Lectures on ergodic theorems

Arijit Ganguly
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Email: aganguly@iitk.ac.in

November 23, 2023

Abstract

We aim to prove the classical ergodic theorems for measure preserving transformations.

1 Introduction

Ergodic theory has its origins in statistical mechanics. While new applications to mathematical physics and other disciplines of science and engineering continued to come in the course of time, the theory soon earned in its own rights as an important place in analysis and probability. Ergodic theory studies *statistical properties* of (deterministic) *dynamical systems*.

A *dynamical system*, at its simplest form, is a map $T : X \longrightarrow X$. We call X the *phase space* or *state space* (as we regard the points $x \in X$ representing the possible states of the system), and T the *law of motion* (as it determines how the system evolves with time: here time is discrete, and from state x it transitions to state Tx in one unit of time). The trajectory $\{T^n x\}_{n\geq 0}$ is the record of the time evolution of the system. Our interest is in the study of the behaviour of the system in the far future, i.e., the behavior of $T^n x$ as $x \to \infty$.

More generally, we consider a semi-group (with identity) G acting on a set X, i.e., a map $G\times X \longrightarrow X$ with the following properties:

(i) $1x = x$, $\forall x \in X$, and

(ii) $(g_1 g_2)x = g_1(g_2 x)$, $\forall g_1, g_2 \in G, x \in X$.

For $g \in G$, we denote the map $X \longrightarrow X, x \mapsto gx$ by T_g. Thus we obtain a semi-group of transformations $\{T_g\}_{g \in G}$ of X. In the particular cases $G = \mathbb{R}_{\geq 0}$ or $\mathbb{R}$, we speak of *semi-flow* or *flow* respectively.

Ergodic theory, like probability theory, is based on general notions of measure theory. We shall throughout assume that X is endowed with a sigma-algebra $\mathscr{B}$ and a probability measure μ on $\mathscr{B}$, the transformation T is measurable and *preserves* μ, i.e., $\mu(T^{-1}(E)) = \mu(E), \forall E \in \mathscr{B}$. In general, an action of a semi-group G on a probability space $(X, \mathscr{B}, \mu)$ is said to be *probability preserving* if each T_g preserves μ.

Probability preserving semi-flows and flows arise naturally in many instances, e.g., the study of conservative Hamiltonian dynamical systems. Their investigation can be facilitated by a further reduction to single probability preserving transformations, on at least two counts:

1. If the flow possesses a global cross-section, then one obtains a probability preserving transformation of the cross-section associated with that flow. Its study will provide considerable information about the total flow.

2. Many a time, for small $\varepsilon > 0$, the 'discrete' version $\{T_{n\varepsilon} = T_\varepsilon^n\}_{n=0,1,\ldots}$ (or $\{T_{n\varepsilon}\}_{n \in \mathbb{Z}}$) happens to be a 'good approximation' of the original semi-flow $\{T_t\}_{t \geq 0}$ (or the flow $\{T_t\}_{t \in \mathbb{R}}$ respectively). Thus it is worthwhile to study the single transformation T_ε and also its iterates.

By *statistical properties* of a dynamical system, we mean properties that are expressed through the behaviour of time averages of (measurable) functions along trajectories of the system, i.e., by $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)$, or for semi-flows or flows, by $\frac{1}{T}\int_0^T f(T_t(x))\,dt$ or $\frac{1}{2T}\int_{-T}^T f(T_t(x))\,dt$ respectively, where f is an integrable function. Roughly speaking, we ask the following:

When do averages of quantities generated in a stationary manner converge?

2 Probability preserving transformations (p.p.t.)

2.1 Induced isometry

Let $(X, \mathscr{B}, \mu)$ be a probability space and $T : X \longrightarrow X$. One can easily see the proposition below.

Proposition 2.1. *The following are equivalent:*

(i) T *preserves* μ.

(ii) $\int_X f \circ T \, d\mu = \int_X f \, d\mu$, *for every positive measurable function.*

(iii) $\int_X f \circ T \, d\mu = \int_X f \, d\mu$, *for every* $f \in L^p(X, \mathscr{B}, \mu), 1 \leq p \leq \infty$.

Assume X is a compact topological space and $\mathscr{B}$ is the Borel sigma-algebra. Then one can also add the following to the above list:

(iv) $\int_X f \circ T \, d\mu = \int_X f \, d\mu$*, for every* $f \in C(X)$.

For any measurable $f : X \longrightarrow \mathbb{C}$, denote $f \circ T$ by U_T. The following is an immediate consequence of Proposition 2.1.

Theorem 2.2. *Let T preserve the probability measure μ. Then, for any $1 \leq p \leq \infty$, U_T defines an isometry of the Banach space $L^p(X, \mathscr{B}, \mu)$. In fact when $p = 2$, it is an isometry of the Hilbert space $L^2(X, \mathscr{B}, \mu)$, i.e., $\forall f, g \in L^2(X, \mathscr{B}, \mu)$, one has $\langle U_T f, U_T g \rangle = \langle f, g \rangle$. Moreover, if T is invertible then so is U_T; and for $p = 2$, it is unitary.*

The isometry U_T of $L^2(X, \mathscr{B}, \mu)$ mentioned above in Theorem 2.2 is sometimes referred to as the *Koopman operator* associated to T. The study of U_T is called the spectral study of T, and it is very useful in formulating concepts like ergodicity, mixing etc. and also in dealing with the isomorphism problems for probability preserving dynamical systems.

2.2 Koopman representation

Assume that X is a Borel subset of a complete separable metric space and $\mathscr{B}$ is the Borel sigma-algebra of X. Let G be a locally compact second countable group having a *measurable* action on X, i.e., the action map $G \times X \longrightarrow X$ is measurable. Suppose that μ is a Borel probability measure on X which is *preserved* by the action of G, i.e., $\mu(gE) = \mu(E)$, for all $g \in G$ and $E \in \mathscr{B}$. For any $g \in G$ and $f \in L^2(X, \mathscr{B}, \mu)$, define the following:

$$U_g(f)(x) = f(g^{-1}x), \forall x \in X.$$

Since the action is measure preserving, it is clear that $U_g(f) \in L^2(X, \mathscr{B}, \mu)$. One can easily verify U_g is a unitary operator on the Hilbert space $L^2(X, \mathscr{B}, \mu)$ and furthermore the following map

$$G \longrightarrow \mathscr{U}(L^2(X, \mathscr{B}, \mu)), g \mapsto U_g, \forall g \in G, \tag{2.1}$$

is a homomorphism. Since $L^2(X, \mathscr{B}, \mu)$ is separable, it follows that $\mathscr{U}(L^2(X, \mathscr{B}, \mu))$ is a separable metrizable group when endowed with the strong operator topology. It can be shown, using Fubini's theorem, that the homomorphism defined in (2.1) is measurable, and hence it is continuous due to a theorem by Mackey (see [16, Theorem B.3]). Thus (2.1) becomes a representation of G, which is called the *Koopman representation* associated to the probability preserving action.

2.3 Examples

Example 2.1. Let G be a compact topological group and μ be the Haar measure on G normalized so that $\mu(G) = 1$. For any $g \in G$, the map $T_g(x) \stackrel{\text{def}}{=} gx, \forall x \in G$, preserves μ. Suppose that $\{g_t\}_{t \in \mathbb{R}}$ is an *one-parameter subgroup* of G. For any $t \in \mathbb{R}$, let $T_t(x) \stackrel{\text{def}}{=} g_t x, \forall x \in G$. Then $\{T_t\}_{t \in \mathbb{R}}$ is a flow on G that preserves μ.

Special cases: Circle rotations, or more generally, translations on n-tori $\mathbb{T}^n$ (written additively), *quasi-periodic* flows on $\mathbb{T}^n$ defined as follows: fix $\mathbf{w} \in \mathbb{T}^n$ and for any $t \in \mathbb{R}$, $T_t(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{x} + t\mathbf{w}, \forall \mathbf{x} \in \mathbb{T}^n$.

Example 2.2. Let G be as above in the Example 2.1. Any continuous surjective homomorphism $T : G \longrightarrow G$ preserves μ. To see this, define the measure $\nu(E) = \mu(T^{-1}(E))$, for every Borel $E \subseteq G$. It is easy to verify that ν is a Borel probability measure invariant under left translations, and hence must be equal to μ.

Special cases: The n-th power map on the unit circle for any $n \in \mathbb{Z} \setminus \{0\}$. More generally, consider $A \in M_n(\mathbb{Z})$ with nonzero determinant and define $\mathbb{T}^n \longrightarrow \mathbb{T}^n, \bar{\mathbf{x}} \mapsto \overline{A\mathbf{x}}, \forall \mathbf{x} \in \mathbb{R}^n$ (Arnold's CAT map for instance).

Remark 2.1. One can combine Examples 2.1 and 2.2 into the following: any surjective affine transformation of G preserves μ.

Example 2.3. Let G be a locally compact group and Γ is a *lattice*, i.e., Γ is a discrete subgroup of G and G/Γ carries a finite regular Borel left G-invariant measure.

If Γ is a lattice in a locally compact group G, then the natural action of any subgroup H of G by left translations preserves μ, where μ is the normalised regular left G-invariant Borel measure on G/Γ so that $\mu(G/\Gamma) = 1$.

Remark 2.2. If a locally compact group G has a lattice then it is *unimodular*, i.e., any left invariant Haar measure on G is also right invariant. For any lattice Γ in G, $\Gamma\backslash G$ (the set of all right cosests of Γ in G) carries a unique right G-invariant Borel probability measure with total mass equal to 1.

Example 2.4. Let $(X, \mathscr{B}, \mu)$ be a probability space. Consider the space $\Omega \stackrel{\text{def}}{=} X^{\otimes\mathbb{N}}$ of sequences in X with the sigma algebra $\mathscr{X} \stackrel{\text{def}}{=} \mathscr{B}^{\otimes\mathbb{N}}$ generated by the finite-dimensional cylinders and the product measure $\beta \stackrel{\text{def}}{=} \mu^{\otimes\mathbb{N}}$. Define the left-shift map T on Ω as follows:

$$T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots), \forall (x_1, x_2, \ldots) \in X^{\otimes\mathbb{N}}.$$

For arbitrary m positive integers $i_1 < i_2 < \cdots < i_m$ and $C_{i_1}, C_{i_2}, \ldots, C_{i_m} \subseteq X$, let

$$C_{i_1, i_2, \ldots, i_m} \stackrel{\text{def}}{=} \{\{x_n\}_{n=1}^{\infty} : x_{i_1} \in C_{i_1}, \ldots, x_{i_m} \in C_{i_m}\}.$$

It is obvious that $T^{-1}(C_{i_1, i_2, \ldots, i_m}) = C_{i_1+1, i_2+1, \ldots, i_m+1}$. From this it follows that T preserves β.

