INTRODUCTION TO ERGODIC THEORY FOCUSED ON SZEMERÉDI'S THEOREM

YOUNG-HO AHN

ABSTRACT. This is a note of prelude lecture for Vitaly Bergelson's mini-course on ergodic Ramsey theory at KIAS in 2007.

Dynamical Systems

(i) Differentiable Dynamical Systems

$(X, \mathcal{B}, T)$ where $\mathcal{B}$ is the differential structure of X and T is a diffeomorphism on X.

(ii) Topological Dynamical Systems

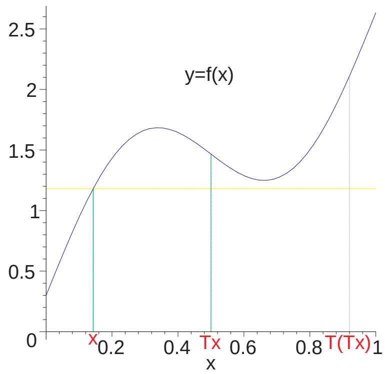
$(X, \mathcal{B}, T)$ where $\mathcal{B}$ is the topological structure on X and T is a homeomorphism(continuous map) on X.

(iii) Measure theoretical Dynamical Systems $\Longrightarrow$ Classical Ergodic Theory

$(X, \mathcal{B}, T)$ where $\mathcal{B}$ is a σ-algebra on X and T is a measurable transformation on X.

Let $(X, \mathcal{B}, \mu)$ be a probability measure space. A measurable transformation $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is said to be measure preserving if $\mu(T^{-1}E) = \mu(E)$ for every measurable subset E.

A measure preserving transformation T on $(X, \mathcal{B}, \mu)$ is called *ergodic* if $f(Tx) = f(x)$ holds only for constant function and it is called weakly mixing if the constant function is the only eigenfunction with respect to T. Imagine the definition of ergodicity by using the following picture.

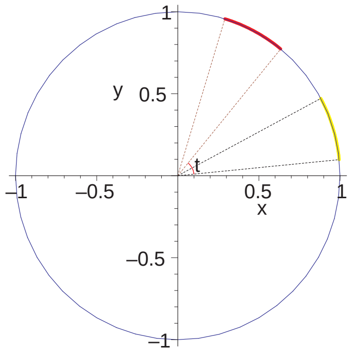


T is a transformation on $[0, 1]$

2000 *Mathematics Subject Classification.* 28D05, 47A35.

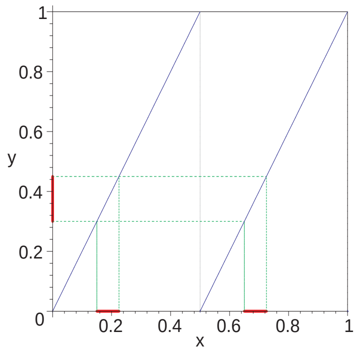
Example 1. Let $X = [0, 1)$, $\mathcal{B}$ Borel σ-algebra, and $\{x\}$ the fractional part of x.

(i) For a given irrational number α, let $T(x) = \{x + \alpha\}$ with Lebesgue measure.



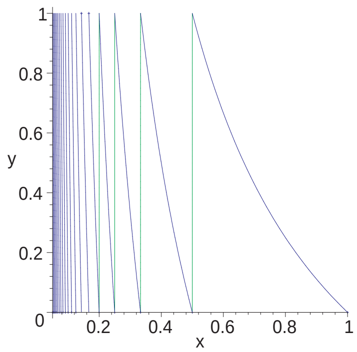
$$T : x \longmapsto \{x + \alpha\},\ t = \exp(2\pi i \alpha)$$

(ii) Let $T(x) = \{kx\}$ where $k = 2, 3, \cdots$ with Lebesgue measure.



$$T : x \longmapsto \{2x\}$$

(iii) Let $T(x) = \{\frac{1}{x}\}$ with $\mu(B) = \frac{1}{\ln 2} \int_B \frac{1}{1+x}\, dx$.



$$T : x \longmapsto \{1/x\}$$

(iv) Let σ be the shift map on $X = \prod_1^\infty \{0, 1\}$ with $(p, 1-p)$ Bernoulli measure where $0 < p < 1$.

$$(x_1, x_2, x_3, \cdots) \Longrightarrow (x_2, x_3, x_4 \cdots)$$

for $x = (x_1, x_2, x_3, \cdots)$.

