Ergodic Theory

Constantine Caramanis

May 6, 1999

1 Introduction

Ergodic theory involves the study of transformations on measure spaces. Interchanging the words "measurable function" and "probability density function" translates many results from real analysis to results in probability theory. Ergodic theory is no exception. Ergodic theory has fundamental applications in probability theory, starting from areas that are very well understood, such as finite state Markov chains. In this paper we define the main objects of interest in the study of ergodic theory, and in particular focus on some applications on so-called interval exchange transformations. Studying this particular class of transformations illuminates the general theory, for as we show, these transformations illustrate the independence of some notions that *a priori* might see connected.

Section 2 contains a discussion of the main definitions needed, including automorphisms, endomorphisms, flows, and semiflows. At this point we present the statement of the ergodic theorem, and the concept of ergodicity. Section 3 discusses several properties of general dynamical systems: 3.1 discusses the mixing properties of dynamical systems, 3.2 describes the automorphism T_E induced by "localizing" an automorphism T to some subset $E \subset M$ of the original measure space, and 3.3 investigates some properties of transformations that are also toplogical maps. Section 4 introduces interval exchange transformations in the context of the discussion in the previous two sections. Section 5 discusses a number of their properties, and also revisits some concepts introduced in sections 2 and 3.

Ergodic theory studies the evolution of dynamical systems, in the context of a measure space. Consider a stochastic process, that is, a series of random variables $\{X_t\}$ whose evolution is governed by some dynamics—say some transformation T. Renewal processes are particular types of stochastic processes such that for certain points in the sequence, say $t \in \{0 = i_0 < i_1 < \dots\}$, the process and its future evolution has the same probabilistic description.

Example 1 Consider a $G/G/1$ queue. The arrival process $\{A(t); t \geq 0\}$ to this queue specifies a renewal process. The associated counting process $\{N(t); t \geq 0\}$

that counts arrivals to an empty system is also a renewal process.

There are two main theorems about renewal processes which are essentially generalizations of the strong law of large numbers for random variables.

Theorem 1 (Strong Law) *If* $\{N(t); t \geq 0\}$ *is a renewal process with mean inter-renewal interval* $\bar{X}$*, then*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\bar{X}},$$

outside a set of measure zero (where the measure is a probability measure).

Note that by $N(t)$ we really mean $N(t,\omega)$ where ω is an element of the sample space. This theorem is a statement about so-called time averages. It says that for any point ω in the sample space, outside a set of measure zero, the time average exists. A similar theorem exists for ensemble averages, that is, averages over the points in the sample space:

Theorem 2 (Elementary Renewal Theorem) *If* $\{N(t); t \geq 0\}$ *is a renewal process with mean inter-renewal interval* $\bar{X}$*, then*

$$\lim_{t\to\infty} \frac{E[N(t)]}{t} = \frac{1}{\bar{X}}.$$

Together these theorems imply that for renewal processes time averages equal ensemble averages, outside a set of measure zero. This is a specialized version of the Ergodic Theorem which we state witout proof in section 2. As we show, the Ergodic Theorem holds for stochastic processes more general than renewal processes. We now turn to the basic definitions of objects involved in the study of ergodic theory.

2 Ergodicity and the Ergodic Theorem

We now define several objects which make the above ideas more precise, before going on to state the Ergodic Theorem.

2.1 Definitions and Basic Results

Whenever we speak of a measure μ, we always take $\mathcal{M}$ to denote the σ-algebra of measurable sets. Ergodic theory takes place in the setting of a measure space $(M, \mathcal{M}, \mu)$. The main objects of study are transformations of the measure space M that respect the measure μ.

Definition 1 *An* **automorphism** $T : M \longrightarrow M$ *is a bijective transformation such that* T, T^{-1} *respect the* σ*-algebra and are* μ*-invariant:*

$$\mu(E) = \mu(TE) = \mu(T^{-1}E), \quad \forall\, E \in \mathcal{M}.$$

Example 2 Any permutation $\pi \in S_n$ of the discrete measure space of n points, is μ-invariant, where μ is the uniform distribution. On the other hand if μ is some nonuniform mass distribution to the n points of M, π need not be measure preserving.

Definition 2 *An* **endomorphism** $T : M \longrightarrow M$ *is a surjective transformation such that* T^{-1} *respects the* σ*-algebra* $\mathcal{M}$*, and is measure preserving, as above.*

Example 3 Consider the Bernoulli shift operator defined on $M = \{0,1\}^{\mathbb{N}}$, the space of $\{0,1\}$ sequences:

$$T(a_0, a_1, a_1, \dots) = (a_1, a_2, \dots).$$

Note that if we identify M with the interval $[0,1]$ by the usual method of dyadic expansions, then the uniform measure on M corresponds to Lebesgue measure on the unit interval. Now let E_1 denote all the sequences (a) such that $a_0 = 1$. Then $\mu(E_1) = 1/2$, where as $\mu(TE_1) = 1$. However T is nonetheless measure preserving, because for endomorphisms, T measure preserving is a statement about the operator's inverse. Indeed for any measurable set E_0, we have $\mu(E_0) = \mu(T^{-1}E_0)$, since we have

$$\mu(\{x : x_{i_1+k} \in A_1, \dots x_{i_r+k} \in A_r\}) = \mu(\{x : x_{i_1} \in A_1, \dots x_{i_r} \in A_r\}),$$

for all $k \geq 0$. Therefore T is in fact μ-invariant.