Similarly one can consider the set of bi-infinite sequences $\Omega \stackrel{\text{def}}{=} X^{\otimes\mathbb{Z}}$ endowed with $\mathscr{X} \stackrel{\text{def}}{=} \mathscr{B}^{\otimes\mathbb{Z}}$ and $\beta \stackrel{\text{def}}{=} \mu^{\otimes\mathbb{Z}}$. The left-shift map map can also be defined on Ω in the natural manner:

$$(\ldots, x_{-2}, x_{-1}, \hat{x}_0, x_1, x_2, \ldots) \mapsto (\ldots, x_{-1}, x_0, \hat{x}_1, x_2, x_3, \ldots), \forall (\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots) \in X^{\otimes\mathbb{N}},$$

and is easily seen to be preserving β.

Special cases: Symbolic spaces.

Remark 2.3. Each coordinate x_i in Example 2.4 is an X valued random variable on of $\mathscr{X}$. The invariance of μ under the shift map means that, for any $k \in \mathbb{N}$, the joint probability distribution of the variables $x_{i_1+k}, x_{i_2+k}, \ldots, x_{i_m+k}$ does not depend on k. Thus here we obtain a *stationary random process*.

Example 2.5. Assume that X is a compact metric space and $T : X \longrightarrow X$ is continuous. Let ν_0 be a Borel probability measure on X. For any $n \in \mathbb{N}$, we let $\nu_n \overset{\text{def}}{=} T_*^n \nu_0$. Define a sequence of probability measures as follows:

$$\mu_n = \frac{1}{n}(\nu_0 + \cdots + \nu_{n-1}), \forall n \in \mathbb{N}.$$

The set of Borel probability measures on X, denoted by $\mathscr{M}(X)$, is endowed with the *weak topology* (see §A). Indeed, this topology is metrizable and furthermore, $\mathscr{M}(X)$ is compact (Theorem A.2 (2)). A typical analytic argument shows that every weak limit of the sequence $\{\mu_n\}_{n=1}^{\infty}$ is preserved under T. Now the existence of a T-invariant Borel probability measure follows from the weak compactness of $\mathscr{M}(X)$.

Example 2.6. Let $T : [0, 1] \longrightarrow [0, 1]$ be the map defined as follows:

$$T(x) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{1}{x} - [\frac{1}{x}] & \text{otherwise.} \end{cases}$$

The above map is called the *Gauss map*. To see how this map is related to the continued fractions, define a map $a_1 : [0, 1] \longrightarrow [0, 1]$ as follows:

$$a_1(x) = \begin{cases} 0 & \text{if } x = 0 \\ [\frac{1}{x}] & \text{otherwise.} \end{cases}$$

It is now a routine verification that, given any $x \in [0, 1]$, $a_1(T^{n-1}(x))$ is the n-th element of the continued fraction expansion of x, where $n \in \mathbb{N}$, whenever it makes sense.

Now for any $n \in \mathbb{N}$, observe that $x \in (\frac{1}{n+1}, \frac{1}{n}] \Leftrightarrow T(x) = \frac{1}{x} - n$. It follows that, for any $a, b \in [0, 1]$ with $a < b$, $T^{-1}([a, b]) = \bigcup_{n=1}^{\infty}[\frac{1}{b+n}, \frac{1}{a+n}]$. From this, one can verify that the preimage of the interval $[\frac{1}{2}, 1]$ has Lebesgue measure $\log 4 - 2 < \frac{1}{2}$, which shows that the Lebesgue measure of $[0, 1]$ is not preserved under T.

We define the *Gauss measure* by $E \mapsto \dfrac{1}{\log 2}\displaystyle\int_E \frac{dx}{1 + x}$, for every Borel $E \subseteq [0, 1]$. An easy computation shows that this measure is invariant under the Gauss map.

Example 2.7. Let $U \subseteq \mathbb{R}^n$ be bounded and open and $\mathbf{f} : U \overset{C^\infty}{\longrightarrow} \mathbb{R}^n$. Assume that the following ODE has solution on $\mathbb{R}$ for any initial $p \in U$:

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}). \tag{2.2}$$

Let $\{\varphi_t\}_{t \in \mathbb{R}}$ be the flow of (2.2). By Liouville's theorem, the Lebesgue measure on U is invariant under the above flow if and only if $\nabla \cdot \mathbf{f} = 0$.

Special cases: Hamiltonian system.

Example 2.8. Consider the upper-half plane $\mathbb{H} \overset{\text{def}}{=} \{x + iy : x, y \in \mathbb{R}, y > 0\}$. The *tangent bundle* of $\mathbb{H}$, denoted by $T\mathbb{H}$, is $\mathbb{C} \times \mathbb{H}$, i.e., the disjoint union of $T_z\mathbb{H} \overset{\text{def}}{=} \{z\} \times \mathbb{C}, z \in \mathbb{H}$. We define the *hyperbolic Riemannian metric* as the collection of inner products $\langle \cdot, \cdot \rangle_z$ on $T_z\mathbb{H}$, for $z \in \mathbb{H}$, given by the following:

$$\langle v, w \rangle_z \overset{\text{def}}{=} \frac{\langle v, w \rangle}{(Im\, z)^2}, \forall v, w \in \mathbb{C}.$$

$PSL_2(\mathbb{R})$ acts on $\mathbb{H}$ by Möbius transformations. This action is transitive and the stabilizer of i is $PSO_2(\mathbb{R})$, hence one has an identification $\mathbb{H} \cong PSL_2(\mathbb{R})/PSO_2(\mathbb{R})$. The *hyperbolic area form* $\frac{1}{y^2}dxdy$ on $\mathbb{H}$ is preserved under the action of $PSL_2(\mathbb{R})$.

Now, for any $g \overset{\text{def}}{=} \pm \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in PSL_2(\mathbb{R})$, we consider the map $\mathbb{H} \longrightarrow \mathbb{H}, z \mapsto \frac{az+b}{cz+d}, \forall z \in \mathbb{H}$, which we also denote by g. This map is holomorphic and in fact, $\forall z \in \mathbb{H}, g'(z) = \frac{1}{(cz+d)^2}$. It is easy to verify that, for any $z \in \mathbb{H}$, the following map preserves the inner products:

$$T_z\mathbb{H} \longrightarrow T_{g(z)}\mathbb{H}, (z, v) \mapsto (g(z), g'(z)v),\ \forall (z, v) \in T_z\mathbb{H}.$$

From this, it follows that $(z, v) \mapsto (g(z), g'(z)v),\ \forall (z, v) \in T\mathbb{H}$, defines an action of $PSL_2(\mathbb{R})$ on the *unit tangent bundle* $T^1\mathbb{H} \overset{\text{def}}{=} \{(z, v) \in T\mathbb{H} : \|v\|_z = 1\}$ of $\mathbb{H}$. One can check that this action is simply transitive, and hence $T^1\mathbb{H} \cong PSL_2(\mathbb{R})$. The *hyperbolic volume form* $\lambda \overset{\text{def}}{=} \frac{1}{y^2}dxdyd\theta$ on $T^1\mathbb{H}$ is preserved under the action of $PSL_2(\mathbb{R})$.

Let Γ be a lattice in $PSL_2(\mathbb{R})$. As mentioned in Remark 2.2, the quotient Haar probability measure on $\Gamma\backslash PSL_2(\mathbb{R})$ is right invariant. The aboe mentioned identification $T^1\mathbb{H} \cong PSL_2(\mathbb{R})$ induces an identification $T^1(\Gamma\backslash\mathbb{H}) = \Gamma\backslash T^1\mathbb{H} \cong \Gamma\backslash PSL_2(\mathbb{R})$. This identification yields a right invariant Borel probability measure $\tilde{\lambda}$ on $T^1(\Gamma\backslash\mathbb{H})$, which is given by the following:

$$\int_{T^1\mathbb{H}} f\, d\lambda = \int_{\Gamma\backslash T^1\mathbb{H}} \sum_{\gamma\in\Gamma} f(\gamma x)\, d\tilde{\lambda}(x),\ \forall f \in C_c(T^1\mathbb{H}).$$

Define the *geodesic flow* on $\Gamma\backslash PSL_2(\mathbb{R})$ (or equivalently $\Gamma\backslash T^1\mathbb{H}$) as follows: for any $t \in \mathbb{R}$,

$$\Gamma\backslash PSL_2(\mathbb{R}) \longrightarrow \Gamma\backslash PSL_2(\mathbb{R}), \Gamma g \mapsto \Gamma g g_t,$$

where $g_t = \begin{pmatrix} e^{t/2} & o \\ 0 & e^{-t/2} \end{pmatrix}$. It follows at once that any right invariant Haar measure on $\Gamma\backslash PSL_2(\mathbb{R})$ in invariant under the flow. The same conclusion also applies to the *horocycle flows* on $\Gamma\backslash PSL_2(\mathbb{R})$ (or equivalently $\Gamma\backslash T^1\mathbb{H}$) defined below:

1. *Positive horocycle flow*: $\Gamma g \mapsto \Gamma g u(t)$, where $u(t) \overset{\text{def}}{=} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \forall t \in \mathbb{R}$.

2. *Negative horocycle flow*: $\Gamma g \mapsto \Gamma g h(t)$, where $h(t) \overset{\text{def}}{=} \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}, \forall t \in \mathbb{R}$.

3 Ergodic theorems

Recall that our primary objective is to study, loosely speaking, when the averages of quantities generated in a stationary manner converge. In the classical situation, the concept of stationarity is probabilistic, i.e., it is determined by a p.p.t. T and we consider averages that are taken along $f, f \circ T$, $f \circ T^2, \cdot$, where f integrable. 'Ergodic theorems' usually refer to the theorems dealing with the convergence of such averages and it's close variants. The mode of converge also varies; for example, it is norm convergence for *mean* ergodic theorems, pointwise almost surely for the *individual* ergodic theorem and both for the *subadditive* ergodic theorem.

First we prove *von Neuman's mean ergodic theorem* where the convergence holds in L^2 norm.

3.1 von Neuman's mean ergodic theorem

Definition 3.1. Let $(X, \mathscr{B}, \mu)$ be a probability space and $T : X \longrightarrow X$. If T preserves μ then, we say that $(X, \mathscr{B}, \mu, T)$ is a *probability preserving system*. A measurable function $f : X \longrightarrow \mathbb{C}$ is said to be *T-invariant* if $f \circ T = f$ a.s.

