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

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

Michaelmas 2018

Contents

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 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 $\mu(A) = \mu(T^{-1}A)$. $\square$

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 $T^n x \in A$?
- What is $\mu(\{\, x \in A \mid T^n x \in A \,\})$ if n is large? (Mixing property)

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 some $x = 0.x_1x_2x_3\ldots_2$, $x \in A$ corresponds to x_1, x_2 both being 0 and the doubling map sends x to $T_2 x = x_2x_3\ldots_2$, giving the required property above.

For example, $x = \frac{1}{6} = 0.00101010\ldots_2$ starts in A but never comes back to A. Also, we have $\mu(\{\, x \in A \mid T_2^n x \,\}) = \frac{1}{16}$ if $n \geq 2$.

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

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

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

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

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

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

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

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

Let

- $X = \{0, 1\}^{\mathbb{Z}}$
- $\mathcal{B}$ = Borel σ-algebra
- σ = the shift map $\mathbf{x} \mapsto (x_{n+1})_n$

Let $\mathbf{x}^S \in X$ be defined by

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

Also let $A \in \beta$ be given by $A = \{\, x \in X \mid x_0 = 1 \,\}$. Observe then that

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

Let $\{M_m\}$ and $\{N_m\}$ be sequences s.t. $M_n - N_m \to \infty$ and

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

Let

$$\mu_m = \frac{1}{M_m - N_m} \sum_{n=N_m}^{M_m-1} \delta_{\sigma^n \mathbf{x}^S}$$

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 how the μ could be different dependent on subsequence choice.

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

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

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.

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

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

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

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 $\mu(B) = \mu(\theta^{-1}B)$. $\square$

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 nice set of sets. So we leave full proof till a more general theorem.

Proposition. Let $S \subseteq \mathbb{Z}$, let $\mathbf{x}^S, A, (X, \mathcal{B}, \mu, \sigma)$ 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 $\mu = \lim \mu_m$ - 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 earlier 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 $\sigma^k \mathbf{x}^S \in B$. Hence

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

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.

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

where the inequality comes from satisfying the conditions of Furstenberg.

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

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

Then the set $A, T^{-1} A, \ldots$ are 'almost pairwise disjoint'. Then

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

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

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

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

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.

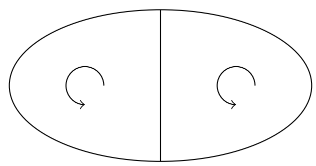
Proof. Let A_0 be the set of points in A that never return to A. We first show $\mu(A_0) = 0$. Note that $\mu(A_0 \cap T^{-n}A_0) \le \mu(A_0 \cap T^{-n}A) = \mu(\varnothing) = 0 \;\forall n > 0$. By the lemma, $\mu(A_0) = 0$. Note that if $x \in (A \setminus \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} T^{-n}A)$, then there is a maximal $m \in \mathbb{Z}_{\ge 0}$ such that $T^m x \in A$. This means that

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

where the right hand side has measure 0. □

This effectively answers one of the questions we asked earlier.

The main issue that can occur is that X splits into parts, which are preserved under T:



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 the MPS is not ergodic, and $A \in \mathcal{B}$ with $0 < \mu(A) < 1$ such that $T^{-1}A = A$, then we can restrict the MPS to A. That is, we consider the MPS: $(A, \mathcal{B}_A, \mu_A, T|_A)$ 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 $f : X \to \mathbb{R}$, $f = f \circ T$ a.e. implies f is constant a.e.

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

Proof. (1) ⇒ (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. $x \in B$ iff x visits A infinitely often $\iff$ Tx visits A infinitely often $\iff$ $T_x \in B$. So we have proved $B = T^{-1}B$.

(2) ⇒ (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) $\implies$ (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 $A_t = \{ x \in X \mid f(x) \le t \}$. 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 $f(x) = c$ a.e.

(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 $\chi(A)$ is constant a.e. □

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. $f(x) = \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n x)$.

$$\begin{aligned} f \circ R_\alpha(x) = f(x + \alpha) &= \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n(x + \alpha)) && (1) \\ &= \sum_{n \in \mathbb{Z}} a_n \exp(2\pi i n\alpha) \exp(2\pi i n x) && (2) \end{aligned}$$

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 $n \neq 0$, then $a_n = 0$.

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

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

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

Proposition. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $f \in L^1$. Let $f_0 = 0, f_1 = f, f_2 = f \circ T + f$,

$$f_n = f \circ T^{n-1} + \cdots + f \circ T + f,$$

and

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

Then

$$\int_{\{ x \in X \mid F_N(x) > 0 \}} f \, d\mu \ge 0 \; \forall N$$

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

$$\int_{\{ x \in X \mid F_N(x) > 0 \}} f(x) \, d\mu \ge \int_{\{ x \in X \mid F_N(x) > 0 \}} (F_n(x) - F_N(Tx)) \, d\mu$$

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

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

$\square$

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

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

Then

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

and $\int_{E_{\alpha,M}} \le \|f\|_1$. Note that $E_\alpha = \bigcup_M E_{\alpha,M}$, and this is an increasing union. $\square$

Proof of pointwise ergodic theorem. Fix $\epsilon > 0$. Then $\exists f_\epsilon \in L^2$, $e_{\epsilon,1} \in L^1$ such that $f = f_\epsilon + e_{\epsilon_1}$, and $\|e_{\epsilon,1}\| < \epsilon$. $\exists g_\epsilon \in L62$, $e_{\epsilon,2} \in L^1$ such that $f_\epsilon = P_T f_\epsilon + g_\epsilon \circ T - g_\epsilon + e_{\epsilon,2}$ and $\|e_{\epsilon,2}\|_1 < \epsilon$.

Also, $\exists h_\epsilon \in L^\infty$, $e_{\epsilon,3} \in L^1$ such that $g_\epsilon = h_\epsilon + e_{\epsilon,3}$ and $\|e_{\epsilon,3}\|_1 < \epsilon$.

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

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

Applying the maximal ergodic theorem for the function e_ϵ:

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

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 $F \subset \cup F_\alpha$ where

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

Notice $F_\alpha \subset E_{\epsilon,\alpha}$ for all $\epsilon > 0$. $\mu(F_\alpha) \le \mu(E_{\epsilon,\alpha}) \le \frac{4\epsilon}{\alpha}$. 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 $f^*(x) = f^*(Tx)$ a.e. For a.e. 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 $f^*(x) - f^*(Tx) = \lim \frac{1}{N} f(x) = 0$. □

Index