Automorphisms and endomorphisms are examples of dynamic systems. For given either, we can look at the trajectories of points or sets in M under repeated application of the transformation or its inverse: $\{T^n\}$ or $\{T^{-n}\}$. If we consider a group (semigroup) of automorphisms (endomorphisms) parameterized by a one parameter group (semigroup) we have the generalization of this notion to continuous time.

Definition 3 *Suppose* $\{T^t\}$ *is a group (semigroup) of automorphisms (endomorphisms) parametrized by a one parameter family* $t \in \mathbb{R}$ *(*$t \in \mathbb{R}_+$*) i.e. we have* $T^{s_1}(T^{s_2}(x)) = T^{s_1+s_2}(x)$*. Suppose further that whever* f *is a measurable function on* M *the function*

$$\begin{array}{rcl} \Phi(x,t) : M \times \mathbb{R} & \longrightarrow & M, \quad \textit{given by} \\ (x,t) & \longmapsto & f(T^t x), \end{array}$$

is measurable on the Cartesian product $M \times \mathbb{R}$*. Then we call* $\{T^t\}$ *a* **flow** *(***semiflow***).*

The trajectories of dynamical systems, either discrete or continuous, are a primary tool for investigating their properties. Often in the theory of Markov processes or diffusion, it is of interest to find how often a particular trajectory of a dynamical system meets a set of positive measure. Along these lines we have the following theorem.

Definition 4 *For some measurable set E, a point $x \in E$ is called a recurrence point of the endomorphism T, if $T^n x \in E$ for some $n > 0$.*

Theorem 3 (Poincaré Recurrence Theorem) *If $T : M \longrightarrow M$ is an endomorphism (with respect to some particular measure μ) then for any set $E \in \mathcal{M}$, the set $E' \subset E$ of recurrence points satisfies:*

$$\mu(E - E') = 0,$$

that is, every point away from a set of measure zero is a recurrence point.

In fact this theorem implies that the points that are infinitely recurrent are of full measure.

Proof. The proof is essentially a consequence of the countable additivity of the measure. For setting $E_0 = E - E'$, we can also write

$$E_0 = A \cap \Big(\bigcup_{n=1}^{\infty} T^{-n}(M - E) \Big),$$

and hence E_0 is measurable. Note further that since for any point $x \in E_0$ the trajectory $\{T^n x\}$ does not meet E, it cannot meet E_0. But then the sets $E_0, T^{-1}E_0, T^{-2}E_0, \ldots$ must be disjoint. Furthermore, since T is an endomorphism, T^{-1} is measure preserving, and therefore we have

$$\mu\Big(\bigcup_{n=0}^{\infty} T^{-n} E_0 \Big) = \sum_{n=0}^{\infty} \mu(T^{-n}E_0) = \sum_{n=0}^{\infty} \mu(E_0),$$

and hence $\mu(E_0) = 0$, concluding the proof. □

In addition to the recurrence of trajectories, another very basic question is whether or not the measure space can somehow be partitioned with respect to a dynamical system and some function—for example whether it is possible to separate a measure space M into two disjoint subsets M_1, M_2, both of positive measure, such that both sets contain their respective trajectories under the given dynamical system. This question is exactly at the core of the concept of ergodicity. To this we now turn.

Definition 5 *If f is a measurable function, we say it is invariant with respect to some dynamical system if it is constant on the trajectories of the dynamical system.*

For example, any radial function is invariant with respect to the automorphism given by rotation of the plane. We say that a set $E \in \mathcal{M}$ is invariant with respect to some dynamical system if its indicator function is invariant. The next lemma shows that it makes sense to define sets and functions that are invariant *almost everywhere.*

Lemma 1 *If f is a function invariant away from a set of measure zero, then there exists a function f_0 that is invariant, and such that $f = f_0$ almost everywhere.*

PROOF. We prove this only for the discrete case, for a dynamical system given by an automorphism. The proof is by construction. Define the sets

$$\begin{aligned} E_0 &= \{x \in M \mid f(x) \neq f(Tx)\} \\ E_1 &= \bigcup_{n=-\infty}^{\infty} T^n E_0, \end{aligned}$$

Then $\mu(E_1) = 0$, and therefore defining

$$f_0(x) = \begin{cases} f(x) & \text{if } x \in M - E_1 \\ 0 & \text{otherwise,} \end{cases}$$

we have the desired function. □

Similarly we can show that if $E \in \mathcal{M}$ is invariant up to a set of measure zero, then we can find a set E_0 that is invariant, and satisfies $\mu(E \triangle E_0) = 0$. This allows us to make a very important definition.

2.2 Ergodicity

We have the following:

Definition 6 *A dynamical system is called* **ergodic** *if it has no nontrivial invariant subsets.*

Then if an ergodic system is ergodic and E is an invariant set, we must have $\mu(E) = 1$ or $\mu(E) = 0$. The following lemma is a refinement of this statement.

Lemma 2 *If a dynamical system is ergodic, then any invariant function must be constant on any set of full measure.*

PROOF. Suppose f is an invariant function for some dynamical system. Then write $E_a = \{x \mid f(x) < a\}$. Since these sets are all invariant they must either be of full measure, or of measure zero. This concludes the proof. □

The following theorem, often known as the Birkhoff-Khinchin Ergodic Theorem, is very basic in the study of ergodic theory. For a proof, see Billingsley [1] or Cornfeld [2].

