

Proof. It is enough to show that $P_{L^c,m}(x) \to 0$ as $L \to \infty$, for some C (to be chosen later) and all $m \neq 0$. Fix $m \neq 0$ and $\epsilon > 0$. For $N \in \mathbb{Z}_{>0}$, we set $k = k(N)$ to be an integer such that $2 \cdot 3^{k-2-\alpha} \leq N < 2 \cdot 3^{k-1-\alpha}$. By Lemma 1, we have:

$$\sum_{a=0}^{3^k-1} |P_{N,m}(x + \frac{a}{3^k})|^2 = \frac{3^k}{N} \leq \frac{N}{2} \cdot 3^{\alpha+2}/N = \frac{3^{\alpha+2}}{2}.$$

By the definition of μ_k, we have

$$\int |P_{N,m}(x)|^2 \, d\mu_k(x) = \int \sum_{a=0}^{3^k-1} |P_{N,m}(x + \frac{a}{3^k})|^2 \, d\mu \leq \frac{3^{\alpha+2}}{2}.$$

By Markov's inequality:

$$\mu_k(x : P_{N,m}(x) > \epsilon) \leq \frac{3^{\alpha+2}}{2\epsilon^2}.$$

By the corollary,

$$\mu(x : |P_{N,m}(x)| > \epsilon, x \in A_k) \leq \exp(-kh_\mu(T_3)/2) \cdot \frac{3^{\alpha+2}}{2\epsilon^2}$$

where $k = k(N)$ defined earlier. Let

$$B_\epsilon = \{\, x \mid |P_{L^c,m}(x)| > \epsilon \text{ for infinitely many } L \,\}.$$

Our goal is to show that $\mu(B_\epsilon) = 0$, and this completes the proof. Write

$$D_{L,\epsilon} = \{\, x \mid x \in A_{k(L)}, |P_{L^c,m}(x)| > \epsilon \,\}.$$

Note that $x \in B_\epsilon$ implies $x \in D_{L,\epsilon}$ for infinitely many Ls. So by B-C lemma, it is enough to show $\sum_{L=1}^{\infty} \mu(D_{L,\epsilon}) < \infty$.

Recall we have:

$$\mu(D_{L,\epsilon}) \leq \exp(-k(L^c)h_\mu(T_3)/2) \cdot \frac{3^{\alpha+2}}{2\epsilon^2}.$$

If we choose c sufficiently large, then

$$k(L^c) > 4 \log L / h_\mu(T_3).$$

So

$$\sum \mu(D_{L,\epsilon}) \leq \sum L^{-2} \cdot \frac{3^{\alpha+2}}{2 \cdot \epsilon^2} < \infty.$$

□

Theorem (Rudolph's Theorem)**.** Let μ be a probability measure on $\mathbb{R}/\mathbb{Z}$ that is invariant and ergodic under the joint action of T_2 and T_3. If $h_\mu(T_3) > 0$ then μ is the Lebesgue measure.

Proof. Write

$$A = \{\, x \mid \text{normal in base } 2 \,\}.$$

By Host's theorem, $\mu(A) > 0$.

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. If $x \in A$, then $\forall f \in C(X)$, we have

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T_2^n x) \to \int f(x)\, dx.$$

Now we can write

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} f(T_2^n T_3 x) &= \frac{1}{N} \sum_{n=0}^{N-1} f(T_3 T_2^n x) \\ &\to \int f \circ T_3(x)\, dx = \int f(x)\, dx. \end{aligned}$$

using $f \circ T_3$ in the role of f This proves that $T_3 x \in A$.

The claim gives $A \subset T_3^{-1} A$. Since $\mu(A) = \mu(T_3^{-1} A)$, we have that $\mu(T_3^{-1} A \setminus A) = 0$.

Using these, it is possible to show that for $B = \bigcup_{n=0}^{\infty} T_3^{-n} A$, we have $T_2^{-1}(B) = B$ and $T_3^{-1} B$ and $\mu(B \setminus A) = 0$. By ergodicity $\mu(B) = 1$, hence $\mu(A) = 1$. Then

$$\frac{1}{N} \sum_{n=0}^{N_1} f(T_2^n x) \to \int f(x)\, dx$$

$\forall f \in C(X)$ and for μ-almost every x. $\square$

Index