Explorations in Ergodic Theory: A Student's Perspective

Kaining Jia
Supervisor: Professor Tserunyan Anush

30th, August, 2024

Abstract

This Honours undergraduate research project aims to provide research experience in ergodic theory. The student learned the Birkhoff Ergodic Theorem and its proof by Professor Tserunyan Anush. One of the primary reading materials was a 2015 paper by Lewis Bowen and Amos Nevo on Amenable Equivalence Relations and the Construction of Ergodic Averages for Group Actions. One of the main goals of the course was to understand the method for constructing pointwise ergodic sequences on countable groups. During the reading process, the student was introduced to ergodic theory, orbit equivalence theory, and other areas of mathematics. This report summarizes some of the concepts and topics that arose naturally.

1 Preliminary Definitions

- **Definitions 1.1**: Let $(X, \mathcal{B}, \mu)$ be a standard probability space and let $T : X \to X$ be a measurable transformation.T is called measure-preserving with respect to the measure μ if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.

  Three classical examples of measure-preserving transformations include the *Baker's map*, the *shift map*, and the *Gauss map*.

- **Definitions 1.2**

  1. A set $D \subseteq \mathcal{B}$ is T-invariant if $T^{-1}(D) = D$.
  2. A function $f : X \to \mathbb{R}$ is called T-invariant if $f = f \circ T$. Equivalently, $f^{-1}(A)$ is T-invariant for each $A \in \mathbb{R}$.

- **Definition 1.3** Let T be a measure-preserving transformation (p.m.p) on a probability space (X, μ). The map $T : X \to X$ is known as ergodic if for every T-invariant measurable set A, $\mu(A) = 0$ or 1.
  Theorem 1.1 Let $(X, \mathcal{B}, \mu)$ be a probability space and T measure preserving. TFAE:

  (i) T is ergodic.
  (ii) If $B \in \mathcal{B}$ with $\mu(T^{-1}(B)\Delta B) = 0$, then $\mu(B) = \{0, 1\}$.
  (iii) If $A \in \mathcal{B}$ with $\mu(A) > 0$, then $\mu\left(\bigcup_{n=1}^{\infty} T^{-n}(A)\right) = 1$.

(iv) If $A, B \in \mathcal{B}$ with $\mu(A) > 0$ and $\mu(B) > 0$, then there exists $n > 0$ such that $\mu(T^{-n}(A) \cap B) > 0$.

The proof for the theorem is omitted.

Note: Examples of ergodic transformations include irrational rotations and the baker's map.

- **Definition 1.4** The action of a countable group Γ on X is free if for all $x \in X$, $\Gamma_x = \{e\}$, i.e., the only stabilizer of x is the identity in Γ. Equivalently, $\gamma x \neq x$ whenever $\gamma \neq 1$.

2 Learning Outcome

2.1 Wondering Sets and Poincaré's Recurrence Theorem

Here we present the proof of the Poincaré recurrence theorem via wondering sets contradiction argument.

Definitions: Let (X, μ) be a standard probability space and let $T : X \to X$ be a measurable p.m.p. transformation.

1. A **wandering set** is a set $W \subseteq X$ such that the sets $T^n(W)$ are pairwise disjoint (more formally, $\mu(T^n(W) \cap T^m(W)) = 0$ for all disjoint $n, m \in \mathbb{Z}$).
2. **The Poincaré Recurrence Theorem** states that for any measurable $A \subseteq X$ with positive measure, almost every $x \in A$ returns to A after some iteration of T (i.e., for almost every $x \in A$, there exists $n > 0$ such that $T^n(x) \in A$).

Proof: We may prove the Poincaré Recurrence Theorem using the wandering set and via a contradiction argument. To start, we construct a wandering set F that consists of points of A which are not A-recurrent:

$$F = \{x \in A \mid T^n(x) \notin A \text{ for all } n \geq 1\}.$$

Now suppose $\mu(F) > 0$. Then observe that F is a wandering set (since $T^{-n}(F) \cap T^{-m}(F) = \emptyset$ for all $n \neq m$). Then we have:

$$1 = \mu(X) \geq \mu\left(\bigcup_{n=0}^{\infty} T^{-n}(F)\right) = \sum_{n=0}^{\infty} \mu(T^{-n}(F)) = \sum_{n=0}^{\infty} \mu(F) = \infty,$$

which is a contradiction. Thus we have shown that $\mu(F) = 0$ and therefore almost every x of A is A-recurrent, and the Poincaré Recurrence Theorem holds.

2.2 Classical Birkhoff Ergodic Theorem

An important part of the research project was the introduction to Professor Tserunyan Anush's proof of the classical pointwise Birkhoff ergodic theorem via the method of tiling. The outline of the proof is presented here.

Theorem: A probability measure preserving transformation $T : X \to X$ is ergodic if and only if for all $f \in L^1(X, \mu)$, and for almost every $x \in X$,

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

Proof: ($\Rightarrow$)
We can prove this theorem using the Local-Global Bridge lemma and the tiling argument. The Local-Global Bridge lemma states that for all $n \in \mathbb{N}$,

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

From now on, we shall define $A_n f(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$. Intuitively, $A_n f(x)$ is the average of $f(x)$ over $x, T(x), T^2(x), \ldots, T^{n-1}(x)$.

Note that since T is a probability measure preserving map, for all measurable subsets $B \subseteq X$,

$$\int_X \mathbf{1}_B(x) \, d\mu = \int_X \mathbf{1}_B(T(x)) \, d\mu,$$

where $\mathbf{1}_A(x)$ is the indicator function for the set A.

