Understanding Equidistribution and the pointwise ergodic theorem

Tomer Guralnik

Nov 2025

If you are interested in a deeper dive on these topics i recommend the following texts:

1. Measure theory - Cohn - Chapter 7 (Measures on locally compact spaces)
2. Ergodic Theory with a view towards Number Theory - Chapters 2, 4. Appendix B.5.

1 The weak* topology

In this section, we will go over some basic definitions and results from functional analysis.

Definition 1. *A topological vector space (TVS) V is a vector space over some field $\mathbb{K}$, to avoid complications we will assume that $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, that is endowed with some topology such that vector addition and scalar multiplication are continuous. We will assume that V is a normed space.*

The examples that will interest us most are:

1. $C(X)$ when X is compact and hausdorff, with the sup-norm. In this case, $C(X)$ is complete with the sup norm.
2. $C_c(X)$ when X is locally compact and hausdorff, with the sup-norm. In this case, $C_c(X)$ is not complete, and it's completion is $C_0(X)$ (continuous functions that vanish at infinity).
   Notice that $C(X)$ is not a TVS with this norm when X is not compact, as the sup-norm of an unbounded function is undefined.

Definition 2. *Given a TVS, V. Define the continuous dual space V' made up of continuous linear functionals, $I : V \to \mathbb{K}$. And the dual space V^* made up of all linear functionals (not necessarily continuous).*

Notice that $V' \subset V^*$ and that V' is the subset of V^* containing all the bounded linear functionals.
The dual space, can be endowed with various topologies. We will be focusing on is the weak* topology.

Definition 3. *The weak* topology is defined as the weakest topology such that for every $v \in V$:*

$$F_v : V^* \to \mathbb{K}, \quad F_v(I) = I(v)$$

is a continuous map.

We can also explicitly define a basis for this topology. Given open sets $U_1, .., U_k \subset \mathbb{K}$ and $v_1, ..., v_k \in V$:

$$U_{U_1,...,U_k}^{v_1,...,v_k} = \{I \in V^* : I(v_i) \in U_i\}$$

Furthermore, it's easy to see that convergence in this topology is pointwise. Hence, the weak* convergence is defined by:

$$\forall v \in V, I_n(v) \to I(v) \iff I_n \underset{w*}{\longrightarrow} I$$

One important result from functional analysis about the weak* topology is the Banach-Alaoglu theorem.

Theorem 1. *For any TVS V with continuous dual V' or dual V^*. The unit ball:*

$$B = \{I \in V' | \|I\|_{op} \leq 1\} = \{I \in V^* | \|I\|_{op} \leq 1\}$$

is compact in the weak topology.*

2 The Riesz Representation Theorem and the space of measures

The Riesz representation theorem is a very important result in measure theory which ties the set of radon, regular, borel measures to the set of positive linear operators on $C_c(X)$.

Definition 4. *A linear operator $I \in V^*$, when V is over $\mathbb{R}$, is considered positive if for all $v \in V$ such that $v \geq 0$, $I(v) \geq 0$. The order relation $\geq$ must be a partial order such that:*

1. $v \leq u \Rightarrow v + z \leq u + z$
2. $v \leq u \Rightarrow rv \leq ru$ *when* $r \geq 0$.

Usually we think of the order defined by the "coordinates" of v, such as, for $f \in C(X)$, $f_1 \leq f_2$ if $f_1(x) \leq f_2(x)$.

Theorem 2. *Let X be a locally compact Hausdorff space, and let I be a positive linear functional on $C_c(X)$. Then there exists a unique regular, radon, Borel measure μ such that for any $f \in C_c(X)$:*

$$I(f) = \int f d\mu$$

Furthermore, this map $C_c^*(X) \to \mathcal{M}(X)$ is a bijection. It's inverse take a measure μ to the linear operator $I(f) = \int f d\mu$.

We can also make the observation that μ is a finite measure if and only if the linear operator $\int f d\mu$ is bounded. Hence, we also get a bijection between the space of finite regular measures and positive continuous linear functionals. The following useful corollaries arise:

Corollary 1. *If two regular, radon, borel measures μ, ν on a locally compact hausdorff space X satisfy $\mu(f) = \nu(f)$ for all $f \in C_c(X)$, then $\mu = \nu$.*

Furthermore, if we combine this with the fact that all measures that are finite on compact sets on a locally compact hausdorff space that has a countable base are regular. We get that this result applies to all reasonable measures in such spaces.

Corollary 2. *If two measures μ, ν that are finite on compact sets on a locally compact Hausdorff second-countable space X satisfy $\mu(f) = \nu(f)$ for all $f \in C_c(X)$, then $\mu = \nu$.*

Since Riesz's theorem gives us a bijection between the space of regular, radon, borel measures and the space of positive linear functionals, we can borrow the weak* topology from $C_c^*(X)$. And define it on the space of regular, radon, borel measures $\mathcal{M}(X)$.

