NOTES ON ERGODIC THEORY

ZHAOSHEN ZHAI

Abstract. Notes on the Summer 2025 Reading Group on Ergodic Theory, following [Tse22], organized by Frédéric Kai, Ludovic Rivet, and Zhaoshen Zhai (website).

Contents

1. Lecture 1 (Samy Lahlou): Crash course on Measure Theory, Part I

Given a set X, our goal is to define a map $\mu : \mathcal{P}(X) \to [0, \infty]$ that assigns to each subset $A \subseteq X$ a *measure* $\mu(A) \in [0, \infty]$ that 'behaves like the volume of A'. This turns out to be impossible in full generality (and we shall see using ergodic-theoretic methods that this impossibility is for good reason; see Exercise 4.20), so we instead restrict to special subsets of X with a nice algebraic (think: 'constructible') structure.

Further reading. [Tse23, Lectures 1 to 5] and [Fol99, Chapter 1].

Definition 1.1. Let X be a set. A *σ-algebra on X* is a collection $\mathcal{B} \subseteq \mathcal{P}(X)$ of subsets of X containing $\varnothing$ and is closed under complements and countable unions. More precisely:

1. (Non-trivial). $\varnothing \in \mathcal{B}$.
2. (Closure under complements). For any $A \in \mathcal{B}$, we have $X \setminus A \in \mathcal{B}$.
3. (Closure under countable unions). For any countable family $\{A_n \in \mathcal{B} : n \in \mathbb{N}\}$, we have $\bigcup_n A_n \in \mathcal{B}$.

Definition 1.2. If $\mathcal{B}$ is a σ-algebra on a set X, the pair $(X, \mathcal{B})$ is said to be a *measurable space*.

A useful way to construct a σ-algebra is to start with an arbitrary family $\mathcal{C} \subseteq \mathcal{P}(X)$ and close[1] it under the above three conditions. Abstractly:

Definition 1.3. The σ-algebra *generated* by $\mathcal{C} \subseteq \mathcal{P}(X)$ is $\langle \mathcal{C} \rangle_\sigma \coloneqq \bigcap\{\mathcal{B} : \mathcal{B} \supseteq \mathcal{C} \text{ is a } \sigma\text{-algebra on } X\}$.

Note that $\langle \mathcal{C} \rangle_\sigma$ is indeed a σ-algebra on X since the intersection of σ-algebras is again a σ-algebra.

Date: June 6, 2025.

[1] This 'closure' operation can be made precise as follows. Starting with $\mathcal{C}_0 \coloneqq \mathcal{C}$, throw in all the subsets of X that is necessary to satisfy Definition 1.1 relativized to $\mathcal{C}_0$ to obtain $\mathcal{C}_1$ (that is, let $\mathcal{C}_1$ contain $\varnothing$ and such that if $A \in \mathcal{C}_0$, then $X \setminus A \in \mathcal{C}_1$, and similarly for condition 3). Then, let $\mathcal{C}_2$ be defined by throwing in all the countable unions and complements of sets in $\mathcal{C}_1$. Doing so infinitely-many times and taking the union $\bigcup_\alpha \mathcal{C}_\alpha$ will give us $\langle \mathcal{C} \rangle_\sigma$, but beware that this process must proceed into the transfinite up to $\alpha < \omega_1$, where ω_1 is the first uncountable ordinal; ask your local set theorist why.

Definition 1.4. Let X be a topological space. The *Borel σ-algebra* on X is $\mathcal{B}(X) \coloneqq \langle \mathcal{T} \rangle_\sigma$, where $\mathcal{T}$ is the topology on X. The elements of $\mathcal{B}(X)$ are called the *Borel sets* of X.

Intuitively, for any topological space X, one would like to 'measure' the Borel sets. This is justified since if one wants a measure compatible with the topology, then one must be able to measure the open sets, and hence also closed sets, countable unions of closed sets (called *F_σ sets*), countable intersections of open sets (called *G_δ sets*), countable intersections of F_σ sets, countable unions of G_δ sets, and so on[2].

Definition 1.5. A *measure* on a measurable space $(X, \mathcal{B})$ is a map $\mu : \mathcal{B} \to [0, \infty]$ such that $\mu(\varnothing) = 0$ and $\mu(\bigsqcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$ for any pairwise disjoint family $\{A_n \in \mathcal{B} : n \in \mathbb{N}\}$.

The triple $(X, \mathcal{B}, \mu)$ is then called a *measure space.* A *Borel measure* is a measure defined on some Borel σ-algebra.

Example 1.6 (Lebesgue)**.** Equip $\mathbb{R}$ with its usual topology. There is[3] a unique measure $\lambda : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ on $\mathbb{R}$, called the *Lebesgue measure*, such that $\lambda([a, b]) = b - a$ for each $a \leq b$.

Example 1.7 (Bernoulli)**.** Equip $2 = \{0, 1\}$ with the discrete topology and consider the product topology on $2^{\mathbb{N}}$. For each $p \in [0, 1]$, is a unique measure $\mu_p : \mathcal{B}(2^{\mathbb{N}}) \to [0, \infty]$ on $2^{\mathbb{N}}$, called the *Bernoulli (p) measure*, such that for each word $w \in 2^{<\mathbb{N}}$, we have $\mu_p([w]) = p^{n_1}(1-p)^{n_0}$ where n_i is the number of $i \in \{0, 1\}$ in w and $[w]$ is the set of all sequences in $2^{\mathbb{N}}$ containing w as a prefix.

If $p = 0$ (similarly if $p = 1$), then $\mu_p(\xi) \in \{0, 1\}$, and we have $\mu_p(\xi) = 1$ iff $(p, p, p, \ldots) \in \xi$. Thus, all of the measure is concentrated at $(p, p, p, \ldots)$. Measures in which this occurs are called *Dirac measures.*

Example 1.8 (Dirac)**.** Let X be a set and fix $x \in X$. The *Dirac measure concentrated* at x is the measure $\delta_x : \mathcal{P}(X) \to \{0, 1\}$ defined by $\delta_x(A) \coloneqq 1$ iff $x \in A$, and $\delta_x(A) \coloneqq 0$ iff $x \notin A$.

Definition 1.9. A measure μ on $(X, \mathcal{B})$ is said to be *finite* if $\mu(X) < \infty$, a *probability measure* if $\mu(X) = 1$, and *σ-finite* if there is a partition $X = \bigsqcup_{n \in \mathbb{N}} X_n$ such that $X_n \in \mathcal{B}$ and $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$.

Unless otherwise stated, all measures are assumed to be σ-finite. In fact, we will usually only deal with probability measures, since we can also normalize a finite measure to a probability measure by $\mu \mapsto \mu/\mu(X)$.

Lastly, even though μ is only defined on the σ-algebra $\mathcal{B}$, we can slightly extend μ to a larger σ-algebra.

Definition 1.10. Let $(X, \mathcal{B}, \mu)$ be a measure space. A subset $Z \subseteq X$ is said to be *μ-null* if there exists some $Z' \in \mathcal{B}$ such that $Z \subseteq Z'$ and $\mu(Z') = 0$. We write Null_μ for the set of all μ-null subsets of X. A subset $A \subseteq X$ is said to be *μ-conull* if $X \setminus A$ is μ-null.

Definition 1.11. Let $(X, \mathcal{B}, \mu)$ be a measure space. A subset $A \subseteq X$ is *μ-measurable*[4] if there exists some $B \in \mathcal{B}$ and some μ-null set Z such that $A = B \cup Z$. We write Meas_μ for the set of all μ-measurable sets.

It is an exercise that $\mathrm{Meas}_\mu = \langle \mathcal{B} \cup \mathrm{Null}_\mu \rangle_\sigma$. Moreover, μ admits a unique extension to a map $\overline{\mu} : \mathrm{Meas}_\mu \to [0, \infty]$, called the *completion* of μ, and this measure satisfies $\mathrm{Meas}_{\overline{\mu}} = \mathrm{Meas}_\mu$. HINT: $\overline{\mu}(B \cup Z) \coloneqq \mu(B)$.

Definition 1.12. A measure μ is *complete* if $\overline{\mu} = \mu$.

For convenience, we will always assume that measures are complete. Neither measures λ nor μ_p in Examples 1.6 and 1.7 are complete, so we tacitly extend them.