This holds as a result of the change of variable formula. Furthermore, since T is probability measure preserving, for all $f \in L^1(X, \mu)$,

$$\int_X f \, d\mu = \int_X f \circ T \, d\mu.$$

$$\int_X f \circ T \, d\mu = \int_X \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(T(x)) \, d\mu = \sum_{i=1}^{n} a_i \mu(T^{-1}(A_i)) = \sum_{i=1}^{n} a_i \mu(A_i) = \int_X f \, d\mu.$$

If $\int_X f \, d\mu = \int_X f \circ T \, d\mu$, then by induction, $\int_X f \, d\mu = \int_X f \circ T^n \, d\mu$ for any $n \in \mathbb{N}$. Then,

$$\int_X f \, d\mu = \frac{1}{n} \int_X \sum_{i=0}^{n-1} f \circ T^i \, d\mu.$$

Thus we have proven the Local-Global Bridge lemma.

Now observe that $\limsup_{n\to\infty} A_n f$ and $\liminf_{n\to\infty} A_n f$ are both T-invariant. Note that

$$A_{n+1} f(x) = \frac{1}{n+1} f(x) + \frac{n}{n+1} A_n f(T(x))$$

and

$$\limsup_{n\to\infty} A_n f(T(x)) = \limsup_{n\to\infty} A_{n+1} f(x) = \limsup_{n\to\infty} A_n f(x).$$

The T-invariance of $\liminf_{n\to\infty} A_n f(x)$ is shown similarly.

One of the key characterizations of ergodicity is that if the probability measure preserving transformation T is ergodic and f is a T-invariant function, then $f(x)$ is constant almost everywhere (i.e., $f(x) = c$ a.e.).

Combining the above results, we obtain that $\limsup_{n\to\infty} A_n f$ and $\liminf_{n\to\infty} A_n f$ are constant a.e. Let us assume for simplicity and without loss of generality that f is bounded and $\int_X f\, d\mu = 0$.
To prove the theorem, we want to show that $\limsup_{n\to\infty} A_n f = \liminf_{n\to\infty} A_n f = \lim_{n\to\infty} A_n f = 0$. Suppose that $\limsup_{n\to\infty} A_n f = c$ and c is a positive constant. Define $\Delta = \frac{c}{2}$ and $\epsilon = \min\left(\frac{c}{16}, \frac{1}{8}\right)$. We can construct a set Z which consists of points $x \in X$ where the average over the first L steps (for some large L) is all smaller than $\frac{c}{2}$. Formally,

$$Z = \{x \in X \mid A_n f(x) < \frac{c}{2} \text{ for all } n \in [1, L]\}.$$

Note that L is large enough that $\|1_Z\| + \|f \circ 1_Z\|_1 < \epsilon$. (i.e., the set Z is small enough such that the set Z supports less than ϵ of the total weight of 1 and f). Now we will define tiling. For all $x \in X$, define $\ell(x) :=$ the smallest integer $n \leq L$ such that the average of f from x to $T^n(x)$ is greater than Δ. Formally,

$$\frac{1}{\ell(x)} \sum_{i=0}^{\ell(x)-1} f(T^i(x)) > \Delta,$$

and if no such n exists, then $\ell(x) = 1$.

Let $I_x = [x, T^{\ell(x)}(x))_T$ be called a **tile** and A T-interval $I = [y, z)$ is tiled if there is a partition into T-intervals in the form of I_x for x. Observe that such a partition is naturally unique because I_x necessarily would be the tile containing x.

Now we can introduce a corollary to help us proceed with the proof.

Corollary: Let T be an ergodic, measure-preserving transformation on a probability space (X, μ). For an arbitrary positive integer L, there exists a Borel set S with arbitrarily small measure such that:

1. The set S does not overlap with any of its first L images under the transformation T. (i.e., S is disjoint from the sets $T(S)$, $T^2(S)$, …, $T^L(S)$).

2. The set S intersects almost every orbit of the transformation T in an infinite number of points in both directions.

The proof of the corollory uses Luzin–Novikov uniformization theorem and is omitted here.
By the measure-preserving nature of the transformation T, the set S can be made arbitrarily small, and the union $\bigcup_{i=1}^{L} T^{-i}(S)$ supports less than ϵ of the total mass of 1 and the function f. Define $s(x)$ to be the closest element of S to the left of x. Define a partial finite equivalence relation F on X by $x \sim_F y$ if and only if there exists $z \notin \bigcup_{i=1}^{L} T^{-i}(S)$ such that $x, y \in I_z$ and $[s(z), z)$ is tiled.

If we let Y be the domain of the partial finite equivalence relation F on X, then we have

$$\int_{Y^c} d\mu + \int_{Y^c} |f|\, d\mu < 2\epsilon.$$

Then

$$0 = \int_X f\,d\mu \geq \int_Y f\,d\mu - 2\epsilon = \int_Y A_N(f)\,d\mu - 2\epsilon > \frac{\delta}{2} \neq 0.$$

Thus, we have our contradiction, and the proof is complete.

($\Leftarrow$) In this direction, we want to show that for any T-invariant measurable set A, $\mu(A) = 0$ or 1. Define the function $f(x) = 1_A(x)$. Since A is a T-invariant set, for any $i \geq 0$, we have

$$f(T^i(x)) = 1_A(T^i(x)) = 1_A(x) = f(x).$$

Then,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) = \frac{1}{n}\sum_{i=0}^{n-1} 1_A(x) = 1_A(x).$$

By assumption, this equals $\int f\,d\mu = \int 1_A\,d\mu = \mu(A)$. Thus, for almost every $x \in X$, $1_A(x) = \mu(A)$. Therefore, $\mu(A) \in \{0,1\}$, and we have proven that T is an ergodic transformation.

2.3 Amenable Groups

Consider (X, μ) where μ is a finitely additive probability measure on X. μ is called **left-invariant** if $\forall \gamma \in \Gamma$, for all sets $A \subseteq \Gamma$, we have $\mu(\gamma * A) = \mu(A)$.

