Name: Tianyu Zhang and Xuankai Wang
Student ID:
Date: 2025-02-27

Lecture Notes for Ergodic Theory (Week 2)

This note is divided into two parts. In the first section, we give two proofs of Krylov-Bogoliubov theorem with the only difference being using weak convergence and using regular Borel measure property. Then we prove the Kac's lemma. In the second section, we introduce the isomorphism between measure preserving systems, and introduce the ergodicity of this system together with a criterion and several important examples.

1 Krylov-Bogoliubov Theorem and Kac's Lemma

We start with two proofs of the famous Krylov-Bogoliubov theorem, also known as the easy Kakutani theorem, which states that given a compact metric space and a continuous transformation (not necessarily measure-preserving), there exists a Borel probability measure that is invariant under this transformation.

Theorem 1.1 (Krylov-Bogoliubov Theorem)**.** *Let (X, d) be a compact metric space and $T : X \to X$ is a continuous transformation. Then there exists a Borel probability measure μ with $\mu \circ T^{-1} = \mu$, i.e., a invariant Borel probability measure.*

Proof. Let $\mu \in \Delta(X)$, where $\Delta(X)$ denotes the space of all Borel probability measures on X. Define

$$\nu_n := \frac{\nu + \nu \circ T^{-1} + \nu \circ T^{-2} + \cdots + \nu \circ T^{-(n-1)}}{n}. \tag{1.1}$$

Since $\Delta(X)$ is convex then $\nu_n \in \Delta(X)$ as well. We need to show that for a suitable subsequence $\{n_k\} \subseteq \mathbb{N}$ we have $\nu_{n_k} \xrightarrow[k\uparrow\infty]{\text{weak*}} \mu$.

Let $C(X)$ be the collection of all continuous real-valued functions f defined on X and let $C_b(X)$ denote the space of all bounded real-valued functions f defined on X. Since X is compact $C_b(X) = C(X)$.

It suffices to show that $\Delta(X)$ is weak*-compact, that is, given any $\{\nu_n\}$ we can find a weak*-convergence sequence. Let $\{f_j\}_{j\geq 1}$ be a countable dense sequence in $\Delta(X)$, if for a subset $S_1 \subseteq \mathbb{N}$ such that

$$\int f_1 d\mu_n \xrightarrow[n\in S_1]{} \text{limit}.$$

Then for $S_2 \subseteq S_1$ we have

$$\int f_2 d\mu_n \xrightarrow[n\in S_2]{} \text{limit}.$$

Continue this fashion, we obtain that for every $j \geq 2$, for $S_k \subseteq S_{k-1}$,

$$\int f_j d\mu_n \xrightarrow[n\in S_j]{} \text{limit}.$$

Let $n_j \in \{j\text{-th element of } S_j\}$ then $\forall i \geq 1$,

$$\int f_i d\mu_{n_j} \xrightarrow[j\uparrow\infty]{} \text{limit}.$$

Denote this limit as $\psi(f)$, an application of Riesz's Representation theorem guarantees that there exists a measure $\mu \in \Delta(X)$ such that $\psi(f) = \int f d\mu$, thus the weak*-compactness of $\Delta(X)$ follows.

Finally, we have

$$\nu_n \circ T^{-1} - \nu_n = \frac{\nu \circ T^{-n} - \nu}{n} \xrightarrow[n\uparrow\infty]{\text{weak*}} 0$$

where the first equality holds by 1.1. Moreover, since

$$\begin{aligned}\int f(d\nu_{n_k} \circ T^{-1}) &= \int (f \circ T) d\nu_{n_k} \\ &\xrightarrow[k\uparrow\infty]{\text{weak*}} \int (f \circ T) d\mu \\ &= \int f d(\mu \circ T^{-1})\end{aligned}$$

where in the first and the last equalities we used the fact that $\int f d(\mu \circ T^{-n}) = \int f \circ T^n d\mu$ for all bounded f, and in the convergence we used the assumption that T is continuous. Therefore,

$$\nu_{n_k} \circ T^{-1} \xrightarrow{\text{weak*}} \mu \circ T^{-1}$$

as desired. □

We also present another version of the proof, where Riesz's representation theorem is not applied to guarantee the weak*-convergence, but to guarantee the existence of regular Borel measure.

