ERGODICITY AND MIXING PROPERTIES

ANTHONY QUAS

1. Article Outline

In this article, we discuss ergodicity (a form of irreducibility) and the many kinds of mixing (independence of behaviour in the long term) for measure-preserving transformations. We discuss the partially understood phenomenon of higher-order mixing and indicate some of the contrast between the situation for single measure-preserving transformations and systems of multiple commuting measure-preserving transformations. We include a complete proof in the special case of a continuous map on a compact metric space of the ergodic decomposition by which a measure-preserving transformation can be split into ergodic parts.

2. Glossary

Bernoulli shift: Mathematical abstraction of the scenario in statistics or probability in which one performs repeated independent identical experiments.

Markov chain: A probability model describing a sequence of observations made at regularly spaced time intervals such that at each time, the probability distribution of the subsequent observation depends only on the current observation and not on prior observations.

Measure-preserving transformation: A map from a measure space to itself such that for each measurable subset of the space, it has the same measure as its inverse image under the map.

Measure-theoretic entropy: A non-negative (possibly infinite) real number describing the complexity of a measure-preserving transformation.

Product transformation: Given a pair of measure-preserving transformations: T of X and S of Y, the product transformation is the map of $X \times Y$ given by $(T \times S)(x, y) = (T(x), S(y))$.

3. Definition

Many physical phenomena in equilibrium can be modeled as measure-preserving transformations. Ergodic theory is the abstract study of these transformations, dealing in particular with their long term average behaviour.

One of the basic steps in analysing a measure-preserving transformation is to break it down into its simplest possible components. These simplest components are its ergodic components, and on each of these components, the system enjoys the ergodic property: the long-term time average of any measurement as the system evolves is equal to the average over the component. Ergodic decomposition gives a precise description of the manner in which a system can be split into ergodic components.

A related (stronger) property of a measure-preserving transformation is mixing. Here one is investigating the correlation between the state of the system at different times. The system is mixing if the states are asymptotically independent: as the times between the measurements increase to infinity, the observed values of the measurements at those times become independent.

4. Introduction

The term ergodic was introduced by Boltzmann [8, 9] in his work on statistical mechanics, where he was studying Hamiltonian systems with large numbers of particles. The system is described at any time by a point of *phase space*, a subset of $\mathbb{R}^{6N}$ where N is the number of particles. The configuration describes the 3-dimensional position and velocity of each of the N particles. It has long been known that the Hamiltonian (i.e. the overall energy of the system) is invariant over time in these systems. Thus, given a starting configuration, all future configurations as the system evolves lie on the same *energy surface* as the initial one.

Boltzmann's *ergodic hypothesis* was that the trajectory of the configuration in phase space would fill out the entire energy surface. The term ergodic is thus an amalgamation of the Greek words for work and path. This hypothesis then allowed Boltzmann to conclude that the long-term average of a quantity as the system evolves would be equal to its average value over the phase space.

Subsequently, it was realized that this hypothesis is rarely satisfied. The ergodic hypothesis was replaced in 1911 by the *quasi-ergodic hypothesis* of the Ehrenfests [16] which stated instead that each trajectory is dense in the energy surface, rather than filling out the entire energy surface. The modern notion of ergodicity (to be defined below) is due to Birkhoff and Smith [7]. Koopman [42] suggested studying a measure-preserving transformation by means of the associated isometry on Hilbert space, $U_T\colon L^2(X) \to L^2(X)$ defined by $U_T(f) = f \circ T$. This point of view was used by von Neumann [50] in his proof of the mean ergodic theorem. This was followed closely by Birkhoff [6] proving the pointwise ergodic theorem. An ergodic measure-preserving transformation enjoys the property that Boltzmann first intended to deduce from his hypothesis: that long-term averages of an observable quantity coincide with the integral of that quantity over the phase space.

These theorems allow one to deduce a form of independence on the average: given two sets of configurations A and B, one can consider the volume of the phase space consisting of points that are in A at time 0 and in B at time t. In an ergodic measure-preserving transformation, if one computes the average of the volumes of these regions over time, the ergodic theorems mentioned above allow one to deduce that the limit is simply the product of the volume of A and the volume of B. This is the weakest mixing-type property. In this article, we will outline a rather full range of mixing properties with ergodicity at the weakest end and the Bernoulli property at the strongest end.