Famously known as the Tarski theorem, it states that Γ **is not amenable if and only if** Γ **is paradoxical** (i.e., there exist disjoint sets $A, B \subseteq \Gamma$ such that there are partitions $A_1, \ldots, A_n$ of A, $B_1, \ldots, B_n$ of B, and $\gamma_1, \ldots, \gamma_n \in \Gamma$ such that

$$\Gamma = \bigcup_{i=1}^{n} \gamma_i A_i = \bigcup_{j=1}^{n} \gamma_j B_j.$$

Proposition: F_2 is paradoxical.

Proof: Suppose a, b are the generators of F_2. Thus, every element of F_2 begins with $\{a, b, a^{-1}, b^{-1}\}$. Define $U(x)$ to be the set of reduced words that begins with x. Clearly, $F_2 \cong U(a) \cup U(a^{-1}) \cong U(b) \cup U(b^{-1})$. Thus, we have demonstrated that F_2 is paradoxical.

Definitions:

1. A countable group Γ is called an amenable group if it admits a left invariant finitely additive probability measure.

2. A countable group Γ satisfies the Reiter condition if $\forall \epsilon > 0$, $\forall \gamma_1, \ldots, \gamma_n \in \Gamma$, there exists $f \in \ell^1(\Gamma)$ such that $f \geq 0$, $\|f\|_1 = 1$, and $\forall 1 \leq i \leq n$, $\|f - \gamma_i f\|_1 < \epsilon$.

3. Γ (the same group as above) satisfies the Følner condition if $\forall \epsilon > 0$, $\forall \gamma_1, \ldots, \gamma_n \in \Gamma$, there exists $F \subseteq \Gamma$, with F finite, such that $\forall 1 \leq i \leq n$,

$$\frac{|\gamma_i F \Delta F|}{|F|} < \epsilon.$$

A Følner sequence is a sequence $\{F_n\}$ of nonempty subsets of Γ such that $\lim_{n\to\infty} \frac{|\gamma F_n \Delta F_n|}{|F_n|} = 0$ for all $\gamma \in \Gamma$. If a group Γ satisfies the Følner condition, then it admits a Følner sequence.

4. A mean on Γ is a linear functional (a linear map from a vector space to its field of scalars) $m : \ell^\infty(\Gamma) \to \mathbb{C}$ (the complex numbers) such that:
   (a) If $f \geq 0$, then $m(f) \geq 0$.
   (b) $m(1) = 1$, where 1 denotes the constant function with value 1.

   A mean m is left-invariant if for all $\gamma \in \Gamma$, $m(f \circ \gamma^{-1}) = m(f)$.

Proposition: Γ is an amenable group if and only if Γ admits a left invariant mean.

Theorem: Suppose Γ is a countable group, then the following are equivalent (TFAE):

1. Γ is an amenable group.
2. Γ satisfies the Reiter condition.
3. Γ satisfies the Følner condition.

Proof of (3) (⇒) (1): Here we will prove that if a group Γ satisfies the Følner condition, then it will be an amenable group. Suppose we have an increasing sequence of subsets of Γ and denote this sequence by $\{A_i\}_{i\in\mathbb{N}}$. Define $\{\epsilon_i\}_{i\in\mathbb{N}}$ to be a sequence of real numbers converging to 0. By the assumption of the Følner condition, we can find a sequence $\{F_n\}_{n\in\mathbb{N}}$ such that

$$\frac{|\gamma F_n \Delta F_n|}{|F_n|} < \epsilon \quad \text{for all } \gamma \in \Gamma \text{ and for any } \epsilon > 0.$$

Thus,

$$|\gamma F_n \Delta F_n| < |F_n| \cdot \epsilon \quad \text{for all } \gamma \in \Gamma.$$

Define $\text{Prob}(\Gamma)$ to be the space of all possible probability measures on Γ, or equivalently,

$$\text{Prob}(\Gamma) = \{\mu \in \ell^1(\Gamma) : \mu \geq 0 \text{ and } \sum_{\gamma\in\Gamma} \mu(\gamma) = 1\}.$$

Then, let $\mu_i = \frac{1}{|F_n|} 1_{F_n} \in \text{Prob}(\Gamma)$. Observe then that

$$\|\gamma\mu_i - \mu_i\|_1 = \frac{1}{|F_i|} |\gamma F_i \Delta F_i|.$$

Let $\mu \in \ell^\infty(\Gamma)$ and define $\mu :=$ the limit point in the weak-topology of the sequence $\{\mu_i\}$. Then μ is an invariant mean and thus Γ is an amenable group.

2.4 Tempered Folner Sequence:

In Elon Lindenstrauss's 2001 *Pointwise Theorem for Amenable Groups*, the author defined a **tempered Folner sequence** as a sequence of sets $\{F_n\}_{n\in\mathbb{N}}$ if there exists a constant $c > 0$ such that for all $n \in \mathbb{N}$, $|\bigcup_{k=1}^{n-1} (F_k)^{-1} F_n| \leq c|F_n|$.

Here, $F_k^{-1} F_n$ represents the set of all elements g of G (where G is a compact group such as the group of integers $\mathbb{Z}$) that can be written as $g = f_k^{-1} f_n$ for some $f_k \in F_k$ and $f_n \in F_n$.

2.5 Radon Nikodym Cocycles

In the paper by Lewis Bowen and Amos Nevo, Radon-Nikodym cocycles are frequently used and we shall present our investigation into the topic.

Let $\Gamma \curvearrowright (X, \mu)$ be a probability measure preserving transformation. A **cocycle** of this action is a measurable function $\omega : \Gamma \times X \to K$ defined by $\omega(st, x) = \omega(s, tx)\omega(t, x)$ for all $s, t \in \Gamma$ and a.e. $x \in X$.

Let E be a countable Borel equivalence relation on a standard probability space (X, μ). E is **measure class preserving** if either of the following equivalence conditions holds:

1. The E-saturation of null sets is null, i.e., if $Z \subset X$ is null, then $[Z]_E = \bigcup_{x\in Z}[x]_E = \{x \in X;\ \text{there exists } z \in Z \text{ such that } zEx\}$ is null.