Proof. (Alternative proof) Let $x \in X$ be fixed and define

$$S_f^N(x) := \frac{1}{N} \sum_{n=0}^{N-1} f\big(T^n(x)\big). \tag{1.2}$$

Let $\mathcal{F}$ be a countable dense subset of $C(X, \mathbb{R})$, the collection of all continuous functions $f : X \to \mathbb{R}$. Then a diagonal argument, together with the fact that $f(X)$ is bounded (since $C_b(X, \mathbb{R}) = C(X, \mathbb{R})$ if X is a compact metrizable space, where $C_b(X, \mathbb{R})$ is the collection of all bounded continuous functions from X into $\mathbb{R}$). It is easy to prove that there exists a strictly increasing sequence N_k such that $S_f^{N_k}(x)$ converges $\forall f \in \mathcal{F}$.
Since $\mathcal{F}$ is dense in the uniform topology, $S_g^{N_k}(x)$ also converges $\forall g \in C(X, \mathbb{R})$. Now let

$$S_g(x) := \lim_{k\uparrow\infty} S_g^{N_k}(x).$$

and denote $L_x(g) := S_g(x)$. Then $L_x : C(X, \mathbb{R}) \to \mathbb{R}$ is a continuous positive linear functional.

Now we can use Riesz's representation theorem to show that there is a regular measure μ on $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra, such that

$$L_x(g) = \int g d\mu$$

for every $g \in C(X, \mathbb{R})$. A simple computation shows that

$$L_x(g \circ T) = L_x(g), \forall g \in C(X, \mathbb{R}), \tag{1.3}$$

which holds by the continuity assumption of T (Indeed, L_x acts on continuous functions, and we need continuity of T to guarantee that so is $g \circ T$). Thus

$$\int_X g d\nu = \int_X g \circ T d\mu = \int_X g d\mu.$$

where $\nu := \mu \circ T^{-1}$. Since ν is a finite measure and X is a compact metrizable space, then $\mu = \nu$, therefore μ is T-invariant as desired. □

Now we introduce the Kac's lemma, which describes the expected return time.

Theorem 1.2 (Kac's Lemma). *Let $(X, \mathcal{F}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{F}$. Define for every $x \in X$*

$$\tau(x) = \tau_A(x) := \inf\{n \geq 1 : T^n(x) \in A\},$$

with $\inf \emptyset = \infty$ *by convention. Then*

$$\int_A \tau(x) d\mu = 1 - \mu(B_\infty), \tag{1.4}$$

where $B_\infty := \{x \in X \setminus A| \text{ where } \tau(x) = \infty\}$.

The restriction $\tau|_A$ is called **return time map** and the Poincáre recurrence theorem tells us that $\tau(x) < \infty$ for μ-almost every $x \in A$.

Proof. Let $A \setminus A_\infty = \cup_{k \geq 1} A_k$, where $A_k := \{x \in A | \tau(x) = k\}$ also for $k = \infty$. Recall that $\mu(A_\infty) = 0$. Similarly, we denote $X \setminus A = B$ and observe that

$$B \setminus B_\infty = \bigcup_{k \geq 1} B_k,$$

where $B_k := \{x \in B | \tau(x) = k\}$. Thus one has

$$1 - \mu(B_\infty) = \mu(X \setminus B_\infty) = \sum_{k \geq 1} \left[\mu(B_k) + \mu(A_k)\right],$$

where the summation in the last equality is over finite values.

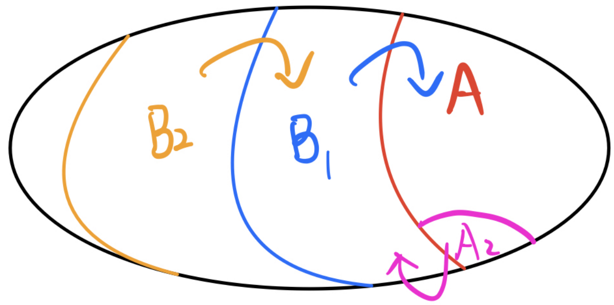

Figure 1: $T^{-1}(B_1) \subseteq B_2 \uplus A_2$ for disjointed B_2 and A_2

We have $T^{-1}(B_1) \subseteq B_2 \uplus A_2$, where $\uplus$ is to stress the "disjoint union". Thus, $T^{-1}(B_k) = B_{k+1} \uplus A_{k+1}$. Since T is measure preserving, applying the measure to each side gives

$$\begin{aligned}\mu(B_k) &= \mu(B_{k+1}) + \mu(A_{k+1}) \\ &= \mu(B_{k+2}) + \mu(A_{k+2}) + \mu(A_{k+1}) \\ &= \cdots \\ &= \mu(B_{k+\ell}) + \sum_{j=1}^{\ell} \mu(A_{k+j})\end{aligned}$$

by induction on ℓ. Therefore, letting $\ell \uparrow \infty$ and the fact that $1 - \mu(B_\infty)$ is convergent give us