We end with some easy exercises on measures; please read/prove them, as they will be used freely in the future; they are roughly ranked by difficulty. Throughout, let $(X, \mathcal{B}, \mu)$ be a measure space and let $A_n \in \mathcal{B}$.

Exercise 1.13 (Monotonicity)**.** If $A_1 \subseteq A_2$, then $\mu(A_1) \leq \mu(A_2)$.

---

[2] This goes up the *Borel hierarchy*, studied in *Descriptive Set Theory*; see [Kec95].

[3] We will not prove this fact, but it is an application of Carathéodory's Extension Theorem; see [Tse23, Lecture 4].

[4] Very confusing terminology. One might think that elements of $\mathcal{B}$ are the 'measurable' ones, but this removes μ from the picture. In general, there are much more μ-measurable sets that there are sets in $\mathcal{B}$. Indeed, there are $2^{\aleph_0}$-many Borel sets on $\mathbb{R}$, but there are $2^{2^{\aleph_0}}$-many λ-measurable sets!

Deduce that if μ is finite, then μ is a bounded function. (Are σ-finite measures bounded?)

Exercise 1.14 (Inclusion-exclusion)**.** For any $A_1, A_2 \in \mathcal{B}$, we have $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$.

Exercise 1.15 (Continuity $\nearrow$)**.** If $(A_n)_{n\in\mathbb{N}}$ is increasing, then $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_n \mu(A_n)$.

Exercise 1.16 (Continuity $\searrow$)**.** If $(A_n)_{n\in\mathbb{N}}$ is decreasing and $\mu(A_1) < \infty$, then $\mu(\bigcap_{n\in\mathbb{N}} A_n) = \lim_n \mu(A_n)$.

Exercise 1.17. Show that $\lambda(\mathbb{Q}) = 0$. Hint: What is the Lebesgue measure of singletons?

Let P be a property of some points in X. We say that P *holds μ-almost everywhere* (or *μ-almost surely*) if $\{x \in X : x \text{ satisfies } P\}$ is μ-conull.

Exercise 1.18 (Borel-Cantelli Lemmas)**.** Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of μ-measurable sets.

1. If $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$, then μ-almost every $x \in X$ lives in at-most finitely-many A_n.
2. (Measure Compactness). If $\mu(X) < \infty$ and there exists $\varepsilon > 0$ such that $\mu(A_n) \geq \varepsilon$ for all $n \in \mathbb{N}$, then at least an ε-measure set of $x \in X$ lives in infinitely-many A_n's.

For measurable spaces $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$, define $\mathcal{B}_1 \otimes \mathcal{B}_2 \coloneqq \langle B_1 \times B_2 : B_i \in \mathcal{B}_i \rangle_\sigma$.

Exercise 1.19. Show that if X_i are second-countable topological spaces, then $\mathcal{B}(X_1 \times X_2) = \mathcal{B}(X_1) \otimes \mathcal{B}(X_2)$.

Exercise 1.20. Let X be a topological space. A *Cantor set* is a subset $C \subseteq X$ homeomorphic to $2^{\mathbb{N}}$.

1. Show that the 'middle-thirds Cantor set' $C \subseteq [0,1]$ is a Cantor set as in the above definition. Moreover, show that $\lambda(C) = 0$. Hint: Recall the construction $C = \bigcap_{n\in\mathbb{N}} C_n$ and use continuity.
2. Define a Cantor set $C \subseteq [0,1]$ with positive Lebesgue measure. Hint: fatten the standard construction.

A measurable set $A \subseteq X$ is said to be an *atom* if there is no subset $B \subseteq A$ with $0 < \mu(B) < \mu(A)$. For example, singletons $\{x\}$ are atoms under the Dirac measure δ_x. More generally:

Exercise 1.21 (Atomic Decomposition)**.** If $(X, \mathcal{B}, \mu)$ is a σ-finite measure space, $\mathcal{B}$ is *countably generated* (i.e., $\mathcal{B} = \langle \mathcal{B}_0 \rangle$ for some countable $\mathcal{B}_0 \subseteq \mathcal{P}(X)$), and *separates points* (i.e., if $x \neq y$, then there exists $B \in \mathcal{B}$ such that $x \in B \not\ni y$.), then every atom $A \in \mathcal{B}$ is a singleton. Moreover, if $\{x_\alpha\}$ are all the atoms (how many can there be?), then $\mu = \mu_0 + \sum_\alpha a_\alpha \delta_{x_\alpha}$ for some atomless measure μ_0 and some $a_\alpha \geq 0$.

2. Lecture 2 (Samy Lahlou): Crash course on Measure Theory, Part II

Let $(X, \mathcal{B}, \mu)$ be a measure space. Our goal is to define the *Lebesgue integral* $\int f \, \mathrm{d}\mu$ for a function $f : X \to \mathbb{R}$. Again, this is not possible in full generality, so we restrict ourselves to the so-called *measurable functions*.

Further reading. [Tse23, Lectures 9 to 13, 17 to 21] and [Fol99, Chapters 2 and 3].

Definition 2.1. A *simple function* is an $\mathbb{R}$-linear combination of characteristic functions on μ-measurable sets, i.e., if $(E_i)_{i\leq n}$ is a collection of pairwise-disjoint μ-measurable sets and $(a_i)_{i\leq n}$ are distinct reals, then $\varphi \coloneqq \sum_{i\leq n} a_i \chi_{E_i}$ is a said to be a *simple function.* Define its *(Lebesgue) integral* as $\int \varphi \, \mathrm{d}\mu \coloneqq \sum_{i\leq n} a_i \mu(E_i)$.

For a (bounded) positive function $f : X \to \mathbb{R}_{\geq 0}$, we might define $\int f \, \mathrm{d}\mu$ by approximating f by simple functions from below, say by an increasing sequence (φ_n) of simple functions such that $f = \lim_n \varphi_n$ uniformly. However, not all functions f admit such an approximation.

To see this, let us attempt to construct such a sequence (φ_n). For each n, we will approximate the cutoff of f at 2^n, i.e., the function $\min(f, 2^n)$. We do so by partitioning the codomain $[0, 2^n]$ into intervals of length 2^{-n}, for a total of $k_n \coloneqq 2^n/2^{-n} = 2^{2n}$ intervals. Set $E_k \coloneqq f^{-1}([2^{-n}k, \infty))$ for each $k \in \{1, \ldots, k_n\}$, and let $\varphi_n \coloneqq \sum_{k\leq k_n} 2^{-n} \chi_{E_k}$. One easily checks that $f = \lim_n \varphi_n$ uniformly.

However, E_k is not guaranteed to be μ-measurable! To fix this, we simply define the issue away.

Definition 2.2. A function $f : X \to Y$ between measurable spaces $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces is said to be $(\mathcal{B}, \mathcal{C})$*-measurable* if $f^{-1}(C) \in \mathcal{B}$ for all $C \in \mathcal{C}$.

A function $f : X \to Y$ between topological spaces is said to be *Borel* if it is $(\mathcal{B}(X), \mathcal{B}(Y))$-measurable. A *Borel isomorphism* is a bijection $f : X \to Y$ such that both f and f^{-1} are Borel.

Exercise 2.3. Continuous maps are Borel. Hint: Define a σ-algebra containing open sets in the codomain.

So far we only dealt with measurable spaces. Let us now bring a measure into the picture.

Definition 2.4. Let (X, μ)[5] be a measure space and Y be a topological space. A function $f : X \to Y$ is said to be *μ-measurable* if it is $(\mathrm{Meas}_\mu, \mathcal{B}(Y))$-measurable.

Remark 2.5. Compositions of μ-measurable functions *need not* be μ-measurable.

The following exercise is one of the main reasons why μ-measurable functions are introduced, and ultimately also why the Lebesgue integral is superior compared to the Riemann integral.

Exercise 2.6. In separable metric spaces, pointwise limits of μ-measurable functions are μ-measurable, i.e., if (f_n) is a sequence of μ-measurable maps $f_n : X \to Y$ from a measure space (X, μ) to a separable space Y, and $f := \lim_n f_n$ (pointwise), then $f : X \to Y$ is μ-measurable.