We will set out in some detail the various mixing properties, basing our study on a number of concrete examples sitting at various points of this hierarchy. Many of the mixing properties may be characterized in terms of the Koopman operators operators mentioned above (i.e. they are *spectral properties*), but we will see that the strongest mixing properties are not spectral in nature.

We shall also see that there are connections between the range of mixing properties that we discuss and measure-theoretic entropy. In measure-preserving transformations that arise in practice, there is a correlation between strong mixing properties and positive entropy, although many of these properties are logically independent.

One important issue for which many questions remain open is that of higher-order mixing. Here, the issue is if instead of asking that the observations at two times separated by a large time T be approximately independent, one asks whether if one makes observations at more times, each pair suitably separated, the results can be expected to be approximately independent. This issue has an analogue in probability theory, where it is well-known that it is possible to have a collection of random variables that are pairwise independent, but not mutually independent.

5. Basics and Examples

In this article, except where otherwise stated, the measure-preserving transformations that we consider will be defined on probability spaces.

More specifically, given a measurable space $(X, \mathcal{B})$ and a probability measure μ defined on $\mathcal{B}$, a *measure-preserving transformation* of $(X, \mathcal{B}, \mu)$ is a $\mathcal{B}$-measurable map $T\colon X \to X$ such that $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$.

While this definition makes sense for arbitrary measures, not simply probability measures, most of the results and definitions below only make sense in the probability measure case. Sometimes it will be helpful to make the assumption that the underlying probability space is a Lebesgue space (that is, the space together with its completed σ-algebra agrees up to a measure-preserving bijection with the unit interval with Lebesgue measure and the usual σ-algebra of Lebesgue measurable sets). Although this sounds like a strong restriction, in practice it is barely a restriction at all, as almost all of the spaces that appear in the theory (and all of those that appear in this article) turn out to be Lebesgue spaces. For a detailed treatment of the theory of Lebesgue spaces, the reader is referred to Rudolph's book [75]. The reader is referred also to the chapter on Measure Preserving Systems.

While many of the definitions that we shall present are valid for both invertible and non-invertible measure-preserving transformations, the strongest mixing conditions are most useful in the case of invertible transformations.

It will be helpful to present a selection of simple examples, relative to which we will be able to explore ergodicity and the various notions of mixing. These examples and the lemmas necessary to show that they are measure-preserving transformations as claimed may be found in the books of Petersen [63], Rudolph [75] and Walters [91]. More details on these examples can also be found in the chapter on Ergodic Theory: Basic Examples and Constructions.

Example 1 (Rotation on the circle)**.** Let $\alpha \in \mathbb{R}$. Let $R_\alpha\colon [0,1) \to [0,1)$ be defined by $R_\alpha(x) = x + \alpha \bmod 1$. It is straightforward to verify that R_α preserves the restriction of Lebesgue measure λ to $[0,1)$ (it is sufficient to check that $\lambda(R_\alpha^{-1}(J)) = \lambda(J)$ for an interval J)

Example 2 (Doubling Map)**.** Let $M_2\colon [0,1) \to [0,1)$ be defined by $M_2(x) = 2x \bmod 1$. Again, Lebesgue measure is invariant under M_2 (to see this, one observes that for an interval J, $M_2^{-1}(J)$ consists of two intervals, each of half the length of J). This may be generalized in the obvious way to a map M_k for any integer $k \geq 2$.

Example 3 (Interval Exchange Transformation)**.** The class of interval exchange transformations was introduced by Sinai [82]. An interval exchange transformation is the map obtained by cutting the interval into a finite number of pieces and permuting them in such a way that the resulting map is invertible, and restricted to each interval is an order-preserving isometry.

More formally, one takes a sequence of positive lengths $\ell_1, \ell_2, \ldots, \ell_k$ summing to 1 and a permutation π of $\{1, \ldots, k\}$ and defines $a_i = \sum_{j<i} \ell_j$ and $b_i = \sum_{\pi(j)<\pi(i)} \ell_j$ (again with $b_0 = 0$). The interval exchange transformation defined by $(\ell_1, \ldots, \ell_k)$ and π is the map $T\colon [0,1) \to [0,1)$ defined by $T|_{[a_i,a_{i+1})}(x) = x + (b_i - a_i)$. It is straightforward to check that any such interval exchange transformation preserves Lebesgue measure on the unit interval.