$$\mu(B_k) = \sum_{\ell=1}^{\infty} \mu(A_{k+\ell}) = \sum_{r>k} \mu(A_r).$$

It follows that

$$\begin{aligned}1 - \mu(B_\infty) &= \sum_{k=1}^{\infty} \sum_{r \geq k} \mu(A_r) \\ &= \sum_{r=1}^{\infty} r \cdot \mu(A_r) \\ &=: \int_A \tau d\mu\end{aligned}$$

since the last equality is just the definition of Lebesgue integral by noting that $\mu(A_r) = \mu\big(\{x \in A | \tau(x) = r\}\big)$. □

Now we give some examples. Let $X = [0, 1)$ and $\mathcal{F} := \mathcal{B}(X)$ be the Borel σ-algebra on X. Then $T_x := 2x \mod 1$ is a measure-preserving system with the measure $\mu = \text{Leb}$, the Lebesgue measure on $[0, 1)$. Indeed, since

$$T_x^{-1}[a, b) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a+1}{2}, \frac{b+1}{2})$$

for $0 \leq a < b < 1$. We have

$$\mu \circ T^{-1}[a, b) = \frac{b-a}{2} + \frac{(b+1) - (a+1)}{2} = b - a = \mu[a, b).$$

Let $I := \bigcup_i I_i$ be a finite or countable union of disjoint intervals $[a_i, b_i)$ we have

$$\mu(I) = \sum_i |b_i - a_i|$$

and $\mu \circ T_x^{-1}(I) = \mu(I)$ holds for all I, then a monotone class argument guarantees that the same holds for $I \in \mathcal{B}(X)$.

2 Isomorphic Measure Preserving System and Ergodicity

In this section we study the isomorphism between measure preserving systems and the ergodicity along with a criteria. We start with the definition of homomorphism.

Definition 2.1. *A* **homomorphism** $\psi : (X, \mathcal{F}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ *between two measure preserving systems is a measurable map* $\psi : X \to Y$*. That is,* $\psi^{-1}(B) \in \mathcal{F}$ *for every* $B \in \mathcal{B}$ *and* $\mu \circ \psi^{-1} = \nu$ *on* B*.*

The diagram
$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ {\scriptstyle\psi}\big\downarrow & & \big\downarrow{\scriptstyle\psi} \\ Y & \xrightarrow[S]{} & Y \end{array}$$
where $\psi \circ T = S \circ \psi$*, is also called a* **factor map** *in ergodic theory.*

For example, if $T : [0,1)^2 \to [0,1)^2$ where $T(x_1, x_2) = (2x_1, 3x_2)$ and $\psi(x_1, x_2) = x_1$ then ψ is a factor map. In general, it suffices that $\psi : X_1 \to Y$ where $\mu(X \setminus X_1) = 0$ for $X_1 \in \mathcal{F}$, that is, the homomorphism only needs to be defined μ-almost everywhere.

Definition 2.2. *A factor map* ψ *is called an* **isomorphism** *between measure preserving systems* $(X, \mathcal{F}, \mu, T)$ *and* $(Y, \mathcal{B}, \nu, S)$ *if there exists a factor map*

$$\widetilde{\psi} : (Y, \mathcal{B}, \nu, S) \to (X, \mathcal{F}, \mu, T)$$

(which can be defined ν*-almost everywhere) such that* $\widetilde{\psi} \circ \psi = Id_X$ μ*-almost everywhere and* $\psi \circ \widetilde{\psi} = Id_Y$ ν*-almost everywhere.*

Let $\mathbb{S} := \{|z| = 1 \big| z \in \mathbb{C}\}$ be the unit circle in $\mathbb{C}$, $\mathcal{B}$ be the Borel sigma-algebra on $\mathbb{S}$, μ be the uniform measure on $\mathbb{S}$, and $S(z) := z^b$ be a transformation. Then this measure preserving system is isomorphic to $([0,1], \mathcal{F}, \text{Leb}, T_b)$, where $T_b(x) := bx \mod 1$. Indeed, the transformation can be defined by $\psi : [0,1) \to \mathbb{S}$, where $\psi(x) = e^{2\pi i x}$.