Hint: Let $\mathcal{C} := \{B \in \mathcal{B}(Y) : f^{-1}(B) \in \mathrm{Meas}_\mu\}$. Show that $\mathcal{C}$ is a σ-algebra containing all open set in Y, so $\mathcal{C} = \mathcal{B}(Y)$, as desired. For each $U \subseteq Y$ open, use separability to write $U = \bigcup_{n \in \mathbb{N}} B_n$, where each B_n is a ball whose closure is contained in U, and show that $f^{-1}(U) \in \mathrm{Meas}_\mu$.

Exercise 2.7. If $f_1, f_2 : (X, \mu) \to \mathbb{R}$ are μ-measurable and $g : \mathbb{R}^2 \to \mathbb{R}$ is Borel, then $g(f_1, f_2) : X \to \mathbb{R}$ is also μ-measurable. In particular, $f_1 + f_2$ and $f_1 \cdot f_2$ are μ-measurable.

Exercise 2.8. If (f_n) is a sequence of μ-measurable functions $f_n : X \to \overline{\mathbb{R}}$, then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, and $\liminf_n f_n$ are also μ-measurable.

Notation 2.9. We write $L(X, \mu)$ for the set of all μ-measurable functions $f : (X, \mu) \to \overline{\mathbb{R}}$, and $L^+(X, \mu)$ for those which are non-negative.

We are finally ready to define the Lebesgue integral.

Definition 2.10. Let (X, μ) be a measure space. The *(Lebesgue) integral* of $f \in L^+(X, \mu)$ is

$$\int f \,\mathrm{d}\mu := \sup \left\{ \int \varphi \,\mathrm{d}\mu : 0 \le \varphi \le f \text{ simple function} \right\}.$$

In general, if $f \in L(X, \mu)$, we decompose $f = f^+ - f^-$ where $f^+ := \max\{f, 0\}$ and $f^- := \max\{-f, 0\}$. The *(Lebesgue) integral* of f is $\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \mathrm{d}\mu$, provided that one of the terms is finite.

If $\int f \,\mathrm{d}\mu < \infty$, we say that f is *μ-integrable*, in which case we write $f \in L^1(X, \mu)$. More generally,

Definition 2.11. Take $p \in [1, \infty]$ and let $L^p(X, \mu)$ be the set of all μ-measurable functions $f : X \to \overline{\mathbb{R}}$ such that $\|f\|_p < \infty$, where $\|f\|_p := (\int |f|^p \,\mathrm{d}\mu)^{1/p}$ if $p < \infty$ and $\|f\|_\infty := \text{ess-sup}\,|f| := \inf\{c \ge 0 : |f| \le c \ \mu\text{-a.e.}\}$.

Exercise 2.12. Let $f, g \in L^p(X, \mu)$. If $f \le g$, then $\|f\|_p \le \|g\|_p$.

Since a μ-measurable function $f : X \to \overline{\mathbb{R}}$ can be approximated from below by simple functions (φ_n), we should be able to calculate $\int f \,\mathrm{d}\mu$ as the limit of $\int \varphi_n \,\mathrm{d}\mu$. Indeed,

Theorem 2.13 (Monotone Convergence Theorem)**.** *If $(f_n) \in L^+(X, \mu)$ and $f_n \nearrow f$, then $\int f_n \,\mathrm{d}\mu \nearrow \int f \,\mathrm{d}\mu$.*

Corollary 2.14. *If $(f_n) \in L^+(X, \mu)$, then $\sum_n \int f_n \,\mathrm{d}\mu = \int \sum_n f_n \,\mathrm{d}\mu$.*

Exercise 2.15. For any $f, g \in L^1(X, \mu)$ and $a, b \in \mathbb{R}$, we have $\int (af + bg) \,\mathrm{d}\mu = a \int f \,\mathrm{d}\mu + b \int g \,\mathrm{d}\mu$.

Hint: Simple $\rightsquigarrow_{\mathrm{MCT}} L^+ \rightsquigarrow L^1$.

Exercise 2.16. Let $f, g \in L^1(X, \mu)$. If $f = g$ μ-a.e., then $\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$. Hint: Consider $\int (f - g) \,\mathrm{d}\mu$.

We list two more convergence theorems that will be useful later on.

Theorem 2.17 (Fatou's Lemma)**.** *If $(f_n) \in L^+(X, \mu)$, then $\int \liminf_n f_n \,\mathrm{d}\mu \le \liminf_n \int f_n \mathrm{d}\mu$.*

[5]Whenever the σ-algebra is not stated, we assume that μ is defined on Meas_μ. In particular, we assume that μ is complete.

Theorem 2.18 (Dominated Convergence Theorem). *Let $(f_n) \in L^1(X,\mu)$. If $f_n \to f$ μ-a.e. and $|f_n| \leq g$ for some $g \in L^1(X,\mu)$, then $\lim_n \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$.*

Let us now discuss differentiation of functions $f : X \to \mathbb{R}$; for convenience, we assume[6] that $f \in L^+(X,\mu)$. For these functions, we can define a new measure ν on $\mathcal{B}$ by $\nu(B) := \int_B f \,\mathrm{d}\mu := \int f \cdot \chi_B \,\mathrm{d}\mu$, which measures the 'area under the curve'. Note that for each $B \in \mathcal{B}$, we have B is ν-null whenever B is μ-null.

It turns out that the 'correct' setting to discuss differentiation is between two measures μ and ν which satisfy the above condition.

Definition 2.19. If μ, ν are measures on a measurable space $(X,\mathcal{B})$ and B is ν-null whenever B is μ-null for each $B \in \mathcal{B}$, we say that ν is *absolutely continuous w.r.t* μ, and write $\nu \ll \mu$.

Theorem 2.20 (Lebesgue-Radon-Nikodym Theorem). *If $\nu \ll \mu$ are σ-finite measures on a measurable space $(X,\mathcal{B})$, then there exists a $\mathcal{B}$-measurable map $f : X \to \mathbb{R}_{\geq 0}$ such that $\nu(B) = \int_B f \,\mathrm{d}\mu$ for all $B \in \mathcal{B}$.*

Such a function $f : X \to \mathbb{R}_{\geq 0}$ is unique μ-a.e., and is called the *Radon-Nikodym derivative of ν w.r.t. μ*, denoted $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Thus, we have $\nu(B) = \int_B \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu$ for all $B \in \mathcal{B}$.

Corollary 2.21. *In the above setting, we have $\int g \,\mathrm{d}\mu = \int g \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\mathrm{d}\nu$ for all $g \in L^1(X,\mu)$.*

To relate $\mathrm{d}\nu/\mathrm{d}\mu$ to derivatives in calculus (say on $\mathbb{R}^n$), we let $\mu := \lambda$ be Lebesgue measure on $\mathbb{R}^n$.

Theorem 2.22 (Lebesgue Differentiation Theorem). *For any locally-integrable function $f : \mathbb{R}^n \to \mathbb{R}$ (i.e. if $f \cdot \chi_K \in L^1(\mathbb{R}^n, \lambda)$ for every compact $K \subseteq \mathbb{R}^n$) and for λ-a.e. $x \in \mathbb{R}^n$, we have*

$$f(x) = \lim_{\varepsilon \to 0} \frac{1}{\lambda(B_\varepsilon(x))} \int_{B_\varepsilon(x)} f \,\mathrm{d}\lambda.$$

Corollary 2.23. *For any locally-finite Borel measure $\mu \ll \lambda$ on $\mathbb{R}^n$ and for λ-a.e. $x \in \mathbb{R}^n$, we have*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x) = \lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))}.$$

We end by briefly mentioning the 'Isomorphism Theorems'. These justify why we only gave three examples in Lecture 1, and allows us to work in concrete spaces like $[0,1]$ or $2^{\mathbb{N}}$.

Definition 2.24. A measurable space $(X,\mathcal{B})$ is said to be *standard Borel* if $\mathcal{B}$ is the Borel σ-algebra of some Polish (i.e. separable and completely metrizable) topology on X.

A probability space $(X,\mathcal{B},\mu)$ is *standard* if $(X,\mathcal{B})$ is standard Borel.

Theorem 2.25 (Borel Isomorphism Theorem). *Any two uncountable standard Borel spaces are Borel isomorphic. In particular, they all have cardinality continuum and are Borel isomorphic to $2^{\mathbb{N}}$.*

Definition 2.26. Let $(X,\mathcal{B})$ and $(Y,\mathcal{C})$ be measurable spaces. If $f : X \to Y$ is $(\mathcal{B},\mathcal{C})$-measurable and μ is a measure on $\mathcal{B}$, the *pushforward measure of μ by f* is the measure $f_*\mu$ on $\mathcal{C}$ defined by $f_*\mu(C) := \mu(f^{-1}(C))$.