Example 4 (Bernoulli Shift)**.** Let A be a finite set and fix a vector $(p_i)_{i\in A}$ of positive numbers that sum to 1. Let $A^{\mathbb{N}}$ denote the set of sequences of the form $x_0x_1x_2\ldots$, where $x_n \in A$ for each $n \in \mathbb{N}$ and let $A^{\mathbb{Z}}$ denote the set of bi-infinite sequences of the form $\ldots x_{-2}x_{-1} \cdot x_0x_1x_2 \ldots$ (the $\cdot$ is a placeholder that allows us to distinguish (for example) between the sequences $\ldots 01010 \cdot 10101 \ldots$ and $\ldots 10101 \cdot 01010 \ldots$).

We define a map (the *shift map*) S on $A^{\mathbb{N}}$ by $(S(x))_n = x_{n+1}$ and define S on $A^{\mathbb{Z}}$ by the same formula. Note that S is invertible as a transformation on $A^{\mathbb{Z}}$ but non-invertible as a transformation on $A^{\mathbb{N}}$.

We need to equip $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ with measures. This is done by defining the measure of a preferred class of sets, checking certain consistency conditions and appealing to the Kolmogorov extension theorem. Here the preferred sets are the *cylinder sets*. Given $m \le n$ in the invertible case and a sequence $a_m \ldots a_n$, we let $[a_m \ldots a_n]_m^n$ denote $\{x \in A^{\mathbb{Z}}\colon x_m = a_m, \ldots, x_n = a_n\}$ and define $\mu([a_m \ldots a_n]_m^n) = p_{a_m}p_{a_{m+1}} \ldots p_{a_n}$. This is then shown to uniquely define a measure μ on the σ-algebra of $A^{\mathbb{Z}}$ generated by the cylinder sets. It is immediate to see that for any cylinder set C, $\mu(S^{-1}C) = \mu(C)$, and it follows that S is a measure-preserving transformation of $(A^{\mathbb{Z}}, \mathcal{B}, \mu)$. The construction is exactly analogous in the non-invertible case. See the chapter on Measure Preserving systems or the books of Walters [91] or Rudolph [75] for more details of defining measures in these systems.

The class of Bernoulli shifts will play a distinguished role in what follows.

Example 5 (Markov Shift)**.** The spaces $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ are exactly as above, as is the shift map. All that changes is the measure.

To define a Markov shift, we need a stochastic matrix P (i.e. a matrix with non-negative entries whose rows sum to 1) with rows and columns indexed by A and a left eigenvector π for P with eigenvalue 1 with the property that the entries of π are non-negative and sum to 1. The existence of such an eigenvector is a consequence of the Perron-Frobenius theory of positive matrices. Provided that the matrix P is irreducible (for each a and a' in A, there is an $n > 0$ such that $P^n_{a,a'} > 0$), the eigenvector π is unique.

Given the pair (P, π), one defines the measure of a cylinder set by $\mu([a_m \ldots a_n]_m^n) = \pi_{a_m} P_{a_m a_{m+1}} \ldots P_{a_{n-1}a_n}$ and extends μ as before to a probability measure on $A^{\mathbb{N}}$ or $A^{\mathbb{Z}}$.

Example 6 (Hard Sphere Gases and Billiards)**.** We wish to model the behaviour of a gas in a bounded region. We make the assumption that the gas consists of a large number N of identical balls which move at constant velocity until two balls collide, whereupon they elastically swap momentum along the direction of contact. The phase space for this system is a region of $\mathbb{R}^{6N}$ (with N 3-dimensional position vectors and N 3-dimensional velocity vectors). More abstractly, the system is equivalent to the motion of a single point particle in a region of $\mathbb{R}^M \times \mathbb{R}^M$ (with the first M-vector representing position and the second representing velocity). The system is constrained in that its position is required to lie in a bounded region S of $\mathbb{R}^M$ with a piecewise smooth boundary. The system evolves by moving the position at a constant rate in the direction of the velocity vector until the point reaches ∂S, at which time the component of the velocity parallel to the normal to ∂S is reversed. This then defines a flow (i.e. a family of maps $(T_t)_{t\in\mathbb{R}}$ satisfying $T_{t+s} = T_t \circ T_s$) on the phase space. Since the magnitude of the velocity is conserved, it is convenient to restrict to flows with speed 1. This system is clearly the closest of the examples that we consider to the situation envisaged by Boltzmann. Perhaps not surprisingly, proofs of even the most basic properties for this system are much harder than the other examples that we consider.