Theorem 3.1 (Mean ergodic theorem)**.** *Let $(X, \mathscr{B}, \mu, T)$ be a probability preserving system. Then for any $f \in L^2(X, \mathscr{B}, \mu)$, there exists a T-invariant $\tilde{f} \in L^2(X, \mathscr{B}, \mu)$ such that the following holds:*

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \xrightarrow[n\to\infty]{L^2} \tilde{f}. \tag{3.1}$$

Before we prove Theorem 3.1, let us observe that, for $i \geq 0$, $f \circ T^i$ is nothing but $U_T^i(f)$, where U_T is the Koopman unitary operator associated to T. This motivates us to consider averages of the following type in the setting of an arbitrary Hilbert space $\mathscr{H}$ and study the convergence behaviour with respect to the strong operator topology:

$$\frac{1}{n}(I + U + \cdots + U^{n-1}), \text{ where } U : \mathscr{H} \longrightarrow \mathscr{H} \text{ is an isometry .} \tag{3.2}$$

To understand the convergence behaviour mentioned just above in (3.2), let us first consider the special case $\mathscr{H} = \mathbb{C}^n$, where $n \in \mathbb{N}$. In that case, U will be a unitary matrix. Without any loss in generality, we may assume that $U = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$. Clearly, for $j = 1, \ldots, n$, we have $|\lambda_j| = 1$. One can easily see that $\frac{1}{n}(I + U + \cdots + U^{n-1}) \xrightarrow[n\to\infty]{} \text{diag}\,(a_1, \ldots, a_n)$, where, for any $j = 1, \ldots, n$,

$$a_j = \begin{cases} 1 & \text{if } \lambda_j \neq 1 \\ 0 & \text{otherwise} \end{cases} .$$

Thus we see that $P \overset{\text{def}}{=} \text{diag}\,(a_1, \ldots, a_n)$ is a projection operator onto the subspace $\ker(U - I)$ (relative to the decomposition $\mathbb{C}^n = ker(U - I) \oplus ker(U - I)^{\perp}$). This suggests that the projection operator onto the closed subspace $ker(U - I)$ should be the right candidate even in the general case. Indeed, we prove the folliowing more general theorem which implies Theorem 3.1 at once:

Theorem 3.2 (Abstract ergodic theorem)**.** *Assume that $\mathscr{H}$ is a Hilbert space and U is a linear operator on $\mathscr{H}$ with $\|U\| \leq 1$. Denote by $\mathscr{H}_0$ and P the closed subspace $ker(U - I)$ and P the orthogonal projection to $\mathscr{H}_0$ respectively. Then for every $x \in \mathscr{H}$, one has the following:*

$$\frac{1}{n}\sum_{i=0}^{n-1} U^i x \xrightarrow[n\to\infty]{} Px. \tag{3.3}$$

Proof. In view of the decomposition $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_0^{\perp}$, it is enough to show (3.3) for vectors in $\mathscr{H}_0$ and $\mathscr{H}_0^{\perp}$. Trivially (3.3) holds when $x \in \mathscr{H}_0$. To prove the same when for $\mathscr{H}_0^{\perp}$, we need to observe the following:

$$\mathscr{H}_0^{\perp} = \overline{\{x - Ux : x \in \mathscr{H}\}}. \tag{3.4}$$

Before we establish (3.4), we first see how that implies (3.3) if $x \in \mathscr{H}_0^{\perp}$. Let $w \in \mathscr{H}$. Then $\sum_{i=0}^{n-1} U^i(w - Uw) = (w + Uw + \cdots + U^{n-1}w) - (Uw + \cdots + U^n w) = w - U^n w$. Since $\|U\} \leq 1$, it follows that

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} U^i(w - Uw)\right\| = \left\|\frac{1}{n}(w - U^n w)\right\| \leq \frac{2\|w\|}{n} \xrightarrow[n\to\infty]{} 0 = P(w - Uw).$$

Thus (3.3) holds for vecors lying in the subspace $\{x - Ux : x \in \mathscr{H}\}$, and hence the same goes on for $\overline{\{x - Ux : x \in \mathscr{H}\}}$ by a routine extension argument. In view of (3.4), (3.3) is now immediate for all $x \in \mathscr{H}_0^{\perp}$.

To prove (3.4), it suffices to show that if $y \in \mathscr{H}$ is such that $\langle y, w - Uw\rangle = 0$, for all $w \in \mathscr{H}$, then $y \in \mathscr{H}_0$. Since, for all $w \in \mathscr{H}$,

$$\langle y, w - Uw\rangle = \langle y, w\rangle - \langle y, Uw\rangle = \langle y, w\rangle - \langle U^* y, w\rangle = \langle y - U^* y, w\rangle,$$

we obtain that $U^* y = y$. Finally we need to use the following lemma to complete the proof. □

Lemma 3.3. *Consider a Hilbert space $\mathscr{H}$ and a linear operator $U : \mathscr{H} \longrightarrow \mathscr{H}$ with $\|U\| \leq 1$. Let $x \in \mathscr{H}$. Then one has the following:*

$$Ux = x \Longleftrightarrow U^* x = x.$$

Proof. It is enough to show the ($\Longleftarrow$) part, since $\|U\| = \|U^*\|$ and $U^{**} = U$. Let $U^* x = x$. The invariance of x under U follows at once from

$$\begin{aligned} 0 \leq \|x - Ux\|^2 = \langle x - Ux, x - Ux\rangle &= \|x\|^2 + \|Ux\|^2 - \langle Ux, x\rangle - \langle x, Ux\rangle \\ &= \|x\|^2 + \|Ux\|^2 - \langle x, U^* x\rangle - \langle U^* x, x\rangle \\ &= \|x\|^2 + \|Ux\|^2 - \|x\|^2 - \|x\|^2 \\ &= \|Ux\|^2 - \|x\|^2 \leq 0. \end{aligned}$$

□

The following proposition shows that the closed subspace of $L^2(X,\mathscr{B},\mu)$ consisting of all T-invariant functions in is again an L^2 space with respect to a sub-sigma-algebra of $\mathscr{B}$.

Proposition 3.4. *The following hold for any probability preserving system $(X,\mathscr{B},\mu,T)$:*

1. $\mathscr{I} \overset{\text{def}}{=} \{A \in \mathscr{B} : T^{-1}(A) = A\}$ *is a sub-sigma-algebra of* $\mathscr{B}$.
2. *Let $f : X \longrightarrow \mathbb{C}$ be measurable. Then f is T-invariant if and only if f is $\mathscr{I}$ measurable. In particular, $f \in L^p(X,\mathscr{B},\mu)$ is T-invariant if and only if $f \in L^p(X,\mathscr{I},\mu)$, for all $1 \le p \le \infty$.*

Proof. The verification of both (1) and (2) are straightfoward, hence omitted. □

In view of the above proposition, we conclude from Theorem 3.1 that

$$\frac{1}{n}(I + U_T + \cdots + U_T^{n-1}) \xrightarrow[n\to\infty]{} E(\cdot|\mathscr{I}) \text{ in the strong operator topology.}$$

For the definition and basic properties of the *Conditional expectation*, see Appendix B.

3.2 Birkhoff's individual ergodic theorem

Theorem 3.5 (Birkhoff's individual ergodic theorem)**.** *Let $(X,\mathscr{B},\mu,T)$ and $\mathscr{I}$ be as above in §3.1. Then, for any $f \in L^1(X,\mathscr{B},\mu)$, one has*

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \xrightarrow[n\to\infty]{a.e.} E(f|\mathscr{I}). \tag{3.5}$$

Proof. We can assume, without any loss in generality, that f is real valued. For any $n \in \mathbb{N}$, consider

$$F_n(x) \overset{\text{def}}{=} \max\{f(x), f(x) + f(T(x)), \dots, f(x) + f(T(x)) + \cdots + f(T^{n-1}(x))\}, \forall x \in X.$$

Observe that, for any $x \in X$, we have

$$F_{n+1}(x) = \max\{f(x), f(x) + F_n(T(x))\} = \begin{cases} f(x) + F_n(T(x)) & \text{if } F_n(T(x)) \ge 0 \\ f(x) & \text{if } F_n(T(x)) < 0 \end{cases}.$$

So $\forall x \in X$, we obtain that

$$F_{n+1}(x) - F_n(T(x)) = f(x) - \min\{F_n(T(x)), 0\} \ge f(x). \tag{3.6}$$

Let A be the set of all $x \in X$ such that

$$\sup\{f(x), f(x) + f(T(x)), \dots, f(x) + f(T(x)) + \cdots + f(T^n(x)), \dots\} = \infty.$$

It is a routine verification that A is T-invariant, and $\forall x \in A$, $F_n(x), F_n(T(x)) > 0$, for all $n \gg 1$. From this and (3.6), it follows that, $\forall x \in A$, the sequence $\{F_{n+1}(x) - F_n(T(x))\}_{n=1}^{\infty}$is decreasing and it decreases to $f(x)$. The Monotone Convergence Theorem now yields that

$$0 \le \int_A (F_{n+1} - F_n)\, d\mu = \int_A (F_{n+1} - F_n \circ T)\, d\mu \xrightarrow[n\to\infty]{} \int_A f\, d\mu. \tag{3.7}$$

On the other hand, it is clear that, $\forall x \notin A$,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \le 0, \tag{3.8}$$

as $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \le \frac{F_n(x)}{n}$, for all $n \in \mathbb{N}$, and $\{F_n(x)\}_{n=1}^{\infty}$ is bounded.

Observation. If $E(f|\mathscr{I}) < 0\, a.e.$ then, as a consequence of (3.7), one obtains

$$0 \ge \int_A E(f|\mathscr{I})\, d\mu = \int_A f\, d\mu \ge 0.$$

Hence $\int_A E(f|\mathscr{I})\, d\mu = 0$, which forces $\mu(A) = 0$, thereby (3.8) holds for μ-a.e. $x \in X$.

We use this observation to deal with the general case. Pick $\varepsilon > 0$ and set $\varphi \overset{\text{def}}{=} f - E(f|\mathscr{I}) - \varepsilon$. By definition $E(\varphi|\mathscr{I}) = -\varepsilon < 0$. In view of the previous observation, one arrives at the following for μ-a.e. $x \in X$:

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) - E(f|\mathscr{I})(x) - \varepsilon \le 0,$$

i.e., $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \le E(f|\mathscr{I})(x) + \varepsilon$, for $\mu - a.e.\, x \in X$. Note that this full measure set under consideration depends upon both f and ε. Now replacing f by $-f$, we get $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \ge E(f|\mathscr{I})(x) - \varepsilon$, for $\mu - a.e.\, x \in X$. The proof of Theorem 3.5 follows at once from this.

□

Remark 3.1. 1. *Birkhoff's individual ergodic theorem for probability preserving semi-flows*: Let $(X, \mathscr{B}, \mu)$ be as above in §3.1. Suppose that $\{\varphi_t\}_{t\ge 0}$ is a *measurable semi-flow* on X, i.e., the map $[0,\infty)\times X \longrightarrow X, (t,x) \mapsto \varphi_t(x), \forall t \ge 0, x \in X$ is measurable. Assume further that μ is preserved under the semi-flow. In this continuous parameter context, the *time average* of an integrable function f on X is given by

$$\frac{1}{T}\int_0^T f(\varphi_t(x))\, dt, \text{ for } T > 0.$$

Note that, for all $T > 0$, the function $[0,T]\times X \longrightarrow X, (t,x) \mapsto f(\varphi_t(x)), \forall t \in [0,T], x \in X$, is integrable, and hence for μ-a.e. $x \in X$, the integral $\frac{1}{T}\int_0^T f(\varphi_t(x))\, dt$ is defined. Analogous to Theorem 3.5, which we usually regard as a 'discrete time' theorem, one can easily have the convergence of $\frac{1}{T}\int_0^T f(\varphi_t(x))\, dt$ as $T \longrightarrow \infty$, for μ-a.e. $x \in X$. Indeed it is an easy consequence of Theorem 3.5. An interested reader is referred to [9, §1.2, **2. Continuous time**] for more details.