Theorem 4 (Ergodic Theorem) *If $f \in L^1(M, \mathcal{M}, \mu)$ and T is an endomorphism, then*

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} f(T^{k-1}x) = \bar{f}(x),$$

outside a set of measure zero. Furthermore, $\bar{f}$ is invariant and integrable, and satisfies

$$\int_M \bar{f}(x)\, d\mu \;=\; \int_M f(x)\, d\mu.$$

Note that by lemma 2 above, if the dynamical system is ergodic, the function $\bar{f}$ must be constant almost everywhere. Therefore we have

$$\bar{f} = \int_M \bar{f}(x)\, d\mu = \int_M f(x)\, d\mu, \quad \text{almost everywhere.}$$

This yields an interpretation of the Ergodic Theorem which gives a generalization of the result for renewal processes given at the start of this paper: *Time means and space means are equal, away from a set of measure zero.*

Example 4 Consider again the example of the discrete, finite measure space with a point mass probability measure, and some permutation $\pi \in S_n$. We can write the permutation as a product of cycles. Then the permutation is measure preserving iff the members of each individual cycle are assigned the same weights. Suppose then that the permutation π is a product of two cycles C_1, C_2 of positive lengths k_1, k_2 respectively. Then for $E = C_1$ the limit in the theorem above is $1/k_1$ on C_2 and $1/k_2$ on C_1. Therefore unless the cycles are the same length, the function $\bar{f}$ is not constant away from a set of measure zero, and hence the dynamical system cannot be ergodic.

Example 5 The Bernoulli shift operator introduced above is ergodic. The fact that the shift operator is ergodic follows from Kolmogorov's *zero-one* law. This states that if $\{E_n\}$ is any sequence of independent measurable sets in some normalized measure space (i.e. some probability space) then the elements of the σ-algebra $\mathcal{F}$ generated by the intersection of the σ-algebras $\mathcal{F}_n$ that are respectively generated by all but the first n sets E_i,

$$\mathcal{F} = \bigcap_{n=1}^{\infty} \mathcal{F}_n = \bigcap_{n=1}^{\infty} \sigma(E_n, E_{n+1}, \dots),$$

are either of measure zero, or of full measure.

If a dynamical system is ergodic, then we have some additional information on its trajectories.

Proposition 1 *If a discrete time dynamical system is ergodic, then the orbit of any set of positive measure, has full measure.*

Proof. If T is our transformation and E our set of positive measure, we need to show that

$$\mu\Big(\bigcup_{n>0} T^{-n} E\Big) = 1.$$

But this is a consequence of ergodicity. For writing $F = \bigcup_{n>0} T^{-n}E$, then F is an invariant set away from a set of measure zero. Therefore we can find some $F' \subset F$ that is an invariant set. By ergodicity, F' must have full measure, and hence so must F. □

The Ergodicity of a dynamical system depends on the particular measure μ on the space. A natural question to ask is what information we have about the ergodicity of the dynamical system with respect to some other (invariant) measure ν. To this end we have the following theorem.

Theorem 5 *If T is a dynamical system invariant with respect to the normalized measures μ and ν, then if T is ergodic with respect to μ, and ν is absolutely continuous with respect to μ, then $\mu = \nu$. If T is ergodic with respect to both measures, then μ and ν are either equal, or they are singular.*

PROOF. We prove the theorem for discrete time dynamical systems. For the first part of the theorem, by theorem 4 we have for any set $E \in \mathcal{M}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k x) = \mu E,$$

away from a set of measure zero. Then by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \int_M \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k x)\, d\nu = \int_M \mu(E)\, d\nu = \mu(E).$$

However, since

$$\int_M \sum_{k=0}^{n-1} \chi_E(T^k x)\, d\nu = \nu(E),$$

for any n, the first part of the theorem is proven.

For the second part, suppose that the two measures are distinct: $\mu \neq \nu$, i.e. for some measurable E, $\mu(E) \neq \nu(E)$. Now let E_μ, E_ν be the sets for which the limit in the ergodic theorem holds, for μ, ν respectively:

$$E_\mu = \left\{ x \in M \mid \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k x) = \mu(E) \right\},$$

and similarly for ν. By ergodicity, E_μ, E_ν both must have full measure with respect to their respective measures. Since they are by definition disjoint sets, the measures μ, and ν must indeed be singular, as required. □

3 Dynamical Systems

In this section we state and prove a number of additional properties of general dynamical systems.

3.1 Mixing Transformations

The Poincaré Recurrence Theorem ensures that the orbit of almost every point meets any set of positive measure infinitely often. A stronger question to answer is of a statistical nature: For two sets E_1, E_2 of positive measure, what can we say about the measure of the intersection $E_1 \cap T^n E_2$ asymptotically as $n \to \infty$? As we now make precise, how much we can say to answer this question depends on the transformation T determining the dynamical system.

Definition 7 *We say that a dynamical system is* **mixing** *if for any functions* $f, g \in L^2(M, \mathcal{M}, \mu)$ *we have*

$$\lim_{n\to\infty} \int_M f(T^n x) g(x)\, d\mu = \int_M f(x)\, d\mu \cdot \int_M g(x)\, d\mu,$$

for discrete dynamic systems ($n \in \mathbb{N}$ for endormorphisms, $n \in \mathbb{Z}$ for automorphisms), and similarly

$$\lim_{t\to\infty} \int_M f(T^t x) g(x)\, d\mu = \int_M f(x)\, d\mu \cdot \int_M g(x)\, d\mu,$$

for continuous time dynamic systems (flows and semiflows).