Definition 5. *The weak* topology is defined as the weakest topology such that for every $f \in C_c(X)$:*

$$F_f : \mathcal{M}(X) \to \mathbb{R}, \quad F_f(\mu) = \mu(f)$$

is a continuous map.

Obviously, the bijection given by Riesz Theorem is a homeomorphism. Hence, the properties of weak* topology from before, apply:

1. convergence in the weak* topology is equivalent to $\mu_n(f) \to \mu(f)$ for all $f \in C_c(X)$.

2. The following is a basis for the weak* topology $U_1, .., U_k \subset \mathbb{K}$ and $f_1, ..., f_k \in C_c(X)$:
$$U^{f_1,...,f_k}_{U_1,...,U_k} = \{\mu \in \mathcal{M} | \mu(f_i) \in U_i\}$$

3 Metrization of measure spaces

Another important fact about the weak* topology on the space $\mathcal{M}(X)$ is that it is metrizable when X is locally compact and second countable. Recall Urysohn's metrization theorem.

Theorem 3. *(Urysohn) Every hausdorff second-countable regular space is metrizable.*

It is an easy exercise to show that $\mathcal{M}(X)$ is regular and hausdorff. To construct a countable base, we use the fact that X is σ-compact (locally compact, second countable spaces are σ-compact) and hence $C_c(X)$ is separable as it is essentially the union of $C(K_n)$ with K_n being compact sets.
Now, denoting the countable base U_n for X and the countable dense set f_n in $C_c(X)$ we construct the following countable sub-basis:

$$U_{U_i}^{f_j} = \{\mu \in \mathcal{M}(X) | \mu(f_i) \in U_i\}$$

4 Compactness of the space of probability measures

We will begin with the simpler case where X is compact. In this case we can show that the space of probability measures is compact as well.

Theorem 4. *Let X be compact and hausdorff. Then the space $\mathcal{P}(X)$ is compact.*

The proof stems from the fact that we can identify $\mathcal{P}(X)$ with the set $\{I \in C_c^*(X) | \|I\|_{op} \leq 1\} \cap \{I \in C_c^*(X) | I \geq 0\} \cap \{I \in C_c^*(X) | I(\mathbb{1}_X) = 1\}$. The first set is is compact due to Banach-Alaoglu and the second and third sets are closed. On the other hand, when taking a space that is not compact this is not true, since $\mathbb{1}_X \notin C_c(X)$.

Example 1. *$X = \mathbb{R}$, let μ denote the lebesgue measure. We can define the probability measures $\nu_n(A) = \mu(A \cap (n, n+1))$. It's obvious that for any compactly supported function f, $\nu_n(f) \to 0$ (since it's support is bounded). Hence $\mu_n \underset{w*}{\longrightarrow} 0 \notin \mathcal{P}(\mathbb{R})$, meaning $\mathcal{P}(\mathbb{R})$ is not closed.*

5 Measure preserving transformations and ergodicity

Definition 6. *A measurable map $\varphi : (X, \mu) \to (Y, \nu)$ is called measure preserving if it satisfies: $\mu(\varphi^{-1}B) = \nu(B)$ for all measurable $B \subset Y$. If a measurable map $T : (X, \mu) \to (X, \mu)$ is measure preserving, then the measure μ is said to be T-invariant, (μ, X, T) a measure preserving system, and T a measure preserving transformation.*

Lemma 1. *A measure μ on X is T-invariant if and only if:*

$$\int f d\mu = \int f \circ T d\mu \tag{5.1}$$

For all $f \in \mathcal{L}^\infty$. Moreover, if μ is T-invariant then (5.1) holds for $f \in \mathcal{L}^1_\mu$. When X is locally compact and hausdorff, it is enough to for (5.1) to hold only for $f \in C_c(X)$ for μ to be T-invariant.

Proof. If (5.1) holds. For any measurable set B we may take $f = \mathbb{1}_B$. And see that

$$\mu(B) = \int \mathbb{1}_B d\mu = \int \mathbb{1}_B \circ T d\mu = \int \mathbb{1}_{T^{-1}B} d\mu = \mu(T^{-1}B)$$

Now, assuming that μ is T-invariant, then (5.1) hold for any indicator and hence for any simple function. Given a non-negative real valued function $f \in \mathcal{L}^1_\mu$ there is a rising sequence of simple functions $f_n \to f$. And this means that $f_n \circ T$ is a rising sequence of simple functions $f_n \circ T \to f \circ T$. Hence:

$$\int f \circ T d\mu = \lim_{n\to\infty} \int f_n \circ T d\mu = \lim_{n\to\infty} \int f_n d\mu = \int f d\mu$$