Definition 2.27. Two measure spaces $(X,\mathcal{B},\mu)$ and $(Y,\mathcal{C},\nu)$ are said to be a *measure isomorphic* if there is a *measure-preserving transformation* $f : X \to Y$, i.e., a map $f : X \to Y$ such that $f_*\mu = \nu$, and such that there is a μ-conull set $X_0 \subseteq X$ and a ν-conull set $Y_0 \subseteq Y$ on which f and f^{-1} restrict to Borel isomorphisms.

Theorem 2.28 (Measure Isomorphism Theorem). *Any two atomless standard probability spaces are measure isomorphic. In particular, they are all measure isomorphic to* $([0,1],\lambda)$.

3. Lecture 3 (Peng Bo): An introduction to Ergodic Theory

Generally speaking, a *dynamical system* is just a group action. For instance, actions of $\mathbb{Z}$ (resp. $\mathbb{R}$) on some space X lead to *discrete* (resp. *continuous*) dynamical systems, and different actions on different spaces are studied from different point of views:

1. Continuous actions on topological spaces lead to *topological dynamics*.

[6] Otherwise, we will need to discuss 'signed measures'.

2. Measure preserving actions on measure spaces and lead to *measured group theory* and *ergodic theory.*
3. Geometric (proper and cocompact) actions on geodesic metric spaces lead to *geometric group theory.*
4. Linear actions on Banach spaces (or TVSs) lead to *functional analysis* and C^**-algebras.*

Here, we will be interested in measure preserving actions on measure spaces. To motivate these actions, we will also mention the theory of continuous actions on topological spaces, which was the original motivation for von-Neumann to introduce measure preserving actions in the first place. For simplicity, we will focus on actions of $\mathbb{Z}$ (so it suffices to provide a single generator) on compact (resp. probability) spaces.

Definition 3.1. A *topological dynamical system* is a pair (X, f), consisting of a compact topological space X and a homeomorphism $f : X \to X$.

Definition 3.2. A *measure-preserving dynamical system* is a tuple (X, μ, T) consisting of a standard probability space (X, μ) and a measure-preserving transformation $T \in \mathrm{MPT}(X, \mu)$.

Definition 3.3. A topological system (X, f) is *minimal* if there is no proper subsystem, i.e. if there is no f-invariant compact subspace $Y \subseteq X$.

Exercise 3.4. A topological system (X, f) is minimal if every f-orbit in X is dense. HINT: If $\{f^n(x)\}_{n\in\mathbb{Z}}$ is not dense, then its closure is a proper subsystem.

The measure-theoretic analogue of minimal system is an *ergodic* system. As we shall see, these are much more amenable than their topological-counterparts.

Definition 3.5. A measure-preserving system (X, μ, T) is *ergodic* if every T-invariant Borel subset is either μ-null or μ-conull.

Example 3.6. Consider the circle $S^1 := \mathbb{R}/\mathbb{Z} \cong [0, 1)$ and an irrational angle $\theta \in [0, 1)$. Then the *(irrational) rotation* $R_\alpha : S^1 \to S^1$ given by $x \mapsto x + \theta$ is both minimal and ergodic.

We will show that R_α is ergodic in Proposition 4.9. To see that R_α is minimal, it suffices to show that the orbit $\{n\theta\}_{n\in\mathbb{Z}}$ approaches 0. Indeed, there exists some $n \in \mathbb{Z}$ such that $[n\theta, (n + 1)\theta]$ contains 0 in the quotient. Without loss of generality, we can assume that $d(0, (n + 1)\theta) \leq d(0, n\theta)$, so $\theta_1 := (n + 1)\theta \leq \theta/2$ in the quotient and is in the orbit of θ. Replacing θ by θ_1 and repeating furnishes a sequence $\theta_n \to 0$.

In contrast to the above example, every homeomorphism on S^2 has a fixed point by Brower's fixed point theorem. Thus, there is no minimal system on S^2. More generally, there is no minimal system in S^{2n}.

Open Question 3.7. Characterize all manifolds supporting a minimal system. REMARK: Negative results include the unit interval and S^{2n}. This question is still open, even in dimension three.

Of course, any such characterization is modulo the isomorphism relation.

Definition 3.8. Two topological dynamical systems (X, f) and (Y, g) are *isomorphic* if there exists a homeomorphism $h : X \to Y$ such that $g = h \circ f \circ h^{-1}$.

The Measure Isomorphism Theorem (Theorem 2.28) asserts that any two atomless standard probability spaces are measure isomorphic. This allows us to reduce the classification of (atomless) measure-preserving dynamical systems to just measure-preserving transformations.

Open Question 3.9. Classify ergodic measure-preserving transformations up to isomorphism.

Here is a celebrated positive result.

Definition 3.10. A *Bernoulli shift* is a measure-preserving dynamical system $(\Sigma^{\mathbb{Z}}, \pi^{\mathbb{Z}}, S)$ where Σ is a finite set, π is a probability measure on Σ, and S is the *left-shift* map $S(x)(n) := x(n + 1)$.

The *entropy* of a Bernoulli shift is $-\sum_{i\leq|\Sigma|} \pi(i) \log(\pi(i))$.

Theorem 3.11 (Ornstein)**.** *Two Bernoulli shifts are isomorphic iff their entropies coincide.*

Lastly, we end with a recurrence question: given a μ-measurable subset $U \subseteq X$ and a point $x \in X$, how many points in the orbit of x are in U?

Theorem 3.12 (Birkhoff's Pointwise Ergodic Theorem; Theorem 5.1)**.** *Let (X, μ, T) be an ergodic system and let $U \subseteq X$ be a μ-measurable subset. Then $\lim_n |\{i < n : T^i x \in U\}|/n = \mu(U)$ for μ-a.e. $x \in X$.*

4. Lecture 4 (Zhaoshen Zhai): Examples of Ergodic Transformations

Following Lecture 3, we begin by studying actions of $\mathbb{N}$ on a standard probability space (X,μ), which is generated by a (probability) measure-preserving transformation $T : X \to X$, called a *pmp transformation*.

Throughout, let (X,μ) be a standard probability space and let $T : X \to X$ be a pmp transformation.

Further reading. [Tse22, Lectures 1 to 4].

Definition 4.1. The *orbit equivalence relation* of T is the equivalence relation $\mathbb{E}_T \subseteq X^2$ defined by $x\mathbb{E}_Ty$ iff $T^n(x) = T^m(y)$ for some $n,m \in \mathbb{N}$. The *forward orbit* of a point $x \in X$ is the set $\{T^n(x)\}_{n\in\mathbb{N}}$.

Definition 4.2. Let E be an equivalence relation on a set X.

1. A subset $A \subseteq X$ is *E-invariant* if A is a union of E-classes. The *E-saturation* of A is $[A]_E :=\bigcup_{x\in A}[x]_E$, which is clearly $\mathbb{E}$-invariant, and A is E-invariant iff $A = [A]_E$.
2. A function $f : X \to Y$ is *E-invariant* if f is constant on each E-class.

We say that A (or f) is *T-invariant* if it is $\mathbb{E}_T$-invariant.

Remark 4.3. Note that $A \subseteq X$ is T-invariant iff $T^{-1}(A) = A$, and $f : X \to Y$ is T-invariant iff $f \circ T = f$.

Observe that $[A]_{\mathbb{E}_T} = \bigcup_{n,m\in\mathbb{Z}} T^{-n}(T^m(A))$, so $[A]_{\mathbb{E}_T}$ is not *a priori* measurable. It turns out that $[A]_{\mathbb{E}_T}$ is measurable by a theorem of Descriptive Set Theory (which one?), but we can avoid it with the following theorem, which is of independent interest.