2. For any/some Borel actions $\Gamma \curvearrowright X$ of a countable group Γ such that $E_\Gamma = E$, the action of Γ is measure class preserving, i.e., each $\gamma \in \Gamma$ maps null sets to null sets. Or equivalently, $\gamma_*\mu \sim \mu$.

3. There exists a unique (up to a null set) Borel function $w : E \to (\mathbb{R}^+, \cdot)$, defined by $w : (x, y) \mapsto w_x(y) \approx \frac{\text{weight}(y)}{\text{weight}(x)}$, called the Radon-Nikodym cocycle, satisfying:

   (a) It is firstly a cocycle.

   (b) For any Borel bijection $\gamma : A \to B$, where A, B are subsets of X such that $\text{graph}(\gamma) \subset E$, we have $\mu(B) = \mu(\gamma_* A) = \int_A w_x(\gamma_x)\, d\mu(x)$.

   (c) "Tilted" mass transport holds: for each Borel function $F : E \to [0, \infty]$,

$$\int \sum_{y\in[x]_E} F(x, y)\, d\mu(x) = \int \sum_{y\in[x]_E} F(y, x) \cdot w_x(y)\, d\mu(x).$$

Consequence: If T is a probability measure preserving map if and only if $w_x(T^{-1}(x)) = 1$ for a.e. $x \in X$.

2.6 Ratio Set, Type and Stable Type of Nonsingular Actions

Let Γ be a countable group and (X, μ) be a standard probability space. A **nonsingular action** of Γ acting on (X, μ) means $\mu(E) = 0$ implies that $\mu(\gamma E) = 0$ for all $\gamma \in \Gamma$ and E a measurable subset of X.

Definition: The ratio set of an action (denoted by $\text{RS}(\Gamma \curvearrowright (X, \mu)) \subseteq [0, \infty]$) is the set of $r \in [0, \infty]$ if and only if for every $A \subseteq X$ such that $\mu(A) > 0$ and $\forall \epsilon > 0$, $\exists A' \subseteq A$ with $\mu(A') > 0$, and $\gamma \in \Gamma \setminus \{e\}$ such that:

1. $\gamma \cdot A' \subseteq A$, and

2. $\left| \frac{d(\mu\circ\gamma)(b)}{d\mu} - r \right| < \epsilon \quad \forall b \in A'$.

Equivalent Definition of the Ratio Set of a Group Action: Suppose Γ is a countable group, and let (X, μ) be a standard probability space. Γ acts on X by p.m.p. transformations (i.e., for each $\gamma \in \Gamma$, $T_\gamma : X \to X$ satisfying the condition that $\mu(T_\gamma^{-1}(A)) = \mu(A)$ $\forall$ measurable sets $A \subseteq X$). The ratio set of the action of Γ on X is the set of all $r > 0$ such that $\exists A \subseteq X$ where $\mu(A) > 0$ and $\gamma_n \in \Gamma$ such that

$$\lim_{n\to\infty} \frac{\mu(T_{\gamma_n}(A) \cap A)}{\mu(A)} = r$$

exists.

We define the **stable ratio set** of the group action Γ acting on (X, μ) by the intersection over all ratio sets of Γ acting on $(X \times Y, \mu \times \nu)$.

Ratio Set / Stable Ratio Set	Type
$\{1\}$	II
$\{0, 1, \infty\}$	III_0
$\{0, \infty, \lambda^n : n \in \mathbb{Z}\}$	III_λ $(0 < \lambda < 1)$
$[0, \infty]$	III_1

Table 1: The Type Classification of The Ratio Sets and Stable Ratio Sets

Note: If for every ergodic action of Γ on (X, μ), the product action $\Gamma \curvearrowright (X \times Y, \mu \times \nu)$ is ergodic, then the action $\Gamma \curvearrowright (X, \mu)$ is **weakly mixing**.

A p.m.p. transformation T is called as **mixing** or **strongly mixing**, if for all $E, F \subseteq X$,

$$\lim_{n\to\infty} \mu(E \cap T^{-n}(F)) = \mu(E)\mu(F).$$

Proposition: The mixing condition implies ergodicity.

Proof: Let A be a T-invariant set $(T^{-1}(A) = A)$. Then, since by assumption T possesses the mixing property, $\mu(A) = \lim_{n\to\infty} \mu(A \cap T^{-n}(A)) = \mu(A)\mu(T^{-n}(A)) = (\mu(A))^2$. The equation $\mu(A) = (\mu(A))^2$ holds only if $\mu(A) = 0$ or 1.

2.7 Boundry Action of a Free Group

Through the research project, the topics of free groups, their characteristics, and the boundary action of the free group are discussed. The paper by Lewis Bowen and Amos Nevo utilizes free group and boundary action in their proof of the ergodic theorem. Here are basic some definitions and results on the boundary actions of the free group.

A **free group** $F(S)$ is a group generated by the set S. $F(S)$ is called a free group because the only relations are those that are absolutely necessary for the group axioms. Let $F(S) = \langle a_1, \ldots, a_r \rangle$ and let $S = \{a_i^{\pm 1} : 1 \leq i \leq r\}$ be a set of free generators. Assume the words are all reduced (i.e., if $g \in F(S)$, then $g = s_1 \ldots s_n$ where $s_i \neq s_{i+1}^{-1}$). The boundary of F, denoted ∂F, is defined as the set of sequences $\zeta = (s_1, s_2, s_3, \ldots) \in S^{\mathbb{N}}$.

In the case of F_2, ∂F_2 is a closed subset of $\{a, b, a^{-1}, b^{-1}\}^{\mathbb{N}} = 4^{\mathbb{N}}$, thus it is a compact Polish set.