If we choose f, g to be the characteristic functions of two measurable sets E_1, E_2 the condition for mixing for discrete time dynamical systems requires:

$$\lim_{n\to\infty} \mu(E_1 \cap T^{-n} E_2) = \mu(E_1) \cdot \mu(E_2).$$

This provides an answer to our initial question: if a transformation T is mixing, E_1, E_2 two measurable sets of positive measure, then the orbit of E_2 intersects E_1 in a region with measure that is asymptotically proportional to the measure of E_1. Note that there are no *a priori* conditions on the dependence or independence of the sets E_1, E_2.

Proposition 2 *If a transformation T is mixing then it is also ergodic.*

Proof. Suppose E_1 is invariant under the mixing automorphism T. Then $T^{-n} E_1 = E_1$. But then by the above, if T is mixing we have

$$\lim_{n\to\infty} \mu(E_2 \cap T^{-n} E_1) = \mu(E_1) \cdot \mu(E_2),$$

and taking $E_1 = E_2$ we have

$$\mu(E_1) = \mu(E_1) \cdot \mu(E_1).$$

The only real numbers that satisfy $\lambda^2 = \lambda$ are 0 and 1, therefore concluding the proof. □

As the following example shows, and as is discussed in detail with the example of interval exchange transformations in section 5, the converse to this proposition is not true, that is, ergodicity and mixing are not equivalent.

Example 6 Consider the rotation transformation on the circle with the Lebesgue measure. If the rotation is by some root of unity, then the transformation is not ergodic. For in this case the transformation is periodic of period n. Then consider the intervals (on the circle) of length $1/2n$ centered at each of the n^{th} roots of unity. These form an invariant set of positive but not full measure. We now show that if T is rotation by some complex number z that is not a root of unity, then T is ergodic. Let α denote the angle through which the circle is translated. Suppose E is some invariant set of positive measure. Then consider the Fourier transform of its characteristic function χ_E. If a_n denotes the n^{th} Fourier coefficient, we have (up to a constant):

$$\begin{aligned} a_n &= \int_0^{2\pi} e^{in\theta}\chi_E(\theta)\,d\theta \\ &= \int_z^{2\pi+\alpha} e^{in\theta+in\alpha}\chi_E(\theta+\alpha)\,d\theta \\ &= e^{in\alpha}\int_0^{2\pi} e^{in\theta}\chi_E(\theta)\,d\theta \\ &= e^{in\alpha}a_n. \end{aligned}$$

Since $\alpha \neq 0$ and z is not a root of unity, this implies that the coefficients are given by:

$$a_n = \begin{cases} \mu(E) & \text{if } n = 0 \\ 0 & \text{otherwise.} \end{cases}$$

But since we can recover integrable functions a.e. from their Fourier coefficients, we must have $\mu(E) = 1$, the desired result.

3.2 Induced Automorphisms

Often it is of interest to understand the action of a dynamical system on some particular subset of the measure space. In this section we investigate this question, and in the process give a method for constructing new automorphisms from old ones. Suppose T is an automorphism of some measure space $(M, \mathcal{M}, \mu)$ with normalized measure, and let $E \subset M$ be some measurable set. We want to define the automorphism T_E induced by T on the new measure space $(E, \mathcal{M}_E, \mu_E)$ where $\mathcal{M}_E = \mathcal{M} \cap E$, and $\mu_E(A) = \mu(A)/\mu(E)$. By the Poincaré Recurrence Theorem the set of recurrent points in E are of full measure in E, so we can assume that all of E is recurrent. Let $k_E : E \to \mathbb{N}$ be the function that gives the first "return time" to E, i.e.

$$k_E(x) = \min\{n \in \mathbb{N} \mid T^n x \in E\}.$$

Then, if we let $E_n = \{x \in E \mid k_E(x) = n\}$ we have

$$\begin{aligned}
\mu\Big(\bigcup_{n\geq 0} T^n E\Big) &= \mu\Big(\bigcup_{n\geq 0}\bigcup_{i=0}^{n-1} T^i E_n\Big) \\
&= \sum_{n=1}^{\infty}\sum_{i=0}^{n-1} \mu(T^i E_n) \\
&= \sum_{n=1}^{\infty} = \int_E k_E(x)\, d\mu.
\end{aligned}$$

Therefore $k_E \in L^1(E, \mathcal{M}_E, \mu_E)$. Then setting $T_E x = T^{k_E(x)} x$ we have an automorphism of E with itself. To show that it is measure preserving, take any measurable set A. Then

$$A = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A \cap E_n.$$

Since the sets $T^n A_n$ are disjoint, we have

$$\mu_E(T_E A) = \sum_{n=1}^{\infty} \mu_E(T^n A_n) = \sum_{n=1}^{\infty} \mu_E(A_n) = \mu_E(A),$$

which is exactly what we need.

3.3 Continuous Transformations

In addition to being a bijective, measure preserving map from the measure space to itself, often an automorphism will have additional structural properties, such as continuity, that allow further investigation of its action on the measure space. Transformations that are homeomorphisms of the measure space to itself are called topological dynamical systems. Perhaps surprisingly, for M a compact metric space, the opposite is also true, namely topological automorphisms, i.e. homeomorphisms, can be realized as measure preserving transformations under a proper choice of normalized measure μ (see Cornfeld [2]). This allows us to analyze the behavior of continuous maps of metric spaces in the context of ergodic theory. The following definitions characterize topological nondecomposability of homeomorphisms on compact metric space M.