Theorem 4.4 (Poincaré Recurrence). *Every measurable set $A \subseteq X$ is a.e.-forward recurrent, i.e., there is a measurable set $A_0 \subseteq X$ such that $A_0 =_\mu A$ and for each $x \in A_0$, we have $T^n(x) \in A_0$ for some $n \geq 1$.*

Proof. Let $W := \bigcap_{n\geq1} \{x \in A : T^n(x) \notin A\}$, which is clearly measurable. Note that $W \cap T^{-n}(W) = \varnothing$ for each $n \geq 1$, so the family $\{T^{-n}(W)\}_{n\in\mathbb{N}}$ is pairwise-disjoint, and hence W is *wandering*:

> **Definition 4.5.** A set $W \subseteq X$ is *T-wandering* if the family $\{T^{-n}(W)\}_{n\in\mathbb{N}}$ is pairwise-disjoint.
>
> **Lemma 4.6.** *Every measurable wandering set T is null.*
>
> *Proof.* $\sum_n \mu(W) = \sum_n \mu(T^{-n}(W)) = \mu(\bigsqcup_n T^{-n}(W)) \leq \mu(X) < \infty$, so $\mu(W) = 0$. □

Set $Z := \bigcup_n T^{-n}(W)$, which is still null, and note that $A_0 := A \setminus Z =_\mu A$ is forward recurrent. ■

Corollary 4.7. *For all measurable sets $A \subseteq X$, there exists $A_0 =_\mu A$ such that $[A_0]_{\mathbb{E}_T} = \bigcup_n T^{-n}(A_0)$.*

Definition 4.8. An equivalence relation $E \subseteq X^2$ on (X,μ) is *ergodic* if every E-invariant measurable set $A \subseteq X$ is either null or conull. A pmp $T : X \to X$ is *ergodic* if $\mathbb{E}_T$ is ergodic.

Proposition 4.9. *The irrational rotation $R_\alpha : S^1 \to S^1$ is ergodic w.r.t. the Lebesgue measure on S^1.*

Proof. We will need the following lemma.

> **Lemma 4.10** (99% Lemma for λ). *For any $A \subseteq [0,1)$ with positive Lebesgue measure, there is an interval $I \subseteq [0,1)$ such that at-least 99% of I is covered by A, i.e., $\lambda(A\cap I)/\lambda(I) \geq 0.99$.*
>
> *Proof.* Fix $\varepsilon > 0$ and (by outer regularity of λ) let $U \subseteq [0,1)$ be open such that $\lambda(A)/\lambda(U) \geq 1-\varepsilon$. Write $U = \bigsqcup_{n\in\mathbb{N}} I_n$ for disjoint open intervals $I_n \subseteq [0,1)$, and observe that
>
> $$\frac{\lambda(A)}{\lambda(U)} = \frac{1}{\lambda(U)}\sum_{n\in\mathbb{N}}\lambda(A\cap I_n) = \sum_{n\in\mathbb{N}}\frac{\lambda(I_n)}{\lambda(U)}\frac{\lambda(A\cap I_n)}{\lambda(I_n)} \geq 1-\varepsilon.$$
>
> Hence a convex combination of $\{\lambda(A\cap I_n)/\lambda(I_n)\}_n$ is at-least $1-\varepsilon$, so the result follows. □

Now, suppose towards a contradiction that R_α is not ergodic, so there exists an R_α-invariant measurable $A \subseteq [0,1)$ such that both A and A^c have positive λ-measure. By the 99% Lemma, let $I \subseteq [0,1)$ (resp. J) be an interval such that 99% of I is covered by A (resp. A^c); without loss of generality, suppose that $|J| \leq |I|$.

By R_α-invariance, 99% of any translate of J is still covered by A^c, and so it suffices to cover at-least half of I by translates of J, for then $99\%/2 > 1\%$ of I is covered by A^c, a contradiction. This can be done by minimality of R_α (i.e., density of any R_α-orbit), since we can translate the left-endpoint of J arbitrarily close to the right-endpoint of previous translates of J. ■

Here is a cute application of the ergodicity of R_α. For any map $f : X \to X$, its *graph* is the set $G_f \subseteq X^2$ of pairs $(x, f(x))$. We can view G_f as an abstract graph with vertex set X and with edges $(x, f(x))$.

What is the chromatic number of G_{R_α}? Since R_α is not periodic, each connected component is a $\mathbb{Z}$-line $\{R_\alpha^n(x)\}_{n\in\mathbb{Z}}$, and G_{R_α} is the disjoint union of continuum-many such $\mathbb{Z}$-lines. Using the Axiom of Choice, we can pick a point in each $\mathbb{Z}$-line, so we can 2-color G_{R_α} by coloring said points, say blue, and alternating.

Note that a (finite) coloring of a graph G is just a map $c : G \to n$ for some $n \in \mathbb{N}$ such that if $(x, y) \in G$, then $c(x) \neq c(y)$. Thus we can ask for the *measurable chromatic number* of G: what is the minimal $n \in \mathbb{N}$ such that there is a measurable colouring $c : G \to n$? Clearly, G_{R_α} is measurably 3-colourable.

Corollary 4.11. *The graph $G_{R_\alpha} \subseteq X^2$ is not measurably 2-colourable.*

Proof. If it is, then there is a measurable colour $A \subseteq X$ of G_{R_α} such that $R_\alpha(A) = A^c$, so $\lambda(A) = \lambda(A^c) = 1/2$. Then $R_{2\alpha}(A) = R_\alpha^2(A) = A$, so A is a measurable $R_{2\alpha}$-invariant set. Since 2α is irrational, we see that $R_{2\alpha}$ is ergodic, and hence $\mu(A) \in \{0, 1\}$, a contradiction. ■

Proposition 4.12. *The Bernoulli shift $T : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ is ergodic w.r.t. the Bernoulli$(1/2)$ measure μ on $2^{\mathbb{N}}$.*

Proof. We will prove a stronger result, which implies that T is ergodic.

Definition 4.13. A pmp $T : X \to X$ is said to be *mixing* if for any measurable $A, B \subseteq X$, we have $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ as $n \to \infty$.

Lemma 4.14. *Mixing implies ergodic.*

Proof. If $A \subseteq X$ is a T-invariant, then $\mu(A) = \mu(A \cap T^{-n}A) \to \mu(A)^2$, so $\mu(A) \in \{0, 1\}$. □

Since μ is defined by extending the $1/2$-measure on cylinder sets, it suffices to show that T mixes cylinders $A := [s]$ and $B := [t]$ for $s, t \in 2^{<\mathbb{N}}$. Indeed, let $n \geq \ell(s)$ so that the translate $T^{-n}([t])$ contains sequences specified at indices disjoint from that of s, so $\mu([s] \cap T^{-n}([t])) = \mu([s])\mu([t])$. ■

To give more examples, we will need the following lemma, which is proved the same way as Lemma 4.10.

Lemma 4.15 (99% Lemma for μ)**.** *For any measurable $A \subseteq 2^{\mathbb{N}}$, there exists a cylinder $[w] \subseteq 2^{\mathbb{N}}$ such that at-least 99% of $[w]$ is covered by A, i.e. $\mu(A \cap [w])/\mu([w]) \geq 0.99$.*

Exercise 4.16. Show that the *odometer* transformation $T : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ is ergodic, where T takes a sequence, thought of as the binary representation of a number written in reverse, and adds 1 to it (carrying over if necessary); for instance, $T(00110\dots) := 10110\dots$ and $T(11100\dots) := 00010\dots$, and by convention, we let $T(11111\dots) := 00000\dots$.

Hint: Use the 99% Lemma for μ, with the observation that for any words $s, t \in 2^{<\mathbb{N}}$ of the same length, there exists $k \in \mathbb{N}$ such that $T^k(sx) = tx$ for all $x \in 2^{\mathbb{N}}$.

Let us generalize ergodicity of transformations $T : (X, \mu) \to (X, \mu)$ to actions of a group G on (X, μ).

Definition 4.17. Let G be a group and let (X, μ) be a standard Borel space. An action $\varphi : G \curvearrowright X$ is said to be *Borel* if for each $g \in G$, the map $x \mapsto gx$ is Borel; *measure-preserving* if it is Borel and $\mu(gB) = \mu(B)$ for each $g \in G$ and each Borel $B \subseteq X$; and *ergodic* if it is measure-preserving and the orbit equivalence relation $\mathbb{E}_\varphi$ of φ, given by $x\mathbb{E}_\varphi y$ iff $y = gx$ for some $g \in G$, is ergodic.