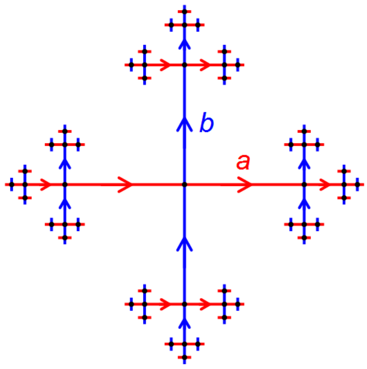

Figure 1: Cayley graph for the free group on 2 generators

We are now ready to define the action of F on ∂F. In the case of F_2, F_2 acts on ∂F_2 continuously by concatenation and cancellation. For $x \in \partial F_2$, the action is defined as follows:

$$ax = \begin{cases} ax & \text{if } x \neq (a^{-1})y, \\ y & \text{if } x = (a^{-1})y. \end{cases}$$

In the general case, the action of F on ∂F is defined as

$$(a_1 \dots a_n)\zeta = (a_1, \dots, a_{n-j}, s_{j+1}, s_{j+2}, \dots)$$

where $a_1, \dots, a_n \in S$ and j is the largest number such that $s_i^{-1} = a_{n+1-i}$ for all $i \leq j \leq n$.

We will demonstrate that a free group like F_2 cannot admit any **invariant Borel probability measure (i.e., $\mu(gA) = \mu(A)$ for all $g \in G$ and all measurable sets A).**

Proof: Suppose that μ is an invariant Borel probability measure on ∂F_2. Then $\partial F_2 = S_a \sqcup S_b \sqcup S_{a^{-1}} \sqcup S_{b^{-1}}$, where $\sqcup$ denotes disjoint union. One of the sets S_a, S_b, $S_{a^{-1}}$, or $S_{b^{-1}}$ must have positive measure. Without loss of generality, assume that S_a has positive measure. Then by the definition of μ being an invariant measure, $\mu(b^n S_a) = \mu(S_a)$ for all n. Therefore,

$$\mu\left(\bigsqcup_{n\in\mathbb{N}} b^n S_a\right) = \sum_{n\in\mathbb{N}} \mu(S_a) = \mu(S_a) \cdot \infty = \infty,$$

which yields a contradiction. Thus, we have shown that there does not exist an invariant Borel probability measure on ∂F_2.

While an invariant Borel probability measure does not exist, there are many quasi-invariant probability measures that do exist. Consider the **uniform measure** on ∂F_2 defined by $\mu([w]) = \frac{1}{4}\left(\frac{1}{3}\right)^{|w|-1}$, where $[w]$ denotes the sequence that begins with w and $|w|$ denotes the length of w (a finite word). In the paper by Lewis Bowen and Amos Nevo, they define a probability measure ν on ∂F as follows: for every finite sequence $a_1, \dots, a_n$ with $a_{i+1} \neq a_i^{-1}$ for all $1 \leq i < n$,

$$\nu(\{(s_1, s_2, \dots) \in \partial F \mid s_i = a_i \text{ for } 1 \leq i \leq n\}) = |S_n|^{-1} = \frac{1}{(2n)(2r-1)^{n-1}}.$$

2.8 Paper by Lewis Bowen and Amos Nevo in 2015

One of the focuses of the research project was understanding the famous theorem proven by Buffetov but using the methods introduced in Lewis Bowen and Amos Nevo's 2015 paper.

Theorem: Let F acts on $(X, \mathcal{B}, \mu)$ be a probability measure-preserving action. Then for all $f \in L\log L(X, \mu)$,

$$\frac{1}{|S_{2n}|}\sum_{\substack{g\in F^2\\ |g|=2n}} f\circ g \xrightarrow{a.e.} \mathbb{E}[f \mid \mathcal{F}^2]$$

almost everywhere on the σ-algebra of $\mathcal{F}^2$-invariant measurable subsets, as $n \to \infty$.

The proof of the theorem by Lewis Bowen and Amos Nevo uses various different topics such as *Radon-Nikodym cocycle*, *ratio sets (and stable types)*, *Maharam extension*, and *Theorem 3.1.*

We will first state an important theorem on the construction of ergodic averages obtained from the paper. In the paper, this theorem is denoted as theorem 3.1.

Suppose (B, ν, R) is a measured equivalence relation.[1] Let $\Omega = \{\omega_i\}_{i\in I}$ be a measurable family of leafwise probability measures $\omega_i : R \to [0, 1]$, and $\alpha : R \to \Gamma$ be a measurable cocycle with Γ a countable group. Suppose there exists a nonsingular [2] compact group action $K \curvearrowright (B, \nu)$ with uniformly bounded Radon-Nikodym derivatives [3] and $\psi \in L^q(B, \nu)$ with $q \in (1, \infty)$ and ψ is a probability density [4]

For $i \in I$, the probability measure ζ_i on the group Γ is defined as

$$\zeta_i(\gamma) = \int_B\int_K \sum_{c:\alpha(c,kb)=\gamma} \omega_i(c, kb)\psi(b)\, dk\, d\nu(b).$$

[1]: R is an equivalence relation on B such that R is a measurable subset of $B\times B$ under the product σ-algebra. It is often assumed that R is a countable equivalence relation, which means its equivalence classes are countable, and that the measure ν is compatible with the equivalence relation.

[2]: Equivalently, for any measurable set $W \subset B$, $\nu(W) = 0$ if and only if $\nu(kW) = 0$ for all $k \in K$.

[3]: Uniformly bounded RN-derivative means if $\{\nu_\gamma\}_{\gamma\in\Gamma}$ is a family of measures on the measure space B, then there exists a constant $C > 0$ such that $\frac{d\nu_\gamma(b)}{d\nu} \leq C$ for all $b \in B$ and $\gamma \in \Gamma$.

[4]: Intuitively, one may think of this as weights.