Another example is the full shift on b-symbols. Let $A = \{0, 1, 2, \ldots, b-1\}$ be a finite alphabet and $\mathbb{W} = A^{\mathbb{N}}$ denote the space of all infinite sequences of symbols from A. The **left shift transformation** $S : \mathbb{W} \to \mathbb{W}$ is defined as: $S(w) := (a_2, a_3, \ldots)$ where $w = (a_1, a_2, \ldots) \in \mathbb{W}$. The measure ν on $\mathbb{W}$ is the product measure $(\frac{1}{b}, \frac{1}{b}, \ldots, \frac{1}{b})^{\mathbb{N}}$. Define the mapping $\psi : \mathbb{W} \to [0,1)$ as: $\psi(a_1, a_2, a_3, \ldots) = \sum_{i=1}^{\infty} a_i b^{-i}$. This mapping takes each sequence $w = (a_1, a_2, \ldots)$ to a b-ary expansion of a real number in $[0,1)$ and one can verify that this is indeed an isomorphism.

Definition 2.3. *A measure preserving system* $(X, \mathcal{B}, \mu, T)$ *is said to be* ***ergodic*** *if* $\forall A \in \mathcal{B}$ *such that* $T^{-1}(A) = A$ *we have* $\mu(A) \in \{0, 1\}$.

Remark 2.4. Suppose in a measure preserving system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$ such that $\mu(A \Delta T^{-1}A) = 0$ then there exists an $\widetilde{A}$ that is T-invariant, that is, $T^{-1}\widetilde{A} = \widetilde{A}$ and $\mu(\widetilde{A}\Delta A) = 0$. Thus the invariance only needs to be defined almost everywhere.

It is useful sometimes to give some criteria for ergodicity. We here offer a stronger terminology called mixing.

Definition 2.5. *An measure preserving system* $(X, \mathcal{B}, \mu, T)$ *is called* ***mixing*** *if* $\forall A \in \mathcal{B}$ *we have* $\mu(T \cap T^{-n}A) \underset{n\uparrow\infty}{\longrightarrow} \mu(A)^2$.

To verify mixing it suffices to verify on an algebra that generates $\mathcal{B}$, or on any dense set in the metric $\mu(A\Delta B)$ on $\mathcal{B}$ (recall that $\mu(A\Delta B)$ identifies a metric). Then using

$$\mu\Big(\big((A \cap T^{-n}(A)\big)\Delta\big((B \cap T^{-n}(B)\big)\Big) \leq \mu(A\Delta B) + \mu\big(T^{-n}(A)\Delta T^{-n}(B)\big) \leq 2\mu(A\Delta B). \quad (2.1)$$

Using mixing we can also use very nice sets that approximate every sets. Recall the shift S on $A^{\mathbb{N}}$ where $A = \{0, 1, \cdots, b-1\}$. A rich class of measure preserving system is obtained by taking $X \subset \mathbb{A}^{\mathbb{N}}$ that is S-invariant and closed in the product topology (hence compact). Then X is a compact metric space under the metric

$$d\big((x_1, x_2, \cdots), (z_1, z_2, \cdots)\big) := \frac{1}{\exp\big\{\min\{n | x_n \neq z_n\}\big\}}$$

and (X, S) is then called a **subshift**.

An example of the subshift is the Golden shift. By taking $X = \{\omega \in \{0,1\}^{\mathbb{N}} | \omega_j \omega_{j-1} = 0 \forall j \geq 1\}$. Indeed, the number of allowed sequences of length n in the Golden Shift is exactly $F_{n+2} = F_{n+1} + F_n$ and $\sqrt[n]{F_n} \underset{n\uparrow\infty}{\longrightarrow} \varphi = \frac{1+\sqrt{5}}{2}$, which is the golden ratio.

Lemma 2.6. *Mixing implies Ergodicity.*

Proof. If $A = T^{-1}A$ then $\mu(A) = \mu(A \cap T^{-n}A) \underset{n\uparrow\infty}{\longrightarrow} \mu(A)^2$ which is either 0 or 1. □

The full shift with any product measure is mixing If:

$$A := \big\{\omega \in \mathbb{W} = \{0, \cdots, b_n\}^{\mathbb{N}} \big| (\omega_1, \cdots, \omega_n) \in \widetilde{A}_n \subset \{0, \cdots, b_{n-1}\}^n\big\},$$

where $\widetilde{A}_n$ is a subset of finite sequences of length n, such sets are called cylinder sets and generate the sigma-algebra of measurable sets in $\mathbb{W}$. Then $\mu(A \cap T^{-m}A) = \mu(A)^2$ for every $m > n$. In fact, we can also deduce the ergodicity of the full-shift from Kolmogorov's 0-1 law.

We finally obtain that $T_b(x) := bx \mod 1$ on $[0, 1)$ is ergodic by the isomorphism. In fact, if X is ergodic and $\psi : X \to Y$ is a factor map, then it suffices to conclude that Y is also ergodic.