Theorem 1. *A measure preserving transformation T on a probability space $(X, \mathcal{B}, \mu)$ is ergodic if and only if the only measurable sets with $\mu(T^{-1}E \triangle E) = 0$ are those with $\mu(E) = 0$ or $\mu(E) = 1$.*

Proof. Suppose $T^{-1}E = E$. Then $\mathbf{1}_{T^{-1}E}(x) = \mathbf{1}_E(Tx) = \mathbf{1}_E(x)$. Hence $\mathbf{1}_E(x)$ is constant. So $\mu(E) = 0$ or $\mu(E) = 1$.

For the converse, let $X(k,n) = \{x : \frac{k}{2^n} \leq f(x) < \frac{k}{2^{n+1}}\}$. Then we have

$$T^{-1}X(k,n) \triangle X(k,n) \subset \{x : f(Tx) \neq f(x)\}.$$

So $\mu(T^{-1}X(k,n) \triangle X(k,n)) = 0$. Thus $\mu(X(k,n)) = 0$ or $\mu(X(k,n)) = 1$ for all $k, n \in \mathbb{N}$. Hence $f(x)$ has to be constant. □

Definition 1. Let T be a continuous transformation on topological space X. A point $x \in X$ is a recurrent point for T if there exist an increasing sequence $n_1 < n_2 < \cdots$ such that $T^{n_k}x \to x$. In other words, for every open set V, there exist $n \geq 1$ with $T^n x \in V$

Theorem 2. *Let T be a measure preserving transformation on a probability measure space $(X, \mathcal{B}, \mu)$. If $\mu(A) > 0$ then there exist $n \geq 1$ such that $\mu(A \cap T^{-n}A) > 0$.*

Proof. Assume $\mu(A \cap T^{-n}A) = 0$ for all $n \geq 1$. Then $\mu(T^{-i}A \cap T^{-j}A) = 0$ for all $0 \leq i < j$. Hence $\bigcup_n T^{-n}A$ is a disjoint union. Since $\mu(A) = \mu(T^{-n}A)$ and

$$1 = \mu(X) \geq \mu\left(\bigcup_n T^{-n}A\right) = \sum_n \mu(T^{-n}A) = \sum_n \mu(A),$$

$\mu(A) = 0$. □

Remark 1. For a given measurable set A, let $\tilde{A} = A - \cup_j T^{-j}A$. Then $\mu(\tilde{A}) = 0$ by the previous theorem and the property $\tilde{A} \cap T^{-n}\tilde{A} \subset (T^{-n}A^c \cap T^{-n}A) = \phi$ for all $n \geq 1$.

Theorem 3. *Let T be a measure preserving transformation on a probability separable metric space $(X, \mathcal{B}, \mu)$. Then almost every point of X is recurrent for T.*

Proof. For a given $n \in \mathbb{N}$, let $\{B_{n,i}\}$ be a countable cover of X with $diam(B_{n,i}) < \frac{1}{n}$. Let

$$X_n = \bigcup_i \left(B_{n,i} - \bigcup_j T^{-j}B_{n,i}\right).$$

Then a nonrecurrent point belongs to some X_n . By the previous remark and countable subadditivity of measure, the conclusion follows. □

Remark 2. If T preserves a measure μ, then $\int_X f(Tx)\,d\mu(x) = \int_X f(x)\,d\mu(x)$ for all measurable functions. Because it is trivial for characteristic functions and step functions, and all measurable function can be approximated by step functions.

Theorem 4 (Mean Ergodic Theorem)**.** *Let T be a measure preserving transformation on a probability separable metric space $(X, \mathcal{B}, \mu)$. Let $M = \{f \in L^2(X) : f(Tx) = f(x)\}$. Then*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = P_M f(x)$$

where P_M is the orthogonal projection onto M

Proof. For a given measure preserving transformation T, let U be a linear operator on $L^2(X)$ defined by $Uf(x) = f(Tx)$. Then $\|Uf\| = \|f\|$. Hence $\|U\| \leq 1$, i.e. U is a contraction on a Hilbert space $L^2(X)$.