Definition 8 *A homeomorphism $T : M \to M$ is called* **topologically transitive** *if the trajectory of some point $x \in M$ is a dense subset of M.*

Example 7 Consider the continuous map from the circle to itself given by $T(x) = 2x \pmod 1$. This transformation is continuous (as viewed on the circle). Furthermore, it is topologically transitive. To see this, consider the point on the circle that corresponds to the point ω on the unit interval with dyadic expansion:

$$\omega = (1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, \dots)$$

This point is constructed by concatenating all the possible n-digit sequences, for all $n \in \mathbb{N}$. When viewed on the space of sequences, T is no more than the Bernoulli shift operator. Therefore the orbit of this point is in fact a dense set on the circle, and the transformation is topologically transitive, as claimed. Note however that the point given by alternating 1's and 0's has an orbit consisting of only two points.

Definition 9 *We call the homeomorphism $T : M \to M$* **minimal** *if the trajectory of any point $x \in M$ forms a dense subset.*

Definition 10 *We call the homeomorphism $T : M \to M$* **uniquely ergodic** *if it is invariant for exactly one normalized Borel measure μ.*

Proposition 3 *If $T : M \to M$ is uniquely ergodic, then it is ergodic with respect to its normalized invariant measure.*

PROOF. If T is not ergodic, then we can find some set E with $0 < \mu(E) < 1$. But then the measure given by

$$\mu_1(A) = \frac{1}{\mu(E)}\mu(A \cap E),$$

is also a normalized measure, and T is also μ_1-measure preserving. □

While many transformations T are either both minimal, and uniquely ergodic, or not minimal or uniquely ergodic, the two concepts are independent. We show this by means of interval exchange transformations in section 5.3. But first we must define this class of transformations. To this we now turn, defining the transformations, and discussing a number of their properties, including some necessary conditions for ergodicity.

4 Interval Exchange Transformations I

We have already seen an example of an automorphism that is an interval exchange transformation, namely, the rotation of the circle. If we identify the unit interval $M = [0, 1)$ with the circle, then a rotation through angle θ corresponds to the exchange of the intervals $\Delta_1 = [0, \theta/2\pi), \Delta_2 = [\theta/2\pi, 1)$. As shown, the rotation is ergodic iff the angle of rotation does not coincide with any root of unity on the unit circle, that is, the action of the transformation is not periodic. General interval exchange transformations can be thought of as generalizations of rotations of the circle.

Definition 11 *Suppose the semi-interval $M = [0, 1)$ is partitioned into n consecutive semi-intervals $\{\Delta_1, \ldots, \Delta_n\}$. An interval exchange transformation T on $M = [0, 1)$ operates on each interval Δ_i by translation by some amount α_i, such that the first application of T corresponds to a permutation of the $\{\Delta_i\}$ by some permutation $\pi \in S_n$.*

Note then that for a general interval exchange transformation (IET) T^2 is not a permutation of the same intervals as T, or T^{-1}. However if T is an IET, all powers of T are also IETs. Specifically, if T permutes the semi-intervals $\{\Delta_i\}$ then T^{-1} permutes the intervals $\{T\Delta_i\}$, and T^n permutes intervals of the form

$$\Delta_i^{'} = \Delta_{i_0} \cap T^{-1}\Delta_{i_1} \cap T^{-2}\Delta_{i_2} \cap \cdots \cap T^{-n+1}\Delta_{i_{n-1}}.$$

Since T and T^{-1} send semi-intervals to semi-intervals, this intersection is either empty, or also a semi-interval. IETs are invertible, and it is clear that they preserve Lebesgue measure. Therefore they are automorphisms of the measure space $(M = [0,1), \mathcal{M}, \mu)$ where μ is Lebesgue measure, and $\mathcal{M}$ the Lebesgue-measurable sets. The next proposition states that just as in the case of rotations of the circle, if an IET has any periodic points, it cannot be ergodic.

Proposition 4 *If T an interval exchange transformation, has any periodic points, then T is not ergodic.*

PROOF. First note that if x is a fixed point of T, then the entire semi-interval containing x must be fixed under T. Similarly, if x is a periodic point of period n, then T^n must in fact fix some semi-interval containing x. But then T cannot be ergodic, for the orbit E of a sufficiently small subinterval of the periodic interval will be an invariant set with $0 < \mu(E) < 1$. □

Therefore if T is ergodic, it must be aperiodic. Unfortunately, ergodic IETs are not quite so easy to characterize: the converse does not hold, i.e. an IET can be aperiodic but not ergodic.

Example 8 Let T_1 be an ergodic IET on $M_1 = [0,1)$ and T_2 be an ergodic IET on $M_2 = [1,2)$. Then if we define on $M = [0,2)$,

$$T = \begin{cases} T_1(x) & \text{for } x \in M_1 \\ T_2(x) & \text{for } x \in M_2, \end{cases}$$

then T is aperiodic but not ergodic.

However we can say some things about aperiodic interval exchange transformations.