2. *Multiplicative version of individual ergodioc theorem*: Let $f : X \longrightarrow \mathbb{C}$ be a measurable function, where $(X, \mathscr{B}, \mu, T)$ is a p.p.s. Assume that $\log |f| \in L^1(X, \mathscr{B}, \mu)$. As a rather trivial implication of Theorem 3.5, we obtain that,

$$|f(T^{n-1}(x)) \cdots f(T(x)) f(x)|^{\frac{1}{n}} \xrightarrow[n\to\infty]{a.e.} \tilde{f},$$

where $\tilde{f}$ is a T-invariant integrable function.

3.3 L^p ergodic theorem

Theorem 3.6. *Let $(X, \mathscr{B}, \mu, T)$ and $\mathscr{I}$ be as in §3.1. Asume that $1 \leq p < \infty$. Then for any $f \in L^p(X, \mathscr{B}, \mu)$, $E(f|\mathscr{I}) \in L^p(X, \mathscr{I}, \mu)$ and*

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \xrightarrow[n\to\infty]{L^p} E(f|\mathscr{I}). \tag{3.9}$$

Proof. It is enough to show that the sequence $\left\{\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right\}_{n=1}^{\infty}$ is Cauchy, since then from the completeness of $L^p(X, \mathscr{B}, \mu)$, we obtain an $\tilde{f} \in L^p(X, \mathscr{B}, \mu)$ such that

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \xrightarrow[n\to\infty]{L^p} \tilde{f}.$$

Consequently, $\tilde{f}$ will be the pointwise limit of a subsequence of $\left\{\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right\}_{n=1}^{\infty}$ a.e. It then follows from Theorem 3.5 that $\tilde{f} \overset{\text{a.e.}}{=} E(f|\mathscr{I})$. Note that this also shows that $E(f|\mathscr{I}) \in L^p(X, \mathscr{I}, \mu)$.

Let $\varepsilon > 0$. Choose $g \in L^\infty(X, \mathscr{B}, \mu)$ such that $\|f - g\|_p < \varepsilon$. Now observe that, $\forall n, k \in \mathbb{N}$,

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} f \circ T^i\right\|_p \leq& \left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i\right\|_p \\ &+ \left\|\frac{1}{n+k}\sum_{i=0}^{n+k-1} f \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i\right\|_p \\ &+ \left\|\frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i\right\|_p\end{aligned} \tag{3.10}$$

It is clear that, the first two summands appearing in the RHS of (3.10) are both $< \varepsilon$. So from (3.10), we have

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} f \circ T^i\right\|_p \leq \varepsilon + \varepsilon + \left\|\frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i\right\|_p, \tag{3.11}$$

for all $n, k \in \mathbb{N}$. In view of (3.11), the proof will be over as soon as we show $\left\{ \frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i \right\}_{n=1}^{\infty}$ is Cauchy in $L^p(X, \mathscr{B}, \mu)$.

From (2) and (4), Theorem B.1, it is clear that $E(g|\mathscr{I}) \in L^\infty(X, \mathscr{I}, \mu) \subseteq L^p(X, \mathscr{I}, \mu)$. We show

$$\frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i \xrightarrow[n\to\infty]{L^p} E(g|\mathscr{I}).$$

Theorem 3.5 provides us with

$$\frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i \xrightarrow[n\to\infty]{a.e.} E(g|\mathscr{I}).$$

From this, we obtain that

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i - E(g|\mathscr{I}) \right|^p \xrightarrow[n\to\infty]{a.e.} 0.$$

To complete the proof, we now apply the Dominated Convergence Theorem to the following sequence

$$\left\{ \left| \frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i - E(g|\mathscr{I}) \right|^p \right\}_{n=1}^{\infty}.$$

□

3.4 Kingman's subadditive ergodic theorem

We retain the setting of §(3.1).

Definition 3.2. A sequence $\{f_n\}_{n=1}^\infty$ in $L^1(X, \mathscr{B}, \mu)$ is said to be an *additive cocycle* on X if $\forall m, n \in \mathbb{N}$, $f_{m+n} = f_m + f_n \circ T^m$.

Example 3.1. For any $f \in L^1(X, \mathscr{B}, \mu)$, the following defined an additive cocycle:

$$f_n \stackrel{\text{def}}{=} \sum_{i=0}^{n} f \circ T^i, \ \forall n \in \mathbb{N}. \tag{3.12}$$

In fact, it is easily seen that any additive cocycle is of the form (3.12) for some $f \in L^1(X, \mathscr{B}, \mu)$.

Definition 3.3. A sequence $\{f_n\}_{n=1}^\infty$ in $L^1(X, \mathscr{B}, \mu)$ is said to be a *subadditive cocycle* if $\forall m, n \in \mathbb{N}$, $f_{m+n} \le f_m + f_n \circ T^m$.

Example 3.2. Every additive cocycle is automatically subadditive.

Example 3.3. Suppose that A is a Banach algebra and $f : X \longrightarrow A$ is a measurable map. Assume that $\log \|f\| \in L^1(X, \mathscr{B}, \mu)$. For $n \in \mathbb{N}$, let

$$f_n(x) \stackrel{\text{def}}{=} \log \|f(T^{n-1}(x)) \cdots f(T(x))f(x)\|, \forall x \in X.$$

Then $\{f_n\}_{n=1}^{\infty}$ is a subadditive cocycle.

Example 3.4. This is a special case of Example 3.3. Take $G \stackrel{\text{def}}{=} GL_d(\mathbb{C})$, where $d \in \mathbb{N}$, $\mathscr{X} \stackrel{\text{def}}{=}$ Borel sigma-algebra on G and ν is a Borel probability measure on G. Consider $X \stackrel{\text{def}}{=} G^{\otimes\mathbb{N}}$, $\mathscr{B} \stackrel{\text{def}}{=} \mathscr{X}^{\otimes\mathbb{N}}$ and $\mu \stackrel{\text{def}}{=} \nu^{\otimes\mathbb{N}}$. Recall from Example 2.4 that the left-shift map $T : X \longrightarrow X$ preserves μ. Define

$$f : X \longrightarrow G,\ f(g_1, g_2, \dots) = g_1,\ \forall (g_1, g_2, \dots) \in X.$$

This yields the following subadditive cocyle, provided $\displaystyle\int_G \log \|g\|\, d\nu(g) < \infty$:

$$\log \|g_n \cdots g_1\|,\ n \in \mathbb{N}. \tag{3.13}$$

Example 3.5. Let G be a topological group and $f : X \longrightarrow G$ be a measurable map. Define the following for all $n \in \mathbb{N}$:

$$\nu_n(x) = f(T^{n-1}(x)) \cdots f(T(x))f(x),\ \forall x \in X,$$

and

$$f_n(x) = \#\{\nu_i(x) : i = 1, \cdots, n\},\ \forall x \in X.$$

One can easily verify that $\{f_n\}_{n=1}^{\infty}$ is a subadditive cocyle.

The following generalization of Theorem 3.5 for subadditive cocycles is due to Kingman ([7]):

Theorem 3.7 (Kingman's subadditive ergodic theorem)**.** *Let $(X, \mathscr{B}, \mu, T)$ and $\mathscr{I}$ be as above in §3.1. Suppose that $\{f_n\}_{n=1}^{\infty}$ be a subadditive cocycle. Assume*

$$\inf_n \frac{1}{n} \int_X f_n\, d\mu > -\infty. \tag{3.14}$$

Then there exists $\tilde{f} \in L^1(X, \mathscr{I}, \mu)$ such that

$$\frac{f_n}{n} \xrightarrow[n\to\infty]{} \tilde{f} \tag{3.15}$$

pointwise a.e. and also in L^1. In fact, $\tilde{f} = \inf_n \frac{1}{n} E(f_n | \mathscr{I})$.

Remark 3.2. The L^1 convergence in (3.15) implies that, $\forall A \in \mathscr{I}$, $\displaystyle\frac{1}{n}\int_A f_n\, d\mu \xrightarrow[n\to]{} \int_A \tilde{f}\, d\mu$. Indeed, the following lemma shows that, $\forall A \in \mathscr{I}$, $\displaystyle\int_A \tilde{f}\, d\mu = \inf_{n\geq 1} \frac{1}{n} \int_A f_n\, d\mu$.

Lemma 3.8 (Fekete's Subadditive Lemma)**.** *Let* $\{a_n\}_{n=1}^{\infty}$ *be a* subadditive *sequence of real numbers, i.e.,* $a_{m+n} \leq a_m + a_n$, $\forall m, n \in \mathbb{N}$. *Then,*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n} \in \mathbb{R} \cup \{-\infty\}.$$

For a proof of the pointwise convergence part of Theorem 3.7, extending the method of Risez, see [6]. Note that, the proof of [6] gives a proof of Theorem 3.5 simultaneously. The proof of L^1 convergence is independent of the pointwise one. The interested reader is referred to [9, §1.5, **Theorem 5.3**]. These proofs have also been explained with all details in Lecture 17. 18 and 19 of the course MTH 669A taught by me in 2021-22-I.

We now see an important application of Theorem 3.7 to the asymptotic behavior of the norm of products of random matrices, established originally by Furstenberg and Kesten in [4] without the subadditive ergodic theorem. Consider the subadditive cocycle considered in Examle 3.4. Let

$$N : G \longrightarrow [0,\infty), N(g) \stackrel{\text{def}}{=} \max\{\|g\|, \|g^{-1}\|\}, \ \forall g \in G.$$

Assume that the probability measure ν satisfies the *finite first moment condition*, i.e.,

$$\int_G \log N(g)\, d\nu(g) < \infty. \tag{3.16}$$

Note that (3.16) is stronger than $\int_G \log\|g\|\, d\nu(g) < \infty$, and furthermore that yields (3.14), where $\forall n \in \mathbb{N}$,

$$f_n(g_1, g_2, \dots) \stackrel{\text{def}}{=} \log\|g_n \cdots g_1\|, \ \forall (g_1, g_2, \cdots) \in G^{\otimes\mathbb{N}}.$$

To see this, observe that, for all $n \in \mathbb{N}$,

$$\begin{aligned} I_d = (g_1^{-1} \cdots g_n^{-1})(g_n \cdots g_1) \Rightarrow 0 &\leq \sum_{i=1}^{n} \log\|g_i^{-1}\| + \log\|g_n \cdots g_1\| \\ &\Rightarrow 0 \leq \int_G \log N(g)\, d\nu(g) \\ &+ \frac{1}{n}\int_X \log\|g_n \cdots g_1\|\, d\mu(g_1, g_2, \dots) \\ &\Rightarrow -\int_G \log N(g)\, d\nu(g) \\ &\leq \inf_n \frac{1}{n}\int_X \log\|g_n \cdots g_1\|\, d\mu(g_1, g_2, \dots). \end{aligned}$$

Hence, from Theorem 3.7, it follows that,

$$\frac{1}{n}\log\|g_n \cdots g_1\| \text{ converges for } \mu - a.e.\, (g_1, g_2, \cdots) \in G^{\otimes\mathbb{N}}. \tag{3.17}$$

Remark 3.3. Note that, the finite first moment condition as in (3.16) is weaker than the 'actual' finite first moment condition on the random variable N, i.e., $\int_G N(g)\,d\nu(g) < \infty$, since

$$\int_G \log N(g)\,d\nu(g) \le \log\left(\int_G N(g)\,d\nu(g)\right),$$

by Jensen's inequality.