Exercise 4.18. For each $n \in \mathbb{N}$, let $\sigma_n : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ be the *n^{th}-bit flip* map, defined by flipping x_n to $1 - x_n$ and fixing all other coordinates. Let $G := \langle \sigma_n \rangle_{n\in\mathbb{N}} \cong \bigoplus_n \mathbb{Z}/2\mathbb{Z}$, which naturally acts on $2^{\mathbb{N}}$.

1. Show that the orbit equivalence relation $\mathbb{E}_\varphi$ is given by *eventual equality* (denoted $\mathbb{E}_0$), where $x\mathbb{E}_0 y$ iff there exists $N \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq N$.
2. Observe that φ is a pmp action and use the 99% Lemma for μ to show that φ is ergodic.

Exercise 4.19. Consider the translation action $\varphi : \mathbb{Q} \curvearrowright (\mathbb{R}, \lambda)$, whose orbit equivalence relation is given by $x\mathbb{E}_{\mathbb{Q}}y$ iff $x - y \in \mathbb{Q}$. Use the 99% Lemma for λ to show that φ is ergodic.

The following exercise shows that ergodicity gives rise to non-measurable transversals.

Exercise 4.20. Let (X, μ) be an atomless measure space and let $\varphi : G \curvearrowright (X, \mu)$ be a μ-null-preserving action. Prove that if φ is ergodic, then every transversal of $\mathbb{E}_\varphi$ is non-measurable.

HINT: Let $T \subseteq X$ be a measurable transversal, so $X = \bigsqcup_{g\in G} gT$. Observe that $\mu(T) > 0$, and use that (X, μ) is atomless to partition $T = S_1 \sqcup S_2$ non-trivially. What can you say about the $\mathbb{E}_\varphi$-saturations of S_i?

To study ergodic transformations further, it would be useful to have alternative characterizations of ergodicity. The *Ergodic Theorems* (see Lecture 5) are the strongest results of this kind; here, we will be content with the following easy reformulations.

Theorem 4.21. *The following are equivalent for a pmp transformation $T : X \to X$.*

1. *T is ergodic.*
2. *Every T-invariant measurable function $f : X \to Y$ to a standard Borel space Y is constant a.e..*
3. *For every positively-measured subset $A \subseteq X$, its saturation $[A]_{\mathbb{E}_T}$ is conull.*

Proof. Note that (2) and (3) each easily imply (1): if $A \subseteq X$ is a measurable T-invariant set, then

2. the characteristic function $\chi_A : X \to \{0, 1\}$ is T-invariant, and hence either A is null or conull;
3. if A is not null, then $\mu(A) > 0$, and hence $A = [A]_{\mathbb{E}_T}$ is conull.

Conversely, suppose that T is ergodic. For (3), note that if A has positive measure, then so must $[A]_{\mathbb{E}_T}$ *if it is measurable in the first place*, so $[A]_{\mathbb{E}_T}$ is conull. Now, either invoke some Descriptive Set Theory and prove that $[A]_{\mathbb{E}_T}$ is measurable, or proceed by letting $A_0 \subseteq_\mu A$ be forward recurrent, so that $[A_0]_{\mathbb{E}_T} \subseteq [A]_{\mathbb{E}_T}$ is measurable by Corollary 4.7; hence $[A]_{\mathbb{E}_T}$ is measurable too, as desired.

It remains to prove (2). By the Borel Isomorphism Theorem (Theorem 2.25), it suffices to prove it in the case when $Y := 2^{\mathbb{N}}$ equipped with the Borel σ-algebra, which is generated by cylinders. Let $f : X \to 2^{\mathbb{N}}$ be a T-invariant measurable function, so $f^{-1}(B)$ is T-invariant for each Borel $B \subseteq 2^{\mathbb{N}}$. In particular, $f^{-1}([w])$ is either null or conull for each word $w \in 2^{<\mathbb{N}}$. We proceed by finding a (necessarily unique) sequence $y \in 2^{\mathbb{N}}$ such that $f^{-1}(y)$ is conull, as follows. Call a word $w \in 2^{<\mathbb{N}}$ *heavy* if $f^{-1}([w])$ is conull; clearly $\varnothing$ is heavy, and if w is heavy, then exactly one of $w0$ and $w1$ is heavy. Thus there is a unique heavy branch $\{y|n\}_{n\in\mathbb{N}}$, which gives rise to the desired sequence $y \in 2^{\mathbb{N}}$ since $f^{-1}(y) = \bigcap_n f^{-1}([y|n])$ is the intersection of countably-many conull sets, hence conull. ∎

5. Lecture 5 (Zhaoshen Zhai): Birkhoff's Pointwise Ergodic Theorem

Throughout, let (X, μ, T) be a measure-preserving dynamical system. We can rephrase $(1 \Leftrightarrow 3)$ in Theorem 4.21 by the statement that T is ergodic iff for every positively-measured subset $A \subseteq X$ and any $x \in X$, we have $[x]_{\mathbb{E}_T} \cap A \neq \varnothing$. This gives us a 'soft'/'qualitative' characterization of ergodicity: every orbit meets every positively-measured set. It turns out that we can boost this and obtain a more quantitative statement too, of, say, how *often* they meet.

Further reading. [Tse22, Lectures 5 to 7].

Theorem 5.1 (Birkhoff's Pointwise Ergodic Theorem)**.** *A measure-preserving dynamical system (X, μ, T) is ergodic iff any of the following statements hold, where $I_n^T(x) := \{T^i x\}_{i<n}$ and $A_n^T f := \frac{1}{n}\sum_{i<n} f \circ T^i$.*

1. *For all $f \in L^1(X, \mu)$, we have $\lim_{n\to\infty} A_n^T f =_\mu \int f \,\mathrm{d}\mu$.*
2. *For all $f \in L^\infty(X, \mu)$, we have $\lim_{n\to\infty} A_n^T f =_\mu \int f \,\mathrm{d}\mu$.*

3. For all measurable $A \subseteq X$, we have $\lim_{n\to\infty} \frac{1}{n}|I_n^T(x) \cap A| = \mu(A)$ for μ-a.e. $x \in X$.

We can interpret (3) by saying that the average number of times the forward-orbit of μ-a.e. $x \in X$ meets A tends to $\mu(A)$, so if $\mu(A) > 0$, then the (forward-)orbit of x will always meet A. Statements (1) and (2) are generalizations of (3); instead of counting the average number of times $I_n^T(x)$ meets A, we consider the average value of a measurement $f : X \to \mathbb{R}$ (i.e., $f \in L^1(X, \mu)$ or $f \in L^\infty(X, \mu)$) in the following two senses:

- (Time average). For $x \in X$, the *time average* value of f evaluated at x is $A_n^T f(x) := \frac{1}{n}\sum_{i<n} f(T^i(x))$.
- (Space average). The *space average* $\int f \,\mathrm{d}\mu$ of f, which is the average evaluated on the entire system.

Birkhoff's Pointwise Ergodic Theorem then states that these two averages coincide iff T is ergodic.

Proof. Clearly (1) $\Rightarrow$ (2). For (2) $\Rightarrow$ (3), take $f := \chi_A$ and note that $\frac{1}{n}|I_n^T(x) \cap A| = A_n^T f(x)$. If (3) holds, then T is ergodic since if $A \subseteq X$ is a measurable T-invariant set, then the orbit of any $x \in X$ either lies in A or A^c, so $\frac{1}{n}|I_n^T(x) \cap A| \in \{0, 1\}$ uniformly for all $n \in \mathbb{N}$. By (3), this shows that $\mu(A) \in \{0, 1\}$.

Suppose now that T is ergodic and let $f \in L^1(X, \mu)$. Replacing f by $f - \int f \,\mathrm{d}\mu$, we can assume that $\int f \,\mathrm{d}\mu = 0$. To show that $\lim_n A_n^T f = 0$, it suffices by symmetry to show that $\limsup_n A_n^T f \le 0$.

To this end, we first note that $l := \limsup_n A_n^T f : X \to \mathbb{R}$ is T-invariant. Indeed, for any $x \in X$, we have

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

so taking $\limsup_n$ gives $l(x) = l(Tx)$ as desired. Thus l is constant a.e. by Theorem 4.21, say at $l_0 \in \mathbb{R}$.