Set $p = \frac{q}{q-1}$, then we have the following results:

1. If Ω satisfies the strong L^p maximal inequality, then $\{\zeta_i\}_{i\in I}$ also satisfies the strong L^p maximal inequality.

Note: We say that a family of probability measures $\{m_i\}_{i\in I}$ defined on a countable group (such as Γ) satisfies the strong L^p maximal inequality if there exists a constant $D_p > 0$ depending only on $\{m_i\}_{i\in I}$ such that

$$\left\| \sup_{i\in I} \sum_{\gamma\in\Gamma} m_i(\gamma)|f|\circ\gamma^{-1} \right\|_p \le D_p\|f\|_p \quad \text{for all } f\in L^p(X,\mu).$$

2. If Ω is a pointwise ergodic family in L^p, then $\{\zeta_i\}_{i\in I}$ is a pointwise convergent family in L^p.

   Note: $\{m_i\}_{i\in I}$, defined on some group Γ, being a pointwise ergodic family in L^p means that for every p.m.p action $\Gamma \curvearrowright (X,\mu)$ and for all $f\in L^p(X,\mu)$, the functions $\sum_{\gamma\in\Gamma} m_i(\gamma) f\circ\gamma^{-1}$ converge pointwise a.e. as $i\to\infty$, and the a.e. pointwise limit is equal to the conditional expectation of f on the σ-algebra of Γ-invariant Borel sets.

3. If α is weakly mixing relative to the K-action, then $\{\zeta_i\}_{i\in I}$ is a pointwise ergodic family in L^p.

4. If Ω satisfies the $(1,1)$-type maximal inequality and $\psi\in L^\infty(B,\nu)$, then $\{\zeta_i\}_{i\in I}$ satisfies the $L\log(L)$ maximal inequality.

   Note:

   (a) Let $\{m_i\}_{i\in I}$ be a family of probability measures defined on some group Γ. $\{m_i\}_{i\in I}$ satisfy the weak $(1,1)$-type maximal inequality if there exists $C_{(1,1)} > 0$ depending only on $\{m_i\}_{i\in I}$ such that

   $$\mu\left(\left\{x\in X : \sup_{i\in I}\sum_{\gamma\in\Gamma} m_i(\gamma) f\circ\gamma^{-1}(x) \ge t\right\}\right) \le \frac{C_{(1,1)}\|f\|_1}{t}$$

   for every $f\in L^1(X,\mu)$ and $t>0$.

   (b) A family of probability measures $\{m_i\}_{i\in I}$ satisfies the $L\log(L)$ maximal inequality if there exists a constant $C_1 > 0$ depending only on $\{m_i\}_{i\in I}$ such that

   $$\left\| \sup_{i\in I} \sum_{\gamma\in\Gamma} m_i(\gamma) f\circ\gamma^{-1} \right\|_1 \le C_1\|f\|_{L\log(L)}$$

   for all $f\in L\log(L)(X,\mu)$. Here,

   $$\|f\|_{L\log(L)} = \int_X |f|\log^+(|f|)\,d\mu.$$

5. If Ω is a pointwise ergodic family in L^1, then $\{\zeta_i\}_{i\in I}$ is a pointwise convergent family in $L\log(L)$.

6. If α is weakly mixing relative to the K-action, then $\{\zeta_i\}_{i\in I}$ is a pointwise ergodic family in $L\log(L)$.

Clearly, this theorem is of great importance and we will omit the full proof for now. However, we shall see how this theorem is used to their proof. Before we discuss how Lewis Bowen and Amos Nevo proved the ergodic theorem for $f \in L \log L$, there are a few more definitions and theorems that deserve to be presented here.

Previously, we defined weakly mixing for a transformation T. Now we want to define **weakly mixing of a cocycle**. A measurable cocycle $\alpha : R \to \Gamma$, where R is a discrete measurable equivalence relation ($R \subseteq B \times B$ where (B, ν) is a standard probability space) and Γ is a countable group, is said to be weakly mixing if for every ergodic p.m.p. action $T \curvearrowright^{\beta} (X, \mu)$, the induced equivalence relation $R_{\beta\alpha}$ on $B \times X$ is ergodic.

Generalization of weakly-mixing property of cocyle: Let K be a compact group with a nonsingular measurable action $K \curvearrowright (B, \nu)$. The measurable cocycle $\alpha : R \to \Gamma$ is said to be weakly mixing relative to the action $K \curvearrowright (B, \nu)$ if for every p.m.p. action $\Gamma \curvearrowright^{\beta} (X, \mu)$ and every $f \in L^1(X) \subseteq L^1(B \times X)$,

$$\int \mathbb{E}[f \mid R_{\beta\alpha}](kb, x)\, dk = \mathbb{E}[f \mid \Gamma](x) \quad \text{for a.e. } (b, x).$$

Here, dk represents the Haar probability measure on the compact group K. The expectation of f given $R_{\beta\alpha}$ is the conditional expectation of f, which is an element of $L^1(B \times X)$, on the σ-algebra $R_{\beta\alpha}$, corresponding to $R_{\beta\alpha}$-saturated measurable sets. The expectation of f given Γ is the conditional expectation of f on the σ-algebra of Γ-invariant sets.

Note: A set $A \subseteq B \times X$ is $R_{\beta\alpha}$-saturated if for all $(b_1, x_1) \in A$ and $(b_2, x_2) \in B \times X$, if (b_1, x_1) is $R_{\beta\alpha}$-equivalent to (b_2, x_2), then (b_2, x_2) belongs to A.

We shall define a few important equalities.