3.5 Oxtoby's ergodic theorem

Let X be acompact metric space and $X \xrightarrow[cts.]{T} X$. Denote by $\mathscr{M}^T(X)$ the set of all T-invariant Borel probability measures on X. Recall from Example 2.5 that $\mathscr{M}^T(X) \neq \emptyset$. Furthermore, $\mathscr{M}^T(X)$ is a closed and convex subset of the set of all complex Borel measures on X. To see this, let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence in $\mathscr{M}(X)$ converging to $\mu \in \mathscr{M}(X)$. Then by definition, for any $f \in C(X)$, $\int_X f d\mu_n \xrightarrow[n\to\infty]{} \int_X f\,d\mu$. This implies that, for any given $f \in C(X)$, $\mu \mapsto \int_X f\,d\mu$ is a continuous map from $\mathscr{M}(X)$ to $\mathbb{C}$. It follows from Proposition 2.1 (iv) that $\mathscr{M}^T(X) = \bigcap_{f\in C(X)} \left\{\mu \in \mathscr{M}(X) : \int_X (f - f\circ T)\,d\mu\right\}$ is closed, being the intersession of a family of closed sets. The verification of convexity is straightforward. Thus $\mathscr{M}^T(X)$, unless it is singleton, is an uncountable set. For many maps, it is difficult to describe the set $\mathscr{M}^T(X)$, while for some map, it can be described completely; for example, in the following:

Example 3.6 (North-South map)**.** Consider the circle C in $\mathbb{R}^2$ centered at $(0,1)$ and having radius 1. Denote $(0,2)$ and $(0,0)$ by N and S respectively. Let ϕ stand for the *stereographic projection* of this circle to $\mathbb{R}$. Consider the following map :

$$T : C \longrightarrow C,\ T(x) = \begin{cases} \phi^{-1}\left(\frac{\phi(x)}{2}\right) & \text{if } x \in X \setminus \{N\} \\ N & \text{otherwise} \end{cases}. \tag{3.18}$$

Obviously N and S are the fixed points of T so that δ_N and δ_S are T-invariant. Suppose that ν is a T-invariant Borel probability measure on C. Observe I is the open arc of C with end points x and $T(x)$, where $x \neq N$, then $T^{-n}(I) \cap T^{-m}(I) = \emptyset$, $\forall m \neq n \in \mathbb{N}$. From finiteness of $\nu(C)$, one obtains that $\nu(I) = 0$, and consequently, the open arc with end points N and x being the disjoint union of $\{T^{-n}(I) : n \in \mathbb{N}\}$ has ν measure 0. It follows that the ν measure of the entire right semi-circle is 0, and similarly one has the same for the entire left semi-circle as well. Hence ν must be supported on N and S. Thus, in this case, $\mathscr{M}^T(C) = \{\lambda\delta_N + (1-\lambda)\delta_S : \lambda \in [0,1]\}$.

We now focus on the case when $\mathscr{M}^T(X)$ is singleton, and witness a stronger behaviour of the time averages of continuous functions. First we prove the following proposition:

Proposition 3.9. *Assume that X be a compact metric space and $T : X \longrightarrow X$ is continuous. Consider $\mathscr{I} \stackrel{def}{=} \{f \in C(X) : f\circ T = f\}$ and $\mathscr{V} \stackrel{def}{=} \{f - f\circ T : f \in C(X)\}$. Then the following are equivalent:*

1. $\forall f \in C(X), \exists \tilde{f} \in \mathscr{I}$ *such that,* $\forall x \in X$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \xrightarrow[n\to\infty]{} \tilde{f}(x). \tag{3.19}$$

2. $C(X) = \mathscr{I} \oplus \overline{\mathscr{V}}$

3. *Same as* (1)*, but in addition to that, the convergence in* (3.19) *is uniform.*

Proof. (3) ⇒ (1) is trivial. (2) ⇒ (3), follows from the observations made below:

Observation.

- $\forall f \in \mathscr{I}$, $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i = f$, for all $n \in \mathbb{N}$. So the uniform convergence of $\left\{\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\}\right\}_{n=1}^{\infty}$ to f is obvious.

- Let f be of the form $g - g \circ T$, where $g \in C(X)$. In this case also, it is easy to see that

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \xrightarrow[n\to\infty]{} 0 \text{ uniformly.} \tag{3.20}$$

- An usual extension argument permits one to deduce (3.20) for any $f \in \overline{\mathscr{V}}$.

Thus we are now left with the proof of (1) ⇒ (2). Define $P : C(X) \longrightarrow C(X)$ by $Pf = \tilde{f}$, where $\tilde{f}$ is defined as above in (3.19). From the definition, we clearly have the following:

(a) P is a bounded linear operator with $\|P\| \leq 1$,

(b) $P\phi = \phi$, for all $\phi \in \mathscr{I}$, and

(c) $\operatorname{Im} P \subseteq \mathscr{I}$.

Note that (b) and (c) implies that P is a projection onto $\mathscr{I}$. Hence $C(X) = \mathscr{I} \oplus \ker P$, In view of this, it is enough to show $\overline{\mathscr{V}} = \ker P$. We already have $\mathscr{V} \subseteq \ker P$, which yields $\overline{\mathscr{V}} \subseteq \ker P$. The other containment can be established if we can show that for any bounded linear functional I with $I(\mathscr{V}) = \{0\}$ (or equivalently $I(\overline{\mathscr{V}}) = \{0\}$), one always have $I(\ker P) = \{0\}$.

Let $I \in C(X)^*$ be such that $I(\varphi - \varphi \circ T) = 0$, for all $\varphi \in C(X)$. For any $f \in \ker P$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \xrightarrow[n\to\infty]{} 0.$$

Applying Corollary C.3, we obtain that

$$\frac{1}{n}\sum_{i=0}^{n-1} I(f \circ T^i) \xrightarrow[n\to\infty]{} 0.$$

$I(f) = 0$ follows at once from this, since $I(f) = I(f \circ T^i)$, for any $i \geq 0$. □

In the rest of this section, we assume $\mathscr{M}^T(X) = \{\mu\}$.

Remark 3.4. If μ is the only T-invariant Borel probability measure on X, it is not difficult to see the following:

$$A \subseteq X \text{ and } T^{-1}A = A \Longrightarrow \mu(A) = 0 \text{ or } 1. \tag{3.21}$$

For, otherwise we would then have a T-invariant subset A of X with $0 < \mu(A) < 1$. For any $\lambda \in [0, 1]$, let

$$\mu_\lambda(E) \stackrel{\text{def}}{=} \lambda \frac{\mu(A \cap E)}{\mu(A)} + (1 - \lambda)\frac{\mu((X \setminus A) \cap E))}{\mu((X \setminus A))}, \ \forall \text{ Borel } E \subseteq X.$$

Since A is T-invariant, so is $X \setminus A$, from this it follows that the Borel probability measure μ_λ is T-invariant, for all $\lambda \in [0, 1]$. This contradicts our assumption.

We now claim that Pf is constant, for all $f \in C(X)$, or equivalently, $\mathscr{I} = \mathbb{C}$, where $\mathbb{C}$ is to be interpreted as the subspace of all constant functions on X. Suppose that there exist $f \in C(X)$ and $x, y \in X$ such that $Pf(x) \neq Pf(y)$. Define, for all $n \in \mathbb{N}$,

$$\nu_n = \frac{\delta_x + \delta_{T(x)} + \cdots + \delta_{T^{n-1}(x)}}{n}. \tag{3.22}$$

Let $\nu_{n_k} \xrightarrow[k\to\infty]{} \nu$, where $\{n_k\}_{k=1}^\infty$ is a subsequnce and $\nu \in \mathscr{M}(X)$. Clearly ν is T-invariant and

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

In the exactly similar way to that of (3.22), one defines a sequence $\{\eta_n\}_{n=1}^\infty$ in $\mathscr{M}(X)$ and obtain a T-invariant Borel probability measure η such that $\int_X f\,d\eta = Pf(y)$. It now follows that $\nu \neq \eta$, which is absurd.

The converse also holds, i.e., $\forall f \in C(X)$, Pf is constant $\Longrightarrow$ there is a unique T-invariant Borel probability measure. To see this, let $\lambda, \xi \in \mathscr{M}^T(X)$. From Proposition 3.9 (3), one immediately has Pf is equal to both of $\int_X f\,d\lambda$ and $\int_X f\,d\xi$, for all $f \in C(X)$. Now $\lambda = \xi$ is obvious in view of Riesz representation theorem.

We thus arrive at the following:

Theorem 3.10 (Oxtoby's erodic theorem)**.** *Assume that X is a compact metric space and $T : X \longrightarrow X$ is continuous. Then the following are equivalent:*

1. *There is a unique T-invariant Borel probability measure on X.*

2. $\forall f \in C(X), \exists c(f) \in \mathbb{C}$ *such that,* $\forall x \in X$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \xrightarrow[n\to\infty]{} c(f). \tag{3.23}$$

3. *Same as* (2)*, but in addition to that, the convergence in* (3.23) *is uniform.*

4. $\mathscr{I} = \mathbb{C}$.

5. $C(X) = \mathbb{C} \oplus \overline{\mathscr{V}}$

Indeed, if μ *is the unique* T*-invariant Borel probability measure on* X *then* $c(f) = \int_X f\, d\mu$ *for all* $f \in C(X)$.

As an imemdiate corollary of Theorem 3.10, we are provided with the following equivalent chatracterization of the unique T-invariant Borel probability measure in terms of equidistribution (see §D for definition and equivalent characterization) of trajectories of T:

Theorem 3.11. *Let* X *and* T *as in Theorem 3.10. Then, for a* T*-invariant Borel probability measure* μ *on* X*, the following are equivalent:*

1. $\mathscr{M}^T(X) = \{\mu\}$.

2. *For every* $x \in X$, $\{T^n(x)\}_{n=0}^{\infty}$ *equidistributes with respect to* μ.