Let N be a closed subspace generated by $\{g - Ug; g \in L^2(X)\}$. At first, we will show that $M = N^\perp = \{h \in L^2(X); < h, f >= 0 \quad \text{for all } f \in N\}$ where $<,>$ is the inner product in $L^2(X)$. Assume $h \in N^\perp$. Then $0 =< h, g - Ug >=< h, g > - < U^*h, g >=< h - U^*g, g >$ for all $g \in L^2(X)$. Hence $h = U^*h$ and

$$\begin{aligned} \|Uh - h\|^2 &=< Uh - h, Uh - h > \\ &= \|Uh\|^2 - < h, Uh > - < Uh, h > + \|h\|^2 \\ &= \|h\|^2 - < U^*h, h > - < h, U^*h > + \|h\|^2 \\ &= \|h\|^2 - < h, h > - < h, h > + \|h\|^2 = 0. \end{aligned}$$

Thus we have $Uh = h$ and $M \supset N^\perp$.

Conversely, take $h \in M$. Then $Uh = h$, and $U^*h = h$ by exactly the same arguments as in U. Hence $< h, g - Ug >=< h, g > - < h, Ug >=< h, g > - < U^*h, g >=< h - U^*h, g >= 0$ for all $g \in L^2(X)$. So $M \subset N^\perp$.

To prove the theorem, we only need to show that for all $f(x)$ with the form $f(x) = g(x) - g(Tx)$,

$$\frac{1}{n} \sum_{k=0}^{n-1} U^k f \to 0$$

by continuity property of U. This is trivial, since

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \right\| = \left\| \frac{1}{n}(g - U^n g) \right\| \leq \frac{2}{n}\|g\|.$$

So we complete the proof. □

Theorem 5 (Birkhoff Ergodic Theorem)**.** *Let T be a measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ and $f \in L^p, 1 \leq p < \infty$. Then $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges to $\overline{f}(x)$ a.e. and L^p-norm where $\overline{f}(x)$ is a T-invariant function.*

Remark 3. Since $\int_X \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\, d\mu(x) = \int_X f(x)\, d\mu(x)$, if T is an ergodic measure preserving transformation, then $\overline{f}(x) = \int_X f(x)\, d\mu(x)$.

Theorem 6. *T is an ergodic measure preserving transformation if and only if*

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \mu(A)\,\mu(B)$$

for all measurable set $A, B \in \mathcal{B}$.

Proof. Let $f(x) = \mathbf{1}_A(x)$. Then $\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_A(T^k x)$ converges to $\mu(A)$. So $\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_A(T^k x)\,\mathbf{1}_B(x)$ converges to $\mu(A)\,\mathbf{1}_B(x)$. Hence by Dominated Convergence Theorem,

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \mu(A)\mu(B).$$

Conversely, suppose that there exist E with $T^{-1}E = E$. Put $A = B = E$ in the convergence property. Then we have $\frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \frac{1}{n}\sum_{k=0}^{n-1} \mu(E) \to \mu(E)^2$. Hence $\mu(E) = \mu(E)^2$ and $\mu(E) = 0$ or 1. □

Theorem 7 (Borel's Normal Number Theorem)**.** *Let $x = 0.x_1, x_2, \cdots$ be a dyadic expansion of x. Then the relative frequency of 1 in dyadic expansion of x is $\frac{1}{2}$ almost everywhere.*

Proof. Note that the map $T(x) = 2x \pmod 1$ is ergodic with respect to Lebesgue measure. Let $E = [\frac{1}{2}, 1)$ and apply Birkhoff Ergodic Theorem to $f(x) = \mathbf{1}_E(x)$. Then we have $\overline{f}(x) = \int_X \mathbf{1}_E(x)\, d\mu(x) = \mu(E) = \frac{1}{2}$. Hence

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \overline{f}(x) = \frac{1}{2}.$$

Since $x_k = 1$ if and only if $\mathbf{1}_E(T^{k-1}(x)) = 1$, the conclusion follows. □

Definition 2. Let T be a continuous transformation on the space X and μ be a T-invariant measure. We say that a point x_0 is a generic point for μ if

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

for every continuous function $f \in C(X)$.