1.
$$\begin{aligned}
\sum_{\gamma \in \Gamma} \zeta_i(\gamma) f \circ \gamma^{-1}(x) &= \sum_{\gamma \in \Gamma} f \circ \gamma^{-1}(x) \int_B \int_K \sum_{c:\alpha(c,kb)=\gamma} \omega_i(c, kb)\psi(b)\, dk\, d\nu(b) \\
&= \int_B \int_K \sum_{c:\alpha(c,kb)\in R} \omega_i(c, kb) f(c, \alpha(kb)^{-1}x)\psi(b)\, dk\, d\nu(b) \\
&= \int_B \int_K \mathcal{A}[f \mid \omega_i](kb, x)\psi(b)\, dk\, d\nu(b).
\end{aligned}$$

2.
$$\Pi(m_i)(F)(x) = \int_B \int_K \mathcal{A}[F \mid \omega_i](kb, x)\psi(b)\, dk\, d\nu(b) \quad \text{for } F \in L^p(B \times X).$$

We now want to present a theorem following the construction of ergodic averages.

Theorem 3.2:

(i.) If Ω (defined as above) satisfies the strong L^p maximal inequality and $\frac{1}{p} + \frac{1}{q} = 1$, then there exists a positive constant Z_p such that

$$\left\| \sup_{i \in I} \Pi(m_i)(|F|) \right\|_p \leq Z_p \|F\|_p \quad \text{for all } F \in L^p(B \times X).$$

(ii.) If Ω satisfies the $L\log(L)$ maximal inequality and $\psi \in L^\infty(B,\nu)$, then there exists a positive constant Z_1 such that

$$\left\| \sup_{i\in I} \Pi(m_i)(|F|) \right\|_1 \leq Z_1 \|F\|_{L\log(L)} \quad \text{for all } F \in L\log(L)(B\times X).$$

Proof:
WLOG, we can assume $F \geq 0$. Consider $p > 1$ first. Since Ω satisfies the strong L^p maximal inequality, by definition,

$$\left\| \sup_{i\in I} \sum_{\gamma\in\Gamma} m_i(\gamma)|F| \circ \gamma^{-1} \right\|_p \leq C_p \|F\|_p,$$

where C_p is some positive constant. This holds for every $F \in L^p(B\times X)$. By Hölder's inequalities,

$$\begin{aligned}
\left\| \sup_{i\in I} \Pi(\zeta_i)(|F|) \right\|_p^p &= \int_X \left| \sup_{i\in I} \int_B \int_K \mathcal{A}[F \mid \omega_i](kb,x)\psi(b)\, dk\, d\nu(b) \right|^p d\mu(x) \\
&\leq \int_X \sup_{i\in I} \left(\int_B \int_K \mathcal{A}[F \mid \omega_i](kb,x)^p\, dk\, d\nu(b) \right) \\
&\quad \cdot \left(\int_B \int_K \psi(b)^q\, dk\, d\nu(b) \right)^{p/q} d\mu(x) \\
&= \|\psi\|_p^p \int_X \sup_{i\in I} \int_B \int_K \mathcal{A}[F \mid \omega_i](kb,x)^p\, dk\, d\nu(b) d\mu(x) \\
&\leq \|\psi\|_p^p \int_X \int_B \int_K \sup_{i\in I} \Pi(\zeta_i)(|F|)(kb,x)^p\, dk\, d\nu(b) d\mu(x) \\
&= \|\psi\|_p^q \int_X \int_B \int_K \sup_{i\in I} \Pi(\zeta_i)(|F|)(b,x)^p \frac{d\nu\circ k^{-1}}{d\nu}(b)\, dk\, d\nu(b) d\mu(x) \\
&\leq C(K)\|\psi\|_p^q \int_X \int_B \int_K \sup_{i\in I} \Pi(\zeta_i)(|F|)(b,x)^p\, d\nu(b) d\mu(x) \\
&= C(K)\|\psi\|_p^q \left\| \sup_{i\in I} \Pi(\zeta_i)(|F|) \right\|_p^p \\
&\leq C(K)\|\psi\|_p^q C_p \|F\|_p^p = Z_p \|F\|_p^p,
\end{aligned}$$

The proof of the $L\log(L)$ case is omitted. We will present one more theorem and another corollory.

Theorem: For a countable group Γ (such that $|\Gamma| = \infty$), there exists a weakly mixing cocycle $\alpha : R \to \Gamma$, where R is an amenable, discrete, ergodic p.m.p. equivalence relation, and $R \subseteq B \times B$.

Proof: To prove the theorem, we select a probability measure $\kappa \in \text{Prob}(\Gamma)$ such that the support of κ [1] generates the group Γ. Consider the product space $\Gamma^{\mathbb{Z}}$ with the product topology, and let it be equipped with the measure $\kappa^{\mathbb{Z}}$. Define a shift action T on $\Gamma^{\mathbb{Z}}$ by $T(x)(n) = x(n+1)$ where $x \in \Gamma^{\mathbb{Z}}$ and $n \in \mathbb{Z}$. Let R be the equivalence relation on $\Gamma^{\mathbb{Z}}$ by considering pairs $(x, T^n(x))$. Thus, we can formally define

$$R = \{(x, T^n(x)) : x \in \Gamma^{\mathbb{Z}}, n \in \mathbb{Z}\},$$

and the equivalence relation R is determined by the orbits of T, which is isomorphic to $\mathbb{Z}$. Thus, we may reasonably conclude that R is the orbit equivalence relation of the shift transformation. Dye's theorem [2] guarantees that R is a hyperfinite II_1 [3] equivalence relation. Since hyperfinite equivalence relations are amenable, R is amenable.

Now, we define the cocycle $\alpha : R \to \Gamma$ as the following:

$$\alpha(x, T^n(x)) = \begin{cases} x(1)x(2)\dots x(n) & \text{if } n > 0, \\ e & \text{if } n = 0, \\ x(0)^{-1}x(-1)^{-1}\dots x(n+1)^{-1} & \text{if } n < 0, \end{cases}$$

where e is the identity in Γ. This α that we just defined is a well-defined and measurable cocycle on an invariant co-null measurable set.