When X second countable, locally compact and hausdorf and μ a finite (or finite on all compact subsets of X) by Riesz Theorem and Corrolary 2, we get that if we define $\mu_T(A) = \mu(T^{-1}A)$:

$$\mu(f) = \int f d\mu = \int f \circ T d\mu = \mu_T(f)$$

Hence the measures μ and μ_T agree on all compactly supported continuous functions which implies that they are the same measure. □

Definition 7. *A measure preserving transformation T of a probability space $(X, \mathcal{B}, \mu)$ is ergodic if for any $B \in \mathcal{B}$:*

$$T^{-1}B = B \Rightarrow \mu(B) = 0 \text{ or } \mu(B) = 1$$

Proposition 1. *The following are equivalent properties for a measure preserving transformation T on $(X, \mathcal{B}, \mu)$:*

1. *T is ergodic.*
2. *For any $B \in \mathcal{B}$, $\mu(T^{-1}B \triangle B) = 0$ implies that $\mu(B) = 0$ or $\mu(B) = 1$.*
3. *For $A \in \mathcal{B}$, $\mu(A) > 0$ implies that $\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$.*
4. *For $A, B \in \mathcal{B}$, $\mu(A)\mu(B) > 0$ implies that there exists $n \geq 1$ with*

   $$\mu((T^{-n}A \cap B) > 0$$

   .
5. *For $f : X \to Y$ measurable with Y being a locally compact and second countable hausdorff space, $f \circ T \underset{a.e}{=} f$ implies that f is equal to a constant almost everywhere.*

This proposition gives us some more intuition as to why ergodic transformations have a "mixing" effect. For example, (3) shows that an ergodic transformation eventually fills the entire space if applied enough times on a set of any positive size.

6 Birkhoff's Pointwise Ergodic Theorem

Theorem 5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. If $f \in \mathcal{L}^1_\mu$ then*

$$A_n^f(x) = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$$

converges almost everywhere and in $\mathcal{L}^1_\mu$ to a T-invariant function $f^* \in \mathcal{L}^1_\mu$, and

$$\int f^* d\mu = \int f d\mu$$

If T is ergodic, then

$$f^*(x) = \int f d\mu$$

almost everywhere.

The proof assumes two theorems, the mean ergodic theorem and the maximal ergodic theorem. The mean ergodic theorem looks at the action of a measure preserving T as a unitary operator $U_T(f) = f \circ T$ on the hilbert space $\mathcal{L}^2_\mu$.

Theorem 6. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system, and let P_T denote the orthogonal projection onto the closed subspace*

$$I = \{g \in \mathcal{L}^2_\mu | U_T g = g\}$$

Then for any $f \in \mathcal{L}^2_\mu$,

$$A_N^f = x\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f \underset{\mathcal{L}^2_\mu}{\to} P_T f$$

Proof. The idea of the proof is:

1. Showing that $B = \{U_T g - g | g \in L^2_\mu\}$, and that $B^\perp = I$, hence $\mathcal{L}^2_\mu = \overline{B} \oplus I$ and $f = P_T f + h$ for some $h \in \overline{B}$.
2. Showing that for $h \in \overline{B}$, $\frac{1}{N}\sum\limits_{n=0}^{N-1} U_T^n h \underset{\mathcal{L}^2_\mu}{\to} 0$

□

Corollary 3. *In a measure preserving system like before, and $f \in \mathcal{L}^1_\mu$ the averages A_N^f converge in $\mathcal{L}^1_\mu$ to a T-invariant function $f' \in \mathcal{L}^1_\mu$.*

Theorem 7. *Given a measure preserving system like before and a real valued function $g \in \mathcal{L}^1_\mu$. Define:*

$$E_\alpha = \{x \in X | \sup_{n \geq 1} \frac{1}{n}\sum_{i=0}^{n-1} g(T^i x) > \alpha\}$$

fro any $\alpha \in \mathbb{R}$. Then:

$$\alpha\mu(E_\alpha) \leq \int_{E_\alpha} g d\mu \leq \|g\|_1.$$

Moreover, $\alpha\mu(E_\alpha \cap A) \leq \int_{E_\alpha \cap A} g d\mu$ whenever $A = T^{-1}A$.

The proof of this theorem is a bit technical, so we will not cover it.

Now, we can begin the proof of the pointwise ergodic theorem.