The equivalent criteria given by Theorem 3.10 are often quite useful to prove that a given topological dynamical system admits a unique invariant Borel probability measure. We now illustrate that with an example. Before that, we note that, in order to establish $\mu \in \mathscr{M}^T(X)$ is the only such measure, it is enough to find a dense $\mathcal{S} \subseteq C(X)$ such that for all $f \in \mathcal{S}$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \xrightarrow[n\to\infty]{} \int_X f\, d\mu, \forall x \in X. \tag{3.24}$$

Example 3.7. Let $\alpha \in \mathbb{R}$ and $R_\alpha : \mathbb{T} \longrightarrow \mathbb{T}, R_\alpha(\bar{x}) \stackrel{\text{def}}{=} \overline{x+\alpha}$, $\forall x \in \mathbb{R}$. Clearly the Haar probability measure λ is invariant under R_α. Assume that α is irrational. For any $k \in \mathbb{Z}$, consider the function $f_k : \mathbb{T} \longrightarrow \mathbb{C}, f_k(\bar{x}) \stackrel{\text{def}}{=} e^{2\pi i k x}$. Then observe that, for any $\bar{x} \in \mathbb{T}$,

$$\frac{1}{n}\sum_{r=0}^{n-1} f_k(R_\alpha^r(\bar{x})) = \frac{1}{n}\sum_{r=0}^{n-1} e^{2\pi i k(x+r\alpha)} = \begin{cases} 1 & \text{if } k = 0 \\ e^{2\pi i k x}\frac{1}{n}\frac{e^{2\pi i n k\alpha}-1}{e^{2\pi i k\alpha}-1} & \text{if } k \neq 0. \end{cases} \tag{3.25}$$

On the other hand,

$$\int_{\mathbb{T}} f_k\, d\lambda = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0. \end{cases} \tag{3.26}$$

Combining (3.25) and (3.26), we obtain that $\frac{1}{n}\sum_{r=0}^{n-1} f_k(R_\alpha^r(\bar{x})) \xrightarrow[n\to\infty]{} \int_{\mathbb{T}} f_k\, d\lambda$, for all $k \in \mathbb{Z}$. From linearity, it is easy to see that (3.24) holds for any trigonometric polynomial f. Since trigonometric polynomials are dense in $C(\mathbb{T})$, one obtains that λ is the unique R_α-invariant Borel probability measure on $\mathbb{T}$, and consequently, the equidistribution of $\{\overline{n\alpha}\}_{n=0}^{\infty}$ with respect to λ.

Remark 3.5. If $\alpha = \frac{p}{q} \in \mathbb{Q}$, where $p, q \in \mathbb{Z}$ and $q > 0$, it is easy to see that, for any $\bar{x} \in \mathbb{T}$,

$$f_q(R_\alpha(\bar{x})) = e^{2\pi i q\left(x+\frac{p}{q}\right)} = e^{2\pi i q x} = f_q(\bar{x}).$$

So the nonconstant continuous function f_q is R_α-invariant. Therefore, from Theorem 3.10, we obtain more R_α-invariant Borel probability measures on $\mathbb{T}$.

One interesting application of the above equidistribution result is the following:

Example 3.8 (Gelfand's question and Benford's law)**.** Let $d \in \{0, 1, \dots, 9\}$. For any $n \in \mathbb{N} \cup \{0\}$, observe that 2^n has the leading digit d, in the decimal representation if and only if

$$10^k d \leq 2^n < 10^k(d+1), \text{ for some } k \in \mathbb{N} \cup \{0\}. \tag{3.27}$$

Now (3.27) is same as $\langle n \log_{10} 2 \rangle \in [\log_{10} d, \log_{10}(d+1))$. Since $\log_{10} 2$ is irrational, from the equidistribution mentioned in Example 3.7 and Theorem D.1, one concludes the following:

$$\frac{\#\{0 \leq k \leq n-1 : 2^k \text{ has leading digit } d\}}{n} \xrightarrow[n\to\infty]{} \log_{10}\left(1 + \frac{1}{d}\right).$$

Thus the d appears as the leading digit with the asymptotic frequency $\log_{10}\left(1 + \frac{1}{d}\right)$, and this shows that the most frequent leading digit in the sequence of powers of 2 is 1.

Appendices

A Weak topology in $\mathscr{M}(X)$

Let X be metric space and $\mathscr{M}(X)$ stand for the set of all Borel probability measures on X. For $\mu \in \mathscr{M}(X), f_1, \dots f_n \in C(X)$ and $\varepsilon_1, \dots, \varepsilon_n > 0$, define the following:

$$V_{\mu;f_1,\dots,f_n;\varepsilon_1,\dots,\varepsilon_n} = \left\{\nu \in \mathscr{M}(X) : \left|\int_X f_i \, d\nu - \int_X f_i \, d\mu\right| < \varepsilon_i, \text{ for } i = 1, \dots, n\right\}.$$

It is a routine verification that the family $\{V_{\mu;f_1,\dots,f_n;\varepsilon_1,\dots,\varepsilon_n}\}_{\mu;f_1,\dots,f_n;\varepsilon_1,\dots,\varepsilon_n}$ satisfies the conditions for a basis of a topology, which is called the *weak topology* on $\mathscr{M}(X)$. Unless otherwise mentioned, we always assume that $\mathscr{M}(X)$ is equipped with this topology. From the definition, it is clear that, a net $\{\mu_\alpha\}_{\alpha\in I}$ in $\mathscr{M}(X)$ converges to $\mu \in \mathscr{M}(X)$ if and only if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu, \ \forall f \in C(X).$$

In that case, it is customary to say μ_α *converges weakly* to μ. The following theorem provides some conditions that are equivalent to weak convergence:

Theorem A.1. *[12, Chapter II, Theorem 6.1] Consider a net* $\{\mu_\alpha\}_{\alpha\in I}$ *in* $\mathscr{M}(X)$ *and* $\mu \in \mathscr{M}(X)$*. Then the following are equivalent:*

1. $\mu_\alpha \to \mu$.
2. $\int_X f\, d\mu_\alpha \to \int_X f\, d\mu$*, for all bounded real valued uniformly continuous function* f *on* X*.*
3. $\limsup_\alpha \mu_\alpha(C) \leq \mu(C)$*, for every closed subset* $C \subseteq X$*.*
4. $\liminf_\alpha \mu_\alpha(U) \geq \mu(U)$*, for every open subset* $U \subseteq X$*.*
5. $\mu_\alpha(A) \to \mu(A)$*, for every Borel* $A \subseteq X$ *satisfying* $\mu(\partial A) = 0$*.*

Theorem A.2. *The following hold for* $\mathscr{M}(X)$*:*

1. *[12, Chapter II, Theorem 6.2]* $\mathscr{M}(X)$ *can be metrized as a separable metric space if and only if* X *is a separable metric space.*
2. *[12, Theorem 6.4]* $\mathscr{M}(X)$ *is a compact metric space if and only if* X *is a compact metric space.*

We refer the reader to [12, Chapter II, §6] for the proofs of Theorems A.1 and A.2.

Remark A.1. For X compact, $\mathscr{M}(X)$ can be identified with a closed subset of the unit ball of $C(X)^*$ by the Riesz representation theorem. Through this identification, the topology that it obtains from the weak* topology of $C(X)^*$ coincides with the existing weak topology.

Remark A.2. In case of Theorem A.2(2), one can define a metric on $\mathscr{M}(X)$ as follows. Choose a countable dense subset $\{f_n : n \in \mathbb{N}\}$ of $C(X)$. For $\mu, \nu \in \mathscr{M}(X)$, let

$$D(\mu, \nu) \overset{\text{def}}{=} \sum_{n=1}^{\infty} \frac{\left|\int_X f_n\, d\mu - \int_X f_n\, d\nu\right|}{2^n \|f_n\|}. \tag{A.1}$$

It is not difficult to verify that D, defines as above in (A.1), is a metric on $\mathscr{M}(X)$ which produces the weak topology.

B Conditional expectation

Assume that $(X, \mathscr{B}, \mu)$ is a probability space and $\mathscr{A} \subseteq \mathscr{B}$ is a sub-sigma-algebra. Then $L^2(X, \mathscr{A}, \mu)$ is a closed subspace of $L^2(X, \mathscr{B}, \mu)$. Denote by P the orthogonal projection onto $L^2(X, \mathscr{A}, \mu)$.

Note B.1. Let $A \in \mathscr{A}$. Then $\chi_A \in L^2(X, \mathscr{A}, \mu)$, so that one has the following for every $f \in L^2(X, \mathscr{B}, \mu)$:

$$\int_A f\, d\mu = \langle f, \chi_A \rangle = \langle Pf, \chi_A \rangle = \int_A Pf\, d\mu. \tag{B.1}$$

We now regard $L^2(X,\mathscr{B},\mu)$ (and $L^2(X,\mathscr{A},\mu)$) as a dense subspace of $L^1(X,\mathscr{B},\mu)$ (and $L^1(X,\mathscr{B},\mu)$ respectively). Let $f : X \longrightarrow \mathbb{R}$ be square integrable. Then the subsets $\{Pf \geq 0\}$ and $\{Pf < 0\}$ are obviously in $\mathscr{A}$. Now in view of Note B.1, one deduces that

$$\begin{aligned} \|Pf\|_1 = \int_X |Pf|\, d\mu &= \int_{\{Pf\geq 0\}} Pf\, d\mu - \int_{\{Pf<0\}} Pf\, d\mu \\ &= \int_{\{Pf\geq 0\}} f\, d\mu + \int_{\{Pf<0\}} (-f)\, d\mu \\ &\leq \|f\|_1. \end{aligned} \tag{B.2}$$

If $f \in L^2(X,\mathscr{B},\mu)$ is complex valued, then splitting it into real and imaginary parts and using (B.2), one obtains that $\|Pf\|_1 \leq 2\|f\|_1$. Hence, P admits a unique continuous extension to $L^1(X,\mathscr{B},\mu)$. For the time being, let us denote this extension also by P by abuse of notation.