Suppose towards a contradiction that $f^* := l_0/2 > 0$, so for each $x \in X$, there is a minimal $\eta(x) \in \mathbb{N}$ such that $A_{\eta(x)}^T f(x) \ge f^*$. Thus, we can cover X by intervals $I_x := I_{\eta(x)}^T x$ so that the average of f on those intervals is no less than f^*. If we can make the lengths of those intervals uniform, i.e., find some $n \in \mathbb{N}$ such that $A_n^T f \ge f^*$, then

$$0 = \int f \,\mathrm{d}\mu \overset{!}{=} \int A_n^T f \,\mathrm{d}\mu \ge \int f^* \,\mathrm{d}\mu = f^* > 0, \tag{$*$}$$

a contradiction, where the equality (!) follows from the following

Lemma 5.2 (Local-global Bridge). *For each $f \in L^1(X, \mu)$ and $n \in \mathbb{N}$, we have $\int f \,\mathrm{d}\mu = \int A_n^T f \mathrm{d}\mu$.*

Proof. Since T is pmp, we have $\int f \,\mathrm{d}\mu = \int f \circ T \,\mathrm{d}\mu$ by the Change of Variables formula, so $\int \sum_{i<n} f \circ T^i = n \int f \,\mathrm{d}\mu$, and hence the desired equality holds. □

This is too much to ask for in general; instead, we try to cover intervals $I_n^T(x)$ for large enough $n \gg 0$ by the 'good' intervals of the form $I_{\eta(y)}^T(y) := I_y$, on which we have the desired inequality $A_{\eta(y)}^T f \ge f^*$:

Lemma 5.3 (Tiling Lemma). *Let $\eta : X \to \mathbb{N}$ be an arbitrary measurable function. For any $\varepsilon > 0$, there exists $n \gg 0$ such that for each $x \in X$ except on a measure-ε set, the interval $I_n^T(x)$ can be tiled, up to an ε-fraction, by intervals of the form $I_y := I_{\eta(y)}^T(y)$ for $y \in X$.*

Proof. Choose $L \gg 0$ such that $B := \{x \in X : \eta(x) > L\}$ has measure at-most $\varepsilon^2/2$, so for each $n \in \mathbb{N}$, the set $Z_n := \{x \in X : A_n^T \chi_B(x) < \varepsilon/2\}$ is co-ε since, by the Local-global Bridge, we have

$$\varepsilon^2/2 \ge \mu(B) = \int \chi_B \,\mathrm{d}\mu \overset{!}{=} \int A_n^T \chi_B \,\mathrm{d}\mu \ge \int_{X \setminus Z_n} A_n^T \chi_B \,\mathrm{d}\mu \ge \frac{\varepsilon}{2}\mu(X \setminus Z_n).$$

For each $x \in Z_n$, we can tile $I_n^T(x)$ from left to right, skipping 'bad' intervals (i.e., intervals I_y with $y \in B$), which leaves out at-most an $(\varepsilon/2 + L/n)$-fraction of $I_n^T(x)$ untiled by the I_y's; choose $n \gg 0$ such that $L/n < \varepsilon/2$. □

With this in hand, we can start to attempt to replicate ($*$). To this end, first choose $M \gg 0$ so that $X_0 := \{x \in X : f \ge -M\}$ is large; more specifically, so that $\int_{X_0^c} (f - f^*)\mathrm{d}\mu \le f^*/2$, and hence $\int_{X_0^c} f \,\mathrm{d}\mu \ge f^*/2$.

Focus on $f_0 := f|_{X_0}$. Applying the Tiling Lemma to some $\varepsilon > 0$ to be chosen later, there is some $n \gg 0$ and some $Z \subseteq X_0$ of measure at-least $\mu(X_0) - \varepsilon$ such that for each $x \in Z$, the interval $I_n^T(x)$ is tiled by I_y's up to an ε-fraction. Since $f_0 \ge f$, we have $A_{\eta(y)}^T f_0(y) \ge f^*$ too, so for all $x \in Z$,

$$A_n^T f_0(x) \geq (1-\varepsilon) f^* + \varepsilon(-M) = f^* - \varepsilon(M + f^*) \geq f^*/2$$

for sufficiently small $\varepsilon > 0$. We can now replicate $(*)$ to obtain

$$0 = \int f \, \mathrm{d}\mu = \int_{X_0^c} f \, \mathrm{d}\mu + \int_Z f_0 \, \mathrm{d}\mu + \int_{X_0 \setminus Z} f_0 \, \mathrm{d}\mu \geq \frac{f^*}{2} + \int_Z A_n^T f_0 \, \mathrm{d}\mu + \mu(X_0 \setminus Z)(-M) \geq f^* - \varepsilon M > 0$$

for sufficiently small $\varepsilon > 0$, as desired. ■

Remark 5.4. If $f \in L^*(X, \mu)$ and η is bounded, we can tile *every* interval $I_n^T(x)$ for sufficiently large $n \gg 0$ by the I_y's, up to an ε-fraction. This in turn simplifies the proof so that $A_n^T f \geq f^*$ for a uniform $n \in \mathbb{N}$, so that $(*)$ holds.

Exercise 5.5. What is the average value of a given digit $0 \leq m \leq 9$, say $m := 7$, to occur in the decimal representation of λ-a.e. $x \in [0,1]$? That is, does $\ell_m(x) := \lim_n \frac{1}{n} |\{i < n : x_i = m\}|$ exist, and what is it?

HINT: Consider the 10-ary Baker's map $b_{10} : [0,1) \to [0,1)$ sending $x \mapsto 10x \pmod 1$, which is isomorphic to the shift map on $10^{\mathbb{N}}$.

Exercise 5.6 (Equidistribution Theorem)**.** A sequence $(x_n)_n$ in S^1 is said to be *equidistributed* if for every interval $I \subseteq S^1$, we have $\lim_n \frac{1}{n} | \{x_i\}_{i<n} \cap I| = \lambda(I)$. Prove that if $x_n = n\alpha$ for some irrational $\alpha \in S^1$, then $(x_n)_n$ is equidistributed. HINT: Don't overthink it.

Exercise 5.7 (Law of Large Numbers)**.** If you know statistics, prove it!

Exercise 5.8 (An ergodic theorem for non-ergodic actions)**.** Intuitively, Birkhoff's Pointwise Ergodic Theorem states that ergodic transformations $T : X \to X$ stir up X so well that they spread any $f \in L^1(X, \mu)$ evenly on X, making it constant at $\int f \, \mathrm{d}\mu$; indeed, '$f \circ T^\infty = \int f \, \mathrm{d}\mu$'.

If T is not ergodic, then there is a non-trivial partition $X = X_1 \sqcup X_2$ into T-invariant pieces. The best that one can hope is at after 'enough' partitions $X = \bigsqcup_i X_i$, T still spreads each $f_i := f \chi_{X_i}$ evenly on X_i. Viewing f from the lens of these T-invariant pieces leads to the *conditional expectation* of f:

Definition 5.9. Let $\mathcal{A} \subseteq \mathcal{B}(X)$ be a sub-σ-algebra of $\mathcal{B}(X)$. For each $f \in L^1(X, \mu)$, there is a unique (up to a μ-null set) $\mathcal{A}$-measurable function $f_{\mathcal{A}}$ such that $\int_A f \, \mathrm{d}\mu = \int_A f_{\mathcal{A}} \, \mathrm{d}\mu$ for each $A \in \mathcal{A}$, called the *conditional expectation of f w.r.t. $\mathcal{A}$*. We write $\mathbb{E}(f|\mathcal{A})$ for $f_{\mathcal{A}}$.

Remark 5.10. If $\mathcal{P} \subseteq \mathcal{B}(X)$ is a countable partition of X, then $\mathbb{E}(f|\langle \mathcal{P} \rangle_\sigma) = \sum_{P \in \mathcal{P}} \left(\frac{1}{\mu(P)} \int_P f \, \mathrm{d}\mu \right) \chi_P$.

Prove that for any (not necessarily ergodic) pmp transformation $T : X \to X$ and any $f \in L^1(X, \mu)$, we have $\lim_n A_n^T f =_\mu \mathbb{E}(f|\mathcal{B}_T)$, where $\mathcal{B}_T \subseteq \mathcal{B}(X)$ is the σ-algebra generated by all T-invariant Borel sets of X.