By Birkhoff Ergodic Theorem, we have the following proposition.

Proposition 1. *If T is a continuous map on the compact metric space X and μ is an ergodic measure, then almost every point of X (with respect to μ) is generic for μ.*

Definition 3. A topological dynamical system (X, T) is uniquely ergodic if T has a unique invariant probability measure.

Theorem 8. *If (X, T) is uniquely ergodic and μ is T-invariant measure, then $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ uniformly converges to $\int_X f(x)\, d\mu(x)$ for all continuous function. Hence all points are generic.*

Proof. Suppose the conclusion is not true. Then there exist a continuous function $g(x)$ and ϵ and a sequence of points $\{x_n\}$ with

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} g(T^k x_n) - \int_X g\, d\mu\right| \geq \epsilon.$$

Let $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k x_n}$ where δ_t is Dirac delta functional at t. Then $|\int g\, d\mu_n - \int g\, d\mu| \geq \epsilon$. By Banach-Alaoglu theorem, there exist a subsequence n_i and T-invariant measure μ_∞ such that $\mu_{n_i} \to \mu_\infty$. But $|\int g\, d\mu_\infty - \int g\, d\mu| \geq \epsilon$. So $\mu_\infty \neq \mu$. It contradicts to the unique ergodicity. □

Remark 4. Indeed, (X, T) is uniquely ergodic if and only if $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges pointwise to a constant for all continuous function $f \in C(X)$ [3].

Remark 5. It is easy to show that an irrational rotation on $[0, 1)$ is uniquely ergodic by using the above remark and isometry property of rotations.

Definition 4. Let X be a compact metric space, T a continuous transformation of X, and μ a T-invariant measure. A point $x_0 \in X$ is quasi-generic for μ if for some sequence of pairs of integers $\{a_k, b_k\}$ with $a_k \leq b_k$ and $b_k - a_k \to \infty$,

$$\frac{1}{b_k - a_k + 1}\sum_{n=a_k}^{b_k} f(T^n x_0) \to \int_X f\, d\mu$$

as $k \to \infty$, for every continuous function $f \in C(X)$.

Proposition 2. *Let T be a continuous transformation on the compact metric space X. For $x_0 \in X$, let $Y = \overline{\{T^n x_0 : n \geq 0\}}$. If μ is an T-invariant ergodic measure with its topological support is Y, then x_0 is quasi-generic for μ.*

Proof. Since μ is ergodic, there exist a point x_1 which is generic for μ. Then for each $f \in C(X)$,

$$\frac{1}{n+1}\sum_{i=0}^{n} f(T^i x_1) \to \int_X f\, d\mu.$$

Let $\{f_k\}$ be a dense subset of functions in $C(X)$, and n_k be an increasing sequence with

$$\left|\frac{1}{n_k+1}\sum_{i=0}^{n} f_j(T^i x_1) - \int_X f_j\, d\mu\right| < \frac{1}{k} \quad \text{for all} \quad 1 \le j \le k.$$

The above inequality still holds if x_1 is replaced by a sufficiently near point, say some $T^{a_k}x_0$. In other words,

$$\left|\frac{1}{n_k+1}\sum_{i=a_k}^{a_k+n_k} f_j(T^i x_1) - \int_X f_j\, d\mu\right| < \frac{1}{k} \quad \text{for all} \quad 1 \le j \le k.$$

Now let $b_k = a_k + n_k$, then the conclusion follows. □

References

[1] H. Furstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory, Princeton Univ. Press, 1981.

[2] K. Petersen, Ergodic Theory, Cambridge Univ. Press, 1983.

[3] P. Walters, An Introduction to Ergodic Theory, Springer-Verlag, New York, 1982.

Department of Mathematics, Mokpo National University, 534-729, South Korea

E-mail address: yhahn@mokpo.ac.kr