Theorem 6 *For T an IET, The following are equivalent:*

(i) *T is aperiodic, i.e. it has no periodic points;*

(ii) *The diameter of the intervals exchanged by T^n goes to zero as $n \to \infty$;*

(iii) *The union of the orbits of the left endpoints of the intervals Δ_i is dense in $M = [0,1)$.*

PROOF. By contrapositive argument, and by proposition 4 the second and third statements both imply the first. Also, the second and third can be seen to be equivalent if we notice that the partition of $[0, 1)$ formed by the intervals exchanged by T^n are a refinement of the partition formed by intervals exchanged by T^m when $m < n$. Finally, if the orbit of the endpoints is not dense, then the complement of its closure is a nonempty open set, and is therefore the countable union of intervals, which T somehow exchanges. Since there are a finite number of intervals of a given length, the possible permutations are finite, and hence T must have periodic points. This concludes the proof. □

5 Interval Exchange Transformations II

We now go on to prove some more properties of IETs. First, we discuss the number of invariant measures, then we show IETs have no mixing properties, and finally, we discuss the concepts of minimality and unique ergodicity as they apply to IETs.

5.1 Invariant Measures

In section 3.3 we defined a uniquely ergodic transformation to be one that is invariant with respect to exactly one normalized measure. In this section we show that there is, *a priori*, an upper bound on the number of normalized measures with respect to which an IET can be measure preserving and ergodic.

Theorem 7 *If T is an aperiodic IET exchanging r intervals, then*

(i) There exist at most r normalized measures with respect to which T is invariant and ergodic;

(ii) If μ is any measure for which T is measure preserving, then $M = [0, 1)$ can be divided into at most r T-invariant subsets of positive measure.

Proof. The first statement is an immediate consequence of the second, so we prove them in reverse order. Now, since T is an IET, it is a Lebesgue measure preserving automorphism. Then T defines a unitary operator U_T on $H = L^2(M, \mu)$ by $U_T f(x) = f(Tx)$. Suppose we can divide M into k T-invariant subsets of positive measure. Then the simple functions that are linear combinations of the indicator functions of these sets are T-invariant. Furthermore, their span is k dimensional. If we can show that the entire space of T-invariant functions in H is at most r dimensional, we will be done. We define the following notation:

$$\begin{aligned} H^{\text{inv}} &:= \{f \in H \mid U_T f = f\}; \\ H(h) &:= \text{Span}\{U_T^k h \mid -\infty < k < \infty\}; \\ h^{\text{inv}} &:= \text{Orthogonal projection of } h \text{ onto } H^{\text{inv}}; \\ h^{\perp} &:= h - h^{\text{inv}}. \end{aligned}$$

Since (by von Neumann)

$$\frac{1}{n}\sum_{k=0}^{n-1} U_T^k h \longrightarrow h^{\text{inv}}, \quad \text{as } n \to \infty,$$

$h^{\text{inv}} \in H(h)$ and hence so is $h^{\perp}$, and therefore $H(h^{\perp}) \subset H(h)$. Since U_T is unitary,

$$\begin{aligned} h^{\perp} \perp H^{\text{inv}} \quad &\Rightarrow \quad U_T^k h^{\perp} \perp H^{\text{inv}}, \quad \forall k \\ &\Rightarrow \quad H(h^{\perp}) \perp H^{\text{inv}}. \end{aligned}$$

This gives the orthogonal decomposition

$$H(h) = H(h^{\perp}) \oplus H(h^{\text{inv}}).$$

Note that $H(h^{\text{inv}})$ is either one or zero dimensional. Therefore if we can write

$$H^{\text{inv}} = H(h_1^{\text{inv}}) + \cdots + H(h_p^{\text{inv}}),$$

we can conclude H^{inv} is at most p dimensional. By the orthogonal decomposition above, this will follow if we can find p functions $\{h_i\}$ such that

$$H = H(h_1) + \cdots + H(h_p).$$

To this end, if T exchanges intervals Δ_i, let $h_i = \chi_{\Delta_i}$. By the continuity of the measure μ and by theorem 6, linear combinations of the indicator functions $\chi_{i_0,i_1,\dots i_m}$ of the intervals

$$\Delta_{i_0,i_1,dotsi_m} := \Delta_{i_0} \cap T\Delta_{i_1} \cap \cdots \cap T^m \Delta_{i_m}, \quad (m = 1, 2, \dots; i_k = 1, \dots, r),$$

(the intervals exchanged by T^m) are dense in H. But then we are done, for it can be verified that

$$\chi_{i_0,i_1,\dots i_m} = \sum_{i=1}^{r}\sum_{k=0}^{m} c_{ik} U_T^{-k} \chi_{\Delta_i},$$

and hence

$$\chi_{i_0,i_1,\dots i_m} \in H(\chi_{\Delta_1}) \oplus \cdots \oplus H(\chi_{\Delta_r}).$$

This concludes the proof of the second part of the theorem. For the first, suppose that $\mu_1, \dots, \mu_p$ $(p > r)$ are distinct measures for which T is ergodic. Then by theorem 5 the measures must be singular with respect to each other. Therefore we can find p sets $A_1, \dots, A_p$ that are invariant under T, and such that $\mu_i(A_j) = \delta_{ij}$. Now take μ to be the average of these measures:

$$\mu = \frac{1}{p}\sum_{i=1}^{p} \mu_i.$$

Then T is μ-invariant because it is μ_i-invariant, and $\mu(A_i) = 1/p$. Therefore there exist $p > r$ T-invariant sets with positive measure, contradicting part 2. This concludes the proof. □

5.2 Mixing of IETs

We have shown that in the class of aperiodic interval exchange transformations, some are ergodic, while others are not. In this section we show that no IET is ever mixing. First we state some lemmas.