Consider $A \in \mathscr{A}$ and $f \in L^1(X,\mathscr{B},\mu)$. Then for any $\varepsilon > 0$, one has $g \in L^2(X,\mathscr{B},\mu)$ such that $\|f-g\|_1 < \varepsilon$. This implies that

$$\left|\int_A Pf\, d\mu - \int_A Pg\, d\mu\right| \leq \int_X |Pf - Pg|\, d\mu = \|P(f-g)\|_1 \leq \|P\|\|f-g\|_1 \leq 2\varepsilon.$$

From this, it now follows that $\left|\int_A Pf\, d\mu - \int_A f\, d\mu\right| \leq 2\varepsilon + \varepsilon = 3\varepsilon$. Thus one obtains that the following holds for $P : L^1(X,\mathscr{B},\mu) \longrightarrow L^1(X,\mathscr{A},\mu)$:

$$\int_A Pf\, d\mu = \int_A f\, d\mu, \forall A \in \mathscr{A} \text{ and } f \in L^1(X,\mathscr{B},\mu). \tag{B.3}$$

Suppose now that $Q : L^1(X,\mathscr{B},\mu) \longrightarrow L^1(X,\mathscr{A},\mu)$ is a bounded linear map satisfying the following property:

$$\int_A Qf\, d\mu = \int_A f\, d\mu, \forall A \in \mathscr{A} \text{ and } f \in L^1(X,\mathscr{B},\mu). \tag{B.4}$$

It is easy to see from (B.4) that $Q\varphi \geq 0$ almost everywhere if $\varphi \geq 0$ is integrable, which in turn implies that, for any real valued integrable function ψ, $Q\psi(x) \in \mathbb{R}$, for μ a.e. $x \in X$. Assume now $f \in L^1(X,\mathscr{B},\mu)$ is real valued. Clearly $\{Pf < Qf\}$ and $\{Pf > Qf\} \in \mathscr{A}$. Since

$$\int_{\{Pf<Qf\}} (Pf - Qf)\, d\mu = \int_{\{Pf<Qf\}} Pf\, d\mu - \int_{\{Pf<Qf\}} Qf\, d\mu = \int_{\{Pf<Qf\}} f\, d\mu - \int_{\{Pf<Qf\}} f\, d\mu = 0,$$

one has $\{Pf < Qf\}$ is μ-null. A similar argument confirms that $\mu(\{Pf > Qf\}) = 0$ as well. Therefore, we have $Pf \overset{\text{a.e.}}{=} Qf$. The same can be immediately extended to any complex valued integrable function by taking real and imaginary parts. This leads us to the following conclusion:

There exists a unique bounded linear map $P : L^1(X,\mathscr{B},\mu) \longrightarrow L^1(X,\mathscr{A},\mu)$ satisfying (B.3).

Definition B.1. Given a probability space $(X,\mathscr{B},\mu)$ and $\mathscr{A} \subseteq \mathscr{B}$ a sub-sigma-algebra, the unique bounded linear map $P : L^1(X,\mathscr{B},\mu) \longrightarrow L^1(X,\mathscr{A},\mu)$ satisfying (B.3) is called the *conditional expectation given $\mathscr{A}$*, and denoted by $E(\cdot|\mathscr{A})$.

Example B.1. Let $(X, \mathscr{B}, \mu)$ be a probability space and $\mathscr{A} \stackrel{\text{def}}{=} \{E \in \mathscr{B} : \mu(E) = 0 \text{ or } 1\}$. Then $E(f|\mathscr{A}) = \int_X f\, d\mu$, for any $f \in L^1(X, \mathscr{B}, \mu)$.

Example B.2. Let $(X, \mathscr{B}, \mu)$ be a probability space and $\xi \stackrel{\text{def}}{=} \{A_1, \ldots, A_n\}$ be a partition of X into measurable subsets such that $\mu(A_i) > 0$, for all $i = 1, \ldots, n$. For any $x \in X$, let $A(x)$ denote the unique A_i such that $x \in A_i$. Consider the sigma-algebra $\mathscr{A}$ generated by ξ. Then, for any $f \in L^1(X, \mathscr{B}, \mu)$, one has

$$E(f|\mathscr{A})(x) = \frac{1}{A(x)} \int_{A(x)} f\, d\mu, \forall x \in X.$$

Example B.3. Let $(X_i, \mathscr{B}_i, \mu_i)$ be a probability space for $i = 1, 2$. Consider the product space $X \stackrel{\text{def}}{=} X_1 \otimes X_2$ equipped with the product sigma algebra $\mathscr{B} \stackrel{\text{def}}{=} \mathscr{B}_1 \otimes \mathscr{B}_2$ and the product probability measure $\mu \stackrel{\text{def}}{=} \mu_1 \otimes \mu_2$. Let $\mathscr{A} \stackrel{\text{def}}{=} \mathscr{B}_1 \times \{\emptyset, X_2\}$. Then it is easy to see that, $\forall f \in L^1(X, \mathscr{B}, \mu)$,

$$E(f|\mathscr{A})(x_1, x_2) = \int_{X_2} f(x_1, t)\, d\mu_2(t), \forall (x_1, x_2) \in X.$$

We now state some basic properties of the conditional expectation operator without proofs. The proofs are quite typical, and can be found in any standard textbook in Probability theory, e.g., [8, §8.2], or [1, §5.1] or [5, §9.2].

Theorem B.1 (Basic properties)**.** *Assume that $(X, \mathscr{B}, \mu)$ is a probability space and $\mathscr{A} \subseteq \mathscr{B}$ is a sub-sigma-algebra. Suppose that $f \in L^1(X, \mathscr{B}, \mu)$. Then the following hold:*

1. *If $f \in L^1(X, \mathscr{A}, \mu)$ then, $E(f|\mathscr{A}) \stackrel{a.e.}{=} f$.*
2. *Positivity: $f \geq 0 \Longrightarrow E(f|\mathscr{A}) \geq 0$ a.e.*
3. *Product rule: $\forall g \in L^\infty(X, \mathscr{A}, \mu)$, $E(gf|\mathscr{A}) \stackrel{a.e.}{=} gE(f|\mathscr{A})$.*
4. *Triangle inequality: $|E(f|\mathscr{A})| \leq E(|f||\mathscr{A})$ a.e.*
5. $\|E(\cdot|\mathscr{A})\| = 1$.
6. *If $\mathscr{A}' \subseteq \mathscr{A}$ is a sub-sigma-algebra then, $E(E(f|\mathscr{A})|\mathscr{A}') \stackrel{a.e.}{=} E(f|\mathscr{A}')$.*
7. *If $T : X \longrightarrow X$ preserves μ then, $E(f|\mathscr{A}) \circ T \stackrel{a.e.}{=} E(f \circ T|T^{-1}\mathscr{A})$, where $T^{-1}\mathscr{A} \stackrel{def}{=} \{T^{-1}(E) : E \in \mathscr{A}\}$.*
8. *Sup/inf property: Consider a sequence $\{f_n\}_{n=1}^\infty$ in $L^1(X, \mathscr{B}, \mu)$. Then*
   - *(i)* $E\left(\sup_n f_n|\mathscr{A}\right) \geq \sup_n E(f_n|\mathscr{A})$ *a.e., provided $\sup_n f_n$ is integrable.*
   - *(ii)* $E\left(\inf_n f_n|\mathscr{A}\right) \leq \inf_n E(f_n|\mathscr{A})$ *a.e., provided $\inf_n f_n$ is integrable.*

9. *Conditional monotone convergence theorem: Let* $\{f_n\}_{n=1}^{\infty}$ *be an increasing sequence of nonnegative functions in* $L^1(X, \mathscr{B}, \mu)$ *and* $f_n \xrightarrow[n\to\infty]{a.e.} f$. *Then* $E(f_n|\mathscr{A}) \xrightarrow[n\to\infty]{a.e.} E(f|\mathscr{A})$.

10. *Conditional Fatou's lemma:* $E\left(\liminf_n f_n|\mathscr{A}\right) \leq \liminf_n E(f_n|\mathscr{A})$ *a.e., for every sequence* $\{f_n\}_{n=1}^{\infty}$ *of nonnegative functions in* $L^1(X, \mathscr{B}, \mu)$ *such that* $\liminf_n f_n \in L^1(X, \mathscr{B}, \mu)$.

11. *Conditional dominated convergence theorem: Let* $\{f_n\}_{n=1}^{\infty}$ *be a sequence in* $L^1(X, \mathscr{B}, \mu)$ *and g be a nonnegative integrable function be such that* $|f_n| \leq g$ *a.e. If* $f_n \xrightarrow[n\to\infty]{a.e.} f$ *then* $E(f_n|\mathscr{A}) \xrightarrow[n\to\infty]{} E(f|\mathscr{A})$ *a.e. and also in* L^1.

12. *Conditional Jensen's inequality: Let* $g : \mathbb{C} \longrightarrow \mathbb{R}$ *be a convex function such that* $g \circ f$ *is integrable. Then one has* $g \circ E(f|\mathscr{A}) \leq E(g \circ f|\mathscr{A})$ *a.e.*

13. *Conditional Holder's inequality:* $\forall p, q \in [1, \infty]$ *satisfying* $\dfrac{1}{p} + \dfrac{1}{q} = 1$, *one has the following:*

$$E(|\varphi\psi||\mathscr{A}) \leq E(|\varphi|^p|\mathscr{A})^{\frac{1}{p}} E(|\psi|^q|\mathscr{A})^{\frac{1}{q}} \text{ a.e.}, \forall \varphi \in L^p(X, \mathscr{B}, \mu) \text{ and } \psi \in L^q(X, \mathscr{B}, \mu).$$

C Decomposition of linear functionals

Throughout this section, we follow the notation and terminologies of [2, §5.2 and §7.3].

Let $\mathcal{X}$ be a vector space over $\mathbb{C}$. Recall that if f is a complex linear functional on $\mathcal{X}$ and $u \stackrel{\text{def}}{=} \Re f$ then $f(x) = u(x) - iu(ix), \forall x \in \mathcal{X}$. Conversely, if u is a real liner functional on $\mathcal{X}$ and $f(x) \stackrel{\text{def}}{=} u(x) - iu(x)$, $\forall x \in \mathcal{X}$, then f is a complex linear functional. Furthermore, if $\mathcal{X}$ is normed and f is bounded then $\|u\| = \|f\|$.

For a locally compact Hausdorff topological space X, denote the bounded real linear functionals on $C_0(X)$ by $C_0(X, \mathbb{R})$. The following shows that such real linear functionals admit a *Jordon decomposition*:

Proposition C.1. *[2, Lemma 7.15]: If* $I \in C_0(X, \mathbb{R})^*$, *there exists positive linear functionals* I^+ *and* I^- *in* $C_0(X, \mathbb{R})^*$ *such that* $I = I^+ - I^-$.

For a proof of Proposition C.1, see [2, §7.3].

Now observe that every $I \in C_0(X)^*$ is uniquely determined by its restriction J to $C_0(X, \mathbb{R})$, and clearly we have bounded real lienar functionals J_1, J_2 such that $J = J_1 + iJ_2$. From Proposition C.1, one concludes the following:

Corollary C.2. *Every* $I \in C_0(X)^*$ *can be written as* $I_1 - I_2 + i(I_3 - I_4)$, *where* I_1, I_2, I_3 *and* I_4 *positive real linear functionals.*

One important consequence of Corollary C.2 is as follows:

Theorem C.3 (Dominated convergence theorem for bounded linear functionals)**.** *Let X be a compact metric space, $\{f_n\}_{n=1}^{\infty}$ be a sequence in $C(X)$ and $f \in C(X)$ be such that $f_n(x) \xrightarrow[n\to\infty]{} f(x)$, for all $x \in X$. Assume that there exists $g \in C(X)$ such that $\forall x \in X$ and $n \in \mathbb{N}$, $|f_n(x)| \leq g(x)$. Then for any $I \in C(X)$, one has $If_n \xrightarrow[n\to\infty]{} If$.*

D Equidistribution of sequences

Let X be a compact metric space and $\mu \in \mathscr{M}(X)$.