To finalize the proof, we need to show that α is a weakly mixing cocycle. I am not entirely certain of the methods used by Lewis Bowen and Amos Nevo to show that this cocycle is weakly mixing.

[1:]

$$\text{supp}(\kappa) = \{\gamma \in \Gamma \mid \kappa(U) > 0 \text{ for every open } U \text{ containing } \gamma\}$$

[2:]

1. A CBER (Countable Borel Equivalence Relation) R on X is called hyperfinite if there is an increasing exhaustive sequence $\{R_n\}_{n\in\mathbb{N}}$ of finite Borel subequivalence relations of R.

2. Dye's theorem states that if R is the union of an increasing sequence of hyperfinite Borel equivalence relations and $\mu \in \mathcal{P}(X)$, then R is hyperfinite μ-a.e.

[3:] I believe that this is the von Neumann classification of algebras Type II factors. However, I am not entirely sure about its definition and implication.

Corollory: Suppose κ is a probability measure on Γ whose support generates Γ. Let $\rho_n = \left(\sum_{k=1}^{n} \kappa^{*k}\right)/n$. Then $\{\rho_n\}_{n\in\mathbb{N}}$ is a pointwise ergodic sequence in $L\log(L)$.

Note: $\kappa^{(*k)}$ represents the k-fold convolution power of κ.

Now we are ready to discuss Lewis Bowen and Amos Nevo's proof of Theorem 6.2, which has been proven separately in a paper by Bufetov. However, the proof in Bufetov uses different techniques.
The authors define a measure on the Borel σ-algebra of ∂F by

$$\nu(\{(s_1, s_2, \dots) \in \partial F \mid s_i = a_i \text{ for } 1 \leq i \leq n\}) = |S_n|^{-1} = \frac{1}{(2n)(2r-1)^{n-1}}.$$

They show that the Radon-Nikodym derivative satisfies

$$\frac{d(\nu \circ g)(\zeta)}{d\nu} = \frac{1}{(2r-1)^{n-2k}}.$$

The ratio set of the action of the free group F on $(\partial F, \nu)$ is categorized as III_λ with $\lambda = \frac{1}{2r-1}$. According to the Geometric Covering Argument and Ergodic Theorem for Free Groups, the stable type of $F \curvearrowright (\partial F, \nu)$ is III_{λ^2}.

Then the author uses the **Maharam extension** to obtain an equivalence relation on $\partial F \times \{0, 1\}$ and a cocycle. Let F act on $\partial F \times \mathbb{Z}$ defined by $(gb, t + R_\lambda(gb))$, which defines a Maharam extension. For simplicity, the authors let R be the orbit equivalence relation restricted to $\partial F \times \{0\}$, which we can naturally identify as ∂F. Thus, b is R-equivalent to b' if there exists $g \in F$ such that $gb = b'$ and $\frac{d(\nu \circ g)(b)}{d\nu} = 1$. This definition of R equivalence is the same as the tail equivalence relation on ∂F (where ζ and η are two elements of ∂F and they are equivalent if and only if there is a j such that $\zeta_n = \eta_n$ for all $n \geq j$).

The paper then proceeds to define the cocycle $\alpha : R \to F^2$ by $\alpha(gb, b) = g$ for $g \in F^2$ and $b \in \partial F$. Since F^2 acts on $(\partial F, \nu)$ with type and stable type III^2_λ, the defined cocycle is weakly mixing for F^2. We then define $\omega_n : R \to [0, 1]$ to be a leafwise probability measure given by $\omega_n(gb, b) = \frac{1}{2r-2}\left(\frac{1}{(2r-1)^{n-1}}\right)$ if $|g| = 2n$, and 0 if $|g| \neq 2n$. The measure $\omega_n(\cdot, b)$ is uniformly distributed over the set of all elements of the form gb with $|g| = 2n$ and $\frac{d(\nu \circ g)(b)}{d\nu} = 1$. Set $\zeta_n(\gamma) = \int_{\partial F} \sum_{c:\alpha(c,b)=\gamma} \omega_n(c, b)\, d\nu(b)$. It was shown that $\{\omega_n\}_{n\in\mathbb{N}}$ is pointwise ergodic in L^1 and satisfies the weak $(1, 1)$-type maximal inequality. Applying Theorem 3.1 gives us that $\{\zeta_n\}_{n\in\mathbb{N}}$ is pointwise ergodic in $L \log(L)$ for F_2-actions.

3 Conclusion

This research project has deepened my understanding of ergodic theory and greatly enriched my mathematical education. Reading the paper by Lewis Bowen and Amos Nevo offered a valuable opportunity not only to explore the intricacies of their proof of the Pointwise Ergodic Theorem but also to experience learning the rigorous process of mathematical research. The process of deciphering complex concepts, techniques, and methodologies presented in their work has provided me with a clearer insight into the challenges and rewards associated with research in mathematics.

References

[1] Bowen, L., Nevo, A. Amenable equivalence relations and the construction of ergodic averages for group actions. JAMA 126, 359–388 (2015). https://doi.org/10.1007/s11854-015-0021-5

[2] Lindenstrauss, E. Pointwise theorems for amenable groups. Invent. math. 146, 259–295 (2001). https://doi.org/10.1007/s002220100162

[3] Anush, T. (2018). A descriptive set theorist's proof of the pointwise ergodic theorem. https://doi.org/10.48550/arXiv.1805.07365

[4] Kerr D., Li H. (2017) Ergodic Theory: Independence and Dichotomies. https://doi.org/10.1007/978-3-319-49847-8

[5] Bowen, L., Hartman, Y., Tamuz, O. (2014) Property (T) and the Furstenberg Entropy of Nonsingular Actions, https://doi.org/10.48550/arXiv.1406.4488

[6] Juschenko, K., Lecture 2: First definitions of amenability, elementary operations that preserve it, https://web.ma.utexas.edu/users/juschenko/files/Lecture2.pdf