Lemma 3 *Suppose we have a measure preserving transformation T, and a sequence of partitions $p_i = \{\Delta_1^{(i)}, \ldots, \Delta_{s_i}^{(i)}\}$ such that $\Delta_j^{(i)} \cap \Delta_k^{(i)} = \emptyset$ for $j \neq k$ and for all i, and such that $\mu(M - \bigcup_{n=1}^{s_i} \Delta_n^{(i)}) = 0$. Suppose further that we have a sequence $r^{(i)} = \{r_1^{(i)}, \ldots, r_{s_i}^{(i)}\}$ such that we have:*

- *(i)* *The number of elements s_i in each partition is bounded above by some number s;*
- *(ii)* *The lower bound of each family $r^{(i)}$ goes to ∞ as $i \to \infty$;*
- *(iii)* *For any $E \in \mathcal{M}$,*

$$lim_{i\to\infty}\mu\Big(E \cap \Big(\bigcup_{j=1}^{s_i} T^{r_j^{(i)}}(\Delta_j^{(i)} \cap E)\Big)\Big) = \mu(E).$$

then T is not mixing.

Proof. By straightforward manipulation we have

$$\begin{aligned}
\mu\Big(E \cap \Big(\bigcup_{j=1}^{s_i} T^{r_j^{(i)}}(\Delta_j^{(i)} \cap E)\Big)\Big) &= \mu\Big(\bigcup_{j=1}^{s_i}(E \cap T^{r_j^{(i)}}(\Delta_j^{(i)} \cap E))\Big) \\
&\leq \sum_{j=1}^{s_i} \mu(E \cap T^{r_j^{(i)}}(\Delta_j^{(i)} \cap E)) \\
&\leq \sum_{j=1}^{s_i} \mu(E \cap T^{r_j^{(i)}} E).
\end{aligned}$$

But since $\min_j\{r_j^{(i)}\} \to \infty$ as $i \to \infty$, and since $s_i \leq s$, if T is mixing we have further,

$$\lim_{i\to\infty} \sum_{j=1}^{s_i} \mu(E \cap T^{r_j^{(i)}} E) \leq s \cdot \lim_{n\to\infty} \mu(E \cap T^n E) = s \cdot \mu(E)^2.$$

But s and E are chosen independently, and therefore if we choose E with $\mu(E) < \frac{1}{s}$, we contradict the third condition above. □

Lemma 4 *If T is an IET exchanging r intervals, and $M_1 = [a, b) \subset M$ is some subinterval, then the induced automorphism T_1 (as described in section 3.2) is an interval exchange of at most $r + 2$ intervals.*

Proof. We omit this proof. □

We now state and prove the main theorem of this section.

Theorem 8 *If T is an ergodic, μ-invariant IET (for some measure μ) then T is not mixing.*

Note that we assume that T is ergodic (and hence aperiodic) without loss of generality, since mixing implies ergodicity.

PROOF. The full proof is notationally very intensive, and aside from the basic ideas, involves set-theoretic manipulations using the properties of the measure. Then here we give only the fundamental ideas of the proof. For a full proof, see Cornfeld [2]. The main idea is to show that in fact any IET T satisfies the hypotheses of lemma 3 and hence is not mixing. To satisfy the hypotheses of lemma 3 we need to construct a sequence of partitions $\xi_i = \{A_i^{(i)}, \ldots, A_{s_i}^{(i)}\}$, and a sequence of families of numbers $r_j^{(i)}$, and show that they satisfy certain properties. The proof is in two parts: first demonstrate a method for constructing these objects, and then verify that they have the desired properties.

We are initially given an ergodic (and hence aperiodic) transformation T, and the measure space $(M = [0,1), \mathcal{M}, \mu)$. The first part of the construction generates a sequence $\{\Delta^{(i)}, T_i\}$ of nested subintervals, and the corresponding transformations induced by the original transformation T. The first element of the sequence, $(T_0 = T, \Delta^{(0)} = M)$ is given. Let $\Delta_j^{(0)}$ be the intervals exchanged by T_0. In general we choose $\Delta^{(i+1)} = \Delta_j^{(i)}$ for some j. We choose this j to be the index of the associated return function $k_l^{(i)}$ with the largest value (where this "return" function is as described in section 3.2). Given this subinterval, T_{i+1} is defined and is an exchange of intervals $\{\Delta_j^{(i+1)}\}$.

Now we are ready to define the numbers and partitions required to satisfy lemma 3. Let $T_{i,j}$ be the transformation induced by the pair $(T, \Delta_j^{(i)})$, and let $\Delta_{j,l}^{(i)} \subset \Delta_j^{(i)}$ be the intervals exchanged by $T_{i,j}$. We then define the desired partitions and numbers by

$$A_{j,l}^{(i)} = \bigcup_{p=0}^{k_j^{(i)}} T^p \Delta_{j,l}^{(i)}, \qquad r_{j,l}^{(i)} = k_{j,l}^{(i)}.$$

From here we have left to show that these partitions and families of numbers satisfy the hypotheses of lemma 3. The verification of the three properties is tedious, but straightforward, so we omit it. □

5.3 Minimality and Unique Ergodicity

Although in general interval exchange transformations are not homeomorphisms, there exist generalizations of the definitions given in section 3.3.

Definition 12 *For $T : M \to M$ an IET, we say it is* **minimal** *if the orbit of any point is a dense subset of M.*

Definition 13 *For $T : M \to M$ an IET, we say it is* **uniquely ergodic** *if it is invariant with respect to exactly one Borel measure.*

Note that since IETs are always invariant with respect to the Lebesgue measure, this must be the measure referred to in the definition. The next theorem shows that these two notions are independent.