Definition D.1. A sequence $\{x_n\}_{n=1}^{\infty}$ in X is said to *equidistributed* with respect to μ if

$$\forall f \in C(X),\ \frac{1}{n}\sum_{i=1}^{n} f(x_i) \xrightarrow[n\to\infty]{} \int_X f\, d\mu, \tag{D.1}$$

or equivalently, $\dfrac{\delta_{x_1} + \cdots + \delta_{x_n}}{n} \xrightarrow[n\to\infty]{\text{weakly}} \mu$.

Note D.1. (D.1) holds if (and only if)

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) \xrightarrow[n\to\infty]{} \int_X f\, d\mu \text{ is true for all } f \text{ belonging to some dense } \mathcal{S} \subseteq C(X). \tag{D.2}$$

As an immediate consequence of Theorem A.1, we obtain

Theorem D.1. *Let X and μ be as above. Then the following are equivalent for a sequence $\{x_n\}_{n=1}^{\infty}$ in X:*

1. *$\{x_n\}_{n=1}^{\infty}$ equidistributes with respect to μ; and*
2. *$\dfrac{\#\{1 \leq i \leq n : x_i \in A\}}{n} \xrightarrow[n\to\infty]{} \mu(A)$, for every Borel $A \subseteq X$ satisfying $\mu(\partial A) = 0$.*

Definition D.2. Let X and μ be as above. We say $\mathscr{C} \in C(X)$ is a *convergence determining class* with respect to μ if, for any sequence $\{x_n\}_{n=1}^{\infty}$,

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) \xrightarrow[n\to\infty]{} \int_X f\, d\mu, \text{ for all } f \in \mathscr{C} \Longrightarrow \{x_n\}_{n=1}^{\infty} \text{ is equidistributed w.r.t. } \mu.$$

It is easy to see from (D.2) that, if for $\mathscr{C} \in C(X)$, $\text{span}_{\mathbb{C}}\mathscr{C}$ is dense in $C(X)$, then $\mathscr{C}$ is a convergence determining class with respect to any $\mu \in \mathscr{M}(X)$. In fact, we have the following in view of Stone-Weirstrass theorem:

Theorem D.2. *Let X be a compact metric space and $\mathscr{C} \subseteq C(X)$. If $\mathscr{C} \in C(X)$ is a unital sub-algebra of $C(X)$ that separates points and closed under complex conjugation, then $\mathscr{C}$ is a convergence deterimining class with respect to any $\mu \in \mathscr{M}(X)$.*

Corollary D.3. *For any compact abelian group G, $\hat{G}$ is a convergence determining class with respect to any Borel probability measure on G.*

Proof. Trivially $\mathrm{span}_{\mathbb{C}}\hat{G}$, i.e., the collection of all *trigonometric polynomials* on G, is sub-algebra of $C(G)$ containing all constant functions. It is obviously closed under complex conjugation. Theorem D.2 now applies since $\hat{G}$ separates points (see [13, §1.5.2]). □

Corollary D.4. *Let G be as above in Theorem D.3. Denote by λ the Haar probability measure on G. Consider a sequence $\{x_n\}_{n=1}^{\infty}$ in G. Then $\{x_n\}_{n=1}^{\infty}$ equidistributes with respect to λ if and only if*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\chi(x_i)=0,\ \textit{for all nontrivial character } \chi \textit{ of } G. \tag{D.3}$$

Proof. 'Only if' is trivial. To show the 'if' part, observe that

$$\int_G \chi\, d\lambda = \begin{cases} 1 & \text{if } \chi \equiv 1 \\ 0 & \text{if } \chi \text{ is nontrivial.} \end{cases}$$

So (D.3) confirms that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\chi(x_i)=\int_G \chi\, d\lambda$, for every $\chi\in\hat{G}$; thereby the conclusion follows in view of Corollary D.3. □

Corollary D.5 (Weyl's criterion)**.** *A sequence $\{\overline{x_n}\}_{n=1}^{\infty}$ in $\mathbb{T}$ is equidistributed w.r.t. the Haar probability measure if and only if*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i k x_j}=0,\ \forall k\in\mathbb{Z}\setminus\{0\}.$$

More generally, a sequence $\{\overline{\mathbf{x}_n}\}_{n=1}^{\infty}$ in $\mathbb{T}^d$ is equidistributed w.r.t. the Haar probability measure if and only if

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i\langle \mathbf{k},\mathbf{x}_j\rangle}=0,\ \forall \mathbf{k}\in\mathbb{Z}^d\setminus\{(0,\dots,0)\}.$$

Proof. Immediate from Corollary D.4. □

Example D.1. For all $\alpha\in\mathbb{R}\setminus\mathbb{Q}$, $\{\overline{n\alpha}\}_{n=1}^{\infty}$ is equidistributed in $\mathbb{T}$ w.r.t. the Haar probability measure. More generally, if $\alpha=(\alpha_1,\dots,\alpha_d)$ is such that $1,\alpha_1,\dots,\alpha_d$ are linearly independent over $\mathbb{Q}$, then, for any $\theta\in\mathbb{R}^d$, the sequence $\{\overline{n\alpha+\theta}\}_{n=1}^{\infty}$ is equidistributed in $\mathbb{T}$ w.r.t. the Haar probability measure.

Remark D.1. In Example D.1, the linear independence assumption is necessary. Otherwise, there exists $\mathbf{k}\in\mathbb{Z}^d\setminus\{(0,\dots,0)\}$ such that $\langle\mathbf{k},\alpha\rangle\in\mathbb{Z}$. Then clearly $\forall n\in\mathbb{N}$, $\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i j\langle\mathbf{k},\alpha\rangle}=1$. Hence by Weyl's criterion, $\{\overline{n\alpha}\}_{n=1}^{\infty}$ does not equidistribute in $\mathbb{T}^d$ w.r.t. the Haar probability measure.

Example D.2 (Normal number)**.** Let $D \geq 2$ be an integer. Recall that, every $x \in [0,1)$ can be represented in base D as $\frac{a_1}{D} + \frac{a_2}{D^2} + \cdots$, where each $a_i \in \{0, 1, \ldots, D-1\}$. A number $x \in [0, 1)$ said to be *normal to base D* if, for any $\ell \in \mathbb{N}$ and every word $w_1 w_2 \ldots w_\ell$ in $\{0, 1, \ldots, D-1\}$ of length ℓ, one has

$$\frac{\#\{1 \leq i \leq n : a_i = w_1, \ldots, a_{i+\ell-1} = w_\ell\}}{n} \xrightarrow[n\to\infty]{} \frac{1}{D^\ell}.$$

We say $x \in \mathbb{R}$ is *normal to base D* if $x - [x]$ is normal to base D. If a real number is normal to every integer base, we call it a *Normal number*. The following are equivalent for any $x \in \mathbb{R}$ and integer $D \geq 2$:

1. x is normal to base D; and
2. $\{\overline{D^n x}\}_{n=1}^{\infty}$ is equidistributed in $\mathbb{T}$ w.r.t. the Haar probability measure.

The reader is referred to [10, Chapter 1, §8, Theorem 8.1] for a proof of the above mentioned two statements.

Example D.3 (Weyl's theorem ([15]))**.** Let $f(x) \in \mathbb{R}[x]$ be such that at least one coefficient of $f(x)$ other than the constant term is irrational. Then $\{\overline{p(n)}\}_{n=1}^{\infty}$ equidistributes in $\mathbb{T}$ w.r.t. the Haar probability measure. A proof of this theorem using topological dynamics is provided in [3, Chapter 3, §3, Theorem 3.13].

The following theorem provides the existence of a desired convergence determining class for any compact metric space, regardless of any particular Borel probability measure.

Theorem D.6. *[10, Chapter 3, §2, Theorem 2.1] For any compact metric space X there is a countable family $\mathscr{C}$ of real valued continuous functions on X such that for any $\mu \in \mathscr{M}(X)$, $\mathscr{C}$ is a convergence determining class with respect to μ.*

For a proof of Theorem D.6, see [10, Chapter 3, §2].

References

[1] Einsiedler, Manfred and Ward, Thomas, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011. xviii+481 pp. ISBN: 978-0-85729-020-5

[2] Folland, Gerald B. *Real analysis. Modern techniques and their applications*, Second Edition. Pure and Applied Mathematics (New York). A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999. xvi+386 pp. ISBN: 0-471-31716-0

[3] Furstenberg, H., *Recurrence in ergodic theory and combinatorial number theory*, M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981. xi+203 pp. ISBN: 0-691-08269-3

[4] Furstenberg, H.; Kesten, H. *Products of random matrices*, Ann. Math. Statist. 31 (1960), 457–469

[5] Hochman, Michel, *Notes on ergodic theory*, January 27, 2013

[6] Karlsson, A., *A proof of the subadditive ergodic theorem*, December, 2014

[7] Kingman, J. F. C. *The ergodic theory of subadditive stochastic processes*, J. Roy. Statist. Soc. Ser. B 30 (1968), 499–510.

[8] Klenke, Achim, *Probability theory. A comprehensive course*, Second edition. Translation from the German edition. Universitext. Springer, London, 2014. xii+638 pp. ISBN: 978-1-4471-5360-3; 978-1-4471-5361-0

[9] Krengel, Ulrich, *Ergodic theorems. With a supplement by Antoine Brunel*, De Gruyter Studies in Mathematics, 6. Walter de Gruyter & Co., Berlin, 1985. viii+357 pp. ISBN: 3-11-008478-3

[10] Kuipers, L.; Niederreiter, H., *Uniform distribution of sequences*, Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974. xiv+390 pp.

[11] Parry, William, *Topics in ergodic theory*, Cambridge Tracts in Mathematics, 75. Cambridge University Press, Cambridge-New York, 1981. x+110 pp. ISBN: 0-521-22986-3

[12] Parthasarathy, K. R., *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3 Academic Press, Inc., New York-London 1967 xi+276 pp.

[13] Rudin, Walter, *Fourier analysis on groups*, Interscience Tracts in Pure and Applied Mathematics, No. 12 Interscience Publishers (a division of John Wiley & Sons, Inc.), New York-London, 1962 ix+285 pp.

[14] Walters, Peter, *An introduction to ergodic theory*, Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982. ix+250 pp. ISBN: 0-387-90599-5

[15] Weyl, Hermann *Über die Gleichverteilung von Zahlen mod. Eins. (German)* Math. Ann. 77 (1916), no. 3, 313–352.

[16] Zimmer, Robert J., *Ergodic theory and semisimple groups*, Monographs in Mathematics, 81. Birkhäuser Verlag, Basel, 1984. x+209 pp. ISBN: 3-7643-3184-4