Proof. It is enough to prove the result for real valued functions $f \in \mathcal{L}^1_\mu$. Define for any $x \in X$,

$$f^*(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \limsup_{n\to\infty} A_n^f(x)$$

$$f_*(x) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \liminf_{n\to\infty} A_n^f(x)$$

Notice that:

$$\frac{n+1}{n} A_{n+1}^f(x) = \frac{n+1}{n} \left(\frac{1}{n} \sum_{i=0}^{n} f(T^i x) \right) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(Tx)) + \frac{1}{n} f(x) = A_n^f(Tx) + \frac{1}{n} f(x)$$

Hence, for any sequence n_k for which $A_{n+1}^f(x)$ converges to the limsup, the sequence $n_k - 1$ for $A_n^f \circ T(x)$ converges to the same value. Hence, $f^* \leq f^* \circ T$. On the other had, if n_k is a sequence for which $A_n^f(Tx)$ converges to the limsup. We get that $A_{n_k+1}^f(x)$ converges to the same value as well. Finally we get that $f^* = f^* \circ T$. A similar argument shows that $f_* = f_* \circ T$.
Let $\alpha, \beta > 0$ be rational numbers. Define

$$E_\alpha^\beta = \{x \in X | f_*(x) < \beta \text{ and } f^*(x) > \alpha\}$$

Since f^*, f_* are T invariant, $T^{-1}E_\alpha^\beta = E_\alpha^\beta$ and notice that E_α (with g = f) as defined in the maximal ergodic theorem contains E_α^β (as if the limsup is greater than α then the sup is also greater than α). Hence by the maximal ergodic theorem:

$$\int_{E_\alpha^\beta} f d\mu \geq \alpha\mu(E_\alpha^\beta)$$

Similarly, by considering $-f$ instead of f we get that:

$$\int_{E_\alpha^\beta} f d\mu \leq \beta\mu(E_\alpha^\beta)$$

Putting these two inequalities together:

$$\beta\mu(E_\alpha^\beta) \geq \int_{E_\alpha^\beta} f d\mu \geq \alpha\mu(E_\alpha^\beta)$$

For $\alpha > \beta$ this means that $\mu(E_\alpha^\beta) = 0$. Since,

$$\{x | f_*(x) < f^*(x)\} = \bigcup_{\alpha,\beta\in\mathbb{Q},\alpha>\beta} E_\alpha^\beta$$

We get that $\mu(\{x|f_*(x) < f^*(x)\}) = 0$. And thus $f_* \underset{a.e}{=} f^*$. And:

$$A_n^f(x) \underset{a.e}{\to} f^*(x)$$

By Corollary 3, we know that there is some $g \in \mathcal{L}_\mu^1$ such that $A_n^f \underset{\mathcal{L}_\mu^1}{\to} g$. This means that there must be some sub-sequence n_k such that $A_{n_k}^f \underset{a.e}{\to} g$. Hence, $f^* \underset{a.e}{=} g$ and $f^* \in \mathcal{L}_\mu^1$ and finally:

$$\int f d\mu = \int A_n^f d\mu = \int f^* d\mu$$

If T is ergodic, Since f^* is a T-invariant function, we get that f^* is equal to a constant almost everywhere. Hence, in this case:

$$f^*(x) \underset{a.e}{=} \int f d\mu$$

□

7 Generic points

Definition 8. *A sequence of elements $x_n \in X$ is equidistributed with respect to a probability measure μ if for any $f \in C_c(X)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(x_j) = \int f d\mu$$

Equivalently, (x_n) is equidistributed if

$$\frac{1}{n} \sum_{j=1}^{n} \delta_{x_j} \underset{w*}{\to} \mu$$

Definition 9. *A point $x \in X$ is generic in a measure preserving system, if the sequence $x_n = T^n x$ is equidistributed with respect to μ.*

Notice that if x is generic with respect to a regular, radon, borel measure μ. Then μ is defined uniquely by the averages (according to Riesz's theorem). Hence, x cannot be generic for two different measures.
Furthermore, the limit of the averages along the orbit of x in a dynamical system is T invariant, hence if x is generic with respect to μ, then μ is also T invariant.

Corollary 4. *Let $(X, \mathcal{B}, \mu, T)$ be a locally compact, second countable, housdorf space equipped with an ergodic, regular, radon measure, with continuous T. Then, μ almost every point is generic with respect to T, μ.*

Proof. Since $C_c(X)$ is a seperable metric space with respect to the $\|\cdot\|_\infty$ norm. We can take a dense sequence $f_n \subset C_c(X)$. Using brikhoff's pointwise ergodic theorem we get a set X' of full measure such that all $x \in X'$ satisfy for all i:

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

Given an $f \in C_c(X)$, for any $\epsilon > 0$ there is an i such that $\|f - f_i\|_\infty < \varepsilon$ hence $\|A_N^f - A_N^{f_i}\|_\infty < \varepsilon$ and $|\int f d\mu - \int f_i d\mu| < \varepsilon$ from which we conclude that for all $x \in X'$:

$$\int f d\mu - 2\varepsilon \leq \liminf_{N\to\infty} A_N^f(x) \leq \limsup_{N\to\infty} A_N^f(x) \leq \int f d\mu + 2\varepsilon$$

This, concludes the proof that for all $x \in X'$, $\lim_{N\to\infty} A_N^f(x) \to \int f d\mu$. □