Theorem 9 *For any positive integer m, there exists an IET that is minimal, and is invariant with respect to m distinct (normalized) measures.*

Proof. We construct the transformation as follows. Let S^1 be the unit circle, T_α the rotation of S^1 through an angle α, $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ with the uniform distribution, and $\sigma : S^1 \to \mathbb{Z}_m$ some measurable function. Then define the transformation of the space $M = S^1 \times \mathbb{Z}_m$ by

$$T(x, k) = (T_\alpha x, k \oplus \sigma(x)),$$

where $\oplus$ denotes the group operation in $\mathbb{Z}_m$. We will show that in fact this transformation possesses m ergodic normalized invariant measures. First we prove a lemma.

Lemma 5 *If $\alpha \in \mathbb{R} - \mathbb{Q}$ in the construction above, and there exists some measurable function $\tau : S^1 \to \mathbb{Z}_m$ such that $\tau(T_\alpha x) = \tau(x) \oplus \sigma(x)$ a.e., and the preimage in any interval, of any integer k, is of positive measure, then the transformation T resulting from the above construcion is an interval exchange transformation, is minimal, and possesses m ergodic normalized invariant measures.*

To see this lemma, we consider sets of the form

$$S_k = \{(x, \tau(x) \oplus k) \mid x \in S^1\}.$$

These sets S_k are T-invariant up to a set of measure zero. We find the m desired measures by restricting μ, the Lebesgue measure, to each of the sets S_k. Now for minimality, suppose to the contrary that T is not minimal. Then we can find some invariant set $E \neq M$ made up of a finite number of semi-intervals. For at least one k, $\mu(S_k \cap E) > 0$, and $\mu(S_k \cap (M - E)) > 0$. Then for this k, $S_k \cap E$ is T-invariant, up to a set of measure zero. Finally, identifying S_k with S_1, the set $S_k \cap E$ is a T_α-invariant set E' such that $0 < \mu(E') < 1$, a contradiction as we assumed α was irrational (and hence the rotation ergodic). The contradiction concludes the proof of the lemma.

Now to finish the proof of the theorem, we have only to concoct functions σ, τ and a point α, that satisfy the lemma's conditions. Let $\{p_s/q_s\}$ be a sequence of irreducible fractions that satisfy, for any given $\varepsilon > 0$,

$$\begin{aligned} \sum_{s=1}^{\infty} \frac{1}{q_s^{q+\varepsilon}} &< \frac{1}{2mq_1}, \\ \sum_{p=s+1}^{\infty} \frac{1}{q_p^{\varepsilon}} &< \frac{1}{2}\{q_s\alpha\}, \quad \text{for } s = 1, 2. \end{aligned}$$

Then choose some irrational α satisfying

$$0 < \alpha - \frac{p_s}{q_s} < \frac{1}{q_s^{2+\varepsilon}}.$$

Continuing, define the sets:

$$\begin{aligned}
A_s^1 &= [0, mq_s\alpha(\text{mod } 1));\\
A_s^2 &= [kq_s\alpha(\text{mod } 1), (k+1)q_s\alpha(\text{mod } 1)),\ 0 \le k < m;\\
A_s^3 &= [j\alpha + kq_s\alpha(\text{mod } 1), j\alpha + (k+1)q_s\alpha(\text{mod } 1)),\ 1 \le j < q_s,\ 0 \le k < m.
\end{aligned}$$

Now define the functions

$$\begin{aligned}
\sigma_s(x) &= \begin{cases} 1 & \text{for } x \in A_s^1 \\ 0 & \text{otherwise;} \end{cases}\\
\tau_s(x) &= \begin{cases} k & \text{for } x \in A_s^2 \\ k \oplus 1 & \text{for } x \in A_s^3. \end{cases}
\end{aligned}$$

Now consider the shifted functions $\hat{\sigma_s}(x)$ and $\hat{\tau_s}(x)$ given by

$$\begin{aligned}
\hat{\sigma}_n(x) &= \sigma_n((x - mq_1\alpha - \cdots - mq_{n-1}\alpha)(\text{mod } 1))\\
\hat{\tau}_n(x) &= \tau_n\Big(\Big(x - \sum_{p=1}^{n-1} mq_p\alpha\Big)(\text{mod } 1)\Big).
\end{aligned}$$

Finally, we have the desired functions:

$$\begin{aligned}
\sigma(x) &= \sum_{s=1}^{\infty} \hat{\sigma}_s(x),\\
\tau(x) &= \sum_{s=1}^{\infty} \hat{\tau}_s(x).
\end{aligned}$$

Some (nontrivial) verification shows that these are indeed the appropriate functions, completing the proof. □

5.4 Conclusion

This theorem demonstrates an important difference between homeomorphisms, and more general measure preserving transformations such as interval exchange transformations. For it can be shown that for so-called left group translation transformations associated with some group G, the concepts of ergodicity, minimality, and unique ergodicity all coincide. As we have seen, however, for interval exchange transformations, and thus for measure preserving transformations in general, these notions are independent.

References

[1] Billingsley, P. “Probability and Measure” 3rd Edition, New York: John Wiley and Sons, 1995.

[2] Cornfeld, I.; Fomin, S.; Sinai, Y.; “Ergodic Theory,” New York: Springer-Verlag, 1982.

[3] Gallager, R., “Discrete Stochastic Processes,” Boston MA: Kluwer Academic Publishers, 1996.