HINT: Same as the regular proof, only that $f^* : X \to \mathbb{R}$ is not necessarily constant, but just T-invariant.

Exercise 5.11 (L^p-ergodic theorem)**.** Prove that for any pmp-transformation $T : X \to X$ and $p \geq 1$, we have $A_n^T f \to_{L^p} \mathbb{E}(f|\mathcal{B}_T)$ for all $f \in L^p(X, \mu)$.

HINT: If f is bounded, then we are done by the DCT. Otherwise, let $f_k \to_{L^p} f$ where each f_k is bounded and triangle-inequality your way through, using that $\|A_n^T f\|_{L^p} \leq \|f\|_{L^p}$ (prove this too).

6. Lecture 6 (Ludovic Rivet): An overview of Szemerédi's Theorem

This is the first in a series of lectures towards Furstenberg's proof of Szemerédi's Theorem.

Further reading. Parts of [Tse22, Lectures 16 and 17] and [Tao08, Lecture 10].

Conjecture 6.1 (Erdős, Turán)**.** *Let $A \subseteq \mathbb{N}$ be a set such that $\sum_{a \in A} 1/a \to \infty$. Then A contains arbitrarily long arithmetic progressions, i.e., for any $k \geq 1$, there exists $n \in \mathbb{N}$ and $r \geq 1$ such that $\{n + ir : i < k\} \subseteq A$.*

Green and Tao proved, in 2004, that this is true when A is the set of prime numbers. Here, we give an overview a much easier positive instance of this conjecture, namely, when the set A is 'dense enough'.

Definition 6.2. The *(Banach) density* of a subset $A \subseteq \mathbb{N}$ is $d_b(A) := \lim_n \frac{1}{n} |A \cap \{0, \ldots, n-1\}|$. Replacing 'lim' with 'lim sup', we get the *upper (Banach) density* $\overline{d}_b(A)$ of A.

Theorem 6.3 (Szemerédi; 1975)**.** *Conjecture 6.1 is true for those $A \subseteq \mathbb{N}$ such that $\overline{d}_b(A) > 0$. In fact, for any $k \geq 1$, there exists $n \in \mathbb{N}$ such that $\overline{d}_b(\bigcap_{i<k}(A - in)) > 0$.*

Exercise 6.4. Square-free integers[7] has positive density.

Exercise 6.5. The density of primes is zero (so the result by Green and Tao does not follow immediately).

In 1977, Furstenberg [Fur77] gave an ergodic theoretic proof of Szemerédi's Theorem. First, he established a naturally, but *very hard*, generalization of the Poincaré Recurrence Theorem, and showed that this result is equivalent to Szemerédi's Theorem via a correspondence principle.

Theorem 6.6 (Furstenberg's Multiple Recurrence)**.** *Let (X, μ, T) be a measure-preserving dynamical system. For any positive-measure $A \subseteq X$ and any $k \geq 1$, there exists $n \geq 1$ such that $\mu(\bigcap_{i<k} T^{-in}A) > 0$.*

The case $k = 1$ is trivial, and the case $k = 2$ is exactly Poincaré's Recurrence Theorem (Theorem 4.4). We will prove this theorem in later lectures, but for now, we can prove it for two examples:

1. For the irrational rotation $R_\alpha : S^1 \to S^1$ (see Proposition 4.9), we can find $n \in \mathbb{N}$ such that R_α^n is arbitrarily closed to the identity by solving the equation $n\alpha = 10^p$ where p denotes the 'precision'. By choosing a good enough approximation, the sets $R_\alpha^{in}(A)$ for $i < k$ intersect on a set of positive measure.
2. For the shift map $T : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ (see Proposition 4.12), note that $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ since T is mixing, so $\mu(\bigcap_{i<k} T^{-in}A) \to \mu(A)^k > 0$.

Note that Theorem 6.6 holds for drastically different reasons in the two examples above; R_α preserves the distance between points, while T mixes points a lot. This hints that the proof of the general case will require some deep idea that connects all measure-preserving systems together. In fact, we will see that they are all 'extensions' of a combination of systems of these two types.

The link between Theorems 6.3 and 6.6 is via *Furstenberg's Correspondence Principle.*

Theorem 6.7 (Furstenberg's Correspondence Principle on $\mathbb{N}$)**.** *For any $A \subseteq \mathbb{N}$, there is a measure-preserving dynamical system (X, μ, T) and a set $B \subseteq X$ such that for any $k \geq 1$ and $n_1, \ldots, n_{k-1} \in \mathbb{N}$, we have*

$$\overline{d}_b(A \cap (A - n_1) \cap (A - n_2) \cap \cdots \cap (A - n_{k-1})) \geq \mu(B \cap T^{-n_1}B \cap T^{-n_2}B \cap \cdots \cap T^{-n_{k-1}}B).$$

Combining Theorems 6.6 + 6.7, we obtain a proof of Theorem 6.3 by setting $n_i \coloneqq in$.

Proof of Theorem 6.7. By passing to a subsequence, we may assume that $\overline{d}_b(A) = \lim_n |A \cap \{0, \ldots, n-1\}|/n$. Consider the shift action $T : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ and let $B \coloneqq [1]$. Viewing $\chi_A : \mathbb{N} \to 2$ as an element of $2^{\mathbb{N}}$, we have that $T^j\chi_A \in B$ iff $j \in A$. We will define a measure μ on $X \coloneqq 2^{\mathbb{N}}$ with the desired property.

For each $n \in \mathbb{N}$, consider the weighted Dirac measure $\mu_n \coloneqq \frac{1}{n}\sum_{j<n} \delta_{T^j\chi_A}$ on X. By the Banach-Alaoglu Theorem, the space of probability measures on X is weak*-compact. Thus, after passing to a subsequence, there is a weak*-limit $\mu_n \to \mu$ so that (X, μ, T) is a measure-preserving dynamical system and

$$\mu(B) = \lim_{n\to\infty} \mu_n(B) = \lim_{n\to\infty} \frac{1}{n}\sum_{j<n} \delta_{T^j\chi_A}(B) = \lim_{n\to\infty} \frac{1}{n}|A \cap \{0, \ldots, n-1\}| = \overline{d}_b(A).$$

For each $i < k$, observe that $j \in A - n_i$ iff $T^{j+n_i}\chi_A \in B$, which occurs iff $\delta_{T^j\chi_A}(T^{-n_j}B) = 1$, and thus

$$\begin{aligned}\overline{d}_b(A \cap (A - n_1) \cap \cdots \cap (A - n_{k-1})) &\geq \liminf_{n\to\infty} \frac{|A \cap (A - n_1) \cap \cdots \cap (A - n_{k-1}) \cap \{0, \ldots, n-1\}|}{n} \\ &= \liminf_{n\to\infty} \frac{1}{n}\sum_{j<n} \delta_{T^j\chi_A}(B \cap T^{-n_1}B \cap \cdots \cap T^{-n_{k-1}}B) \\ &= \liminf_{n\to\infty} \mu_n(B \cap T^{-n_1}B \cap \cdots \cap T^{-n_{k-1}}B) \\ &= \mu(B \cap T^{-n_1}B \cap \cdots \cap T^{-n_{k-1}}B),\end{aligned}$$

as desired. ∎

[7]Squares don't work; Euler proved that the longest arithmetic progression in squares has length 3.

References

[Fol99] Gerald B. Folland. *Real Analysis. Modern Techniques and Their Applications*. 2nd ed. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, 1999.

[Fur77] Harry Furstenberg. "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions". In: *Journal d'Analyse Mathématique* 31 (1977), pp. 204–256.

[Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Springer New York, NY, 1995.

[Tao08] Terence Tao. *The Furstenberg correspondence principle*. URL: https://terrytao.wordpress.com/2008/02/10/254a-lecture-10-the-furstenberg-correspondence-principle/.

[Tse23] Anush Tserunyan. *MATH564: Advanced Real Analysis 1*. URL: https://www.math.mcgill.ca/atserunyan/Courses/2023F.Math564.Analysis%201/.

[Tse22] Anush Tserunyan. *MATH594: Topics in Ergodic Theory and Measured Group Theory*. URL: https://www.math.mcgill.ca/atserunyan/Courses/2022_W.Math594.Erg&MsGrp/.

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, H3A 0B9, Canada

Email address: zhaoshen.zhai@mail.mcgill.ca