We will need to make use of the concept of measure-theoretic isomorphism. Two measure-preserving transformations T of $(X, \mathcal{B}, \mu)$ and S of $(Y, \mathcal{F}, \nu)$ are *measure-theoretically isomorphic* (or just *isomorphic*) if there exist measurable maps $g\colon X \to Y$ and $h\colon Y \to X$ such that

(1) $g \circ h$ and $h \circ g$ agree with the respective identity maps almost everywhere;
(2) $\mu(g^{-1}F) = \nu(F)$ and $\nu(h^{-1}B) = \mu(B)$ for all $F \in \mathcal{F}$ and $B \in \mathcal{B}$; and
(3) $S \circ g(x) = g \circ T(x)$ for μ-almost every x (or equivalently $T \circ h(y) = h \circ S(y)$ for ν-almost every y).

Measure-theoretic isomorphism is the basic notion of 'sameness' in ergodic theory. It is in some sense quite weak, so that systems may be isomorphic that feel very different (for example, as we discuss later, the time one map of a geodesic flow is isomorphic to a Bernoulli shift). For comparison, the notion of sameness in topological dynamical systems (topological conjugacy) is far stronger.

As an example of measure-theoretic isomorphism, it may be seen that the doubling map is isomorphic to the one-sided Bernoulli shift on $\{0, 1\}$ with $p_0 = p_1 = 1/2$ (the map g takes an $x \in [0, 1)$ to the sequence of 0's and 1's in its binary expansion (choosing the sequence ending with 0's, for example, if x is of the form $p/2^n$) and the inverse map h takes a sequence of 0's and 1's to the point in $[0, 1)$ with that binary expansion.

Given a measure-preserving transformation T of a probability space $(X, \mathcal{B}, \mu)$, T is associated to an isometry of $L^2(X, \mathcal{B}, \mu)$ by $U_T(f) = f \circ T$. This operator is known as the *Koopman Operator*. In the case where T is invertible, the operator U_T is unitary. Two measure-preserving transformations T and S of $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{F}, \nu)$ are *spectrally isomorphic* if there is a Hilbert space isomorphism Θ from $L^2(X, \mathcal{B}, \mu)$ to $L^2(Y, \mathcal{F}, \nu)$ such that $\Theta \circ U_T = U_S \circ \Theta$. As we shall see below, spectral isomorphism is a strictly weaker property than measure-theoretic isomorphism.

Since in ergodic theory, measure-theoretic isomorphism is the basic notion of sameness, all properties that are used to describe measure-preserving systems are required to be invariant under measure-theoretic isomorphism (i.e. if two measure-preserving transformations are measure-theoretically isomorphic, the first has a given property if and only if the second does). On the other hand, we shall see that some mixing-type properties are invariant under spectral isomorphism, while others are not. If a property is invariant under spectral isomorphism, we say that it is a *spectral property.*

There are a number of mixing type properties that occur in the probability literature (α-mixing, β-mixing, ϕ-mixing, ψ-mixing etc) (see Bradley's survey [11] for a description of these conditions). Many of these are stronger than the Bernoulli property, and are therefore not preserved under measure-theoretic isomorphism. For this reason, these properties are not widely used in ergodic theory, although β-mixing turns out to be equivalent to the so-called *weak Bernoulli* property (which turns out to be stronger than the Bernoulli property that we discuss in this article - see Smorodinsky's paper [87]) and α-mixing is equivalent to strong-mixing.

A basic construction (see the article on Ergodic Theory: Basic Examples and Constructions) that we shall require in what follows is the product of a pair of measure-preserving transformations: given transformations T of $(X, \mathcal{B}, \mu)$ and S of $(Y, \mathcal{F}, \nu)$, we define the *product transformation* $T \times S\colon (X \times Y, \mathcal{B} \otimes \mathcal{F}, \mu \times \nu)$ by $(T \times S)(x, y) = (Tx, Sy)$.

One issue that we face on occasion is that it is sometimes convenient to deal with invertible measure-preserving transformations. It turns out that given a non-invertible measure-preserving transformation, there is a natural way to uniquely associate an invertible measure-preserving transformation transformation sharing almost all of the ergodic properties of the original transformation. Specifically, given a non-invertible measure-preserving transformation T of $(X, \mathcal{B}, \mu)$, one lets $\overline{X} = \{(x_0, x_1, \ldots)\colon x_n \in X \text{ and } T(x_n) = x_{n-1} \text{ for all } n\}$, $\overline{\mathcal{B}}$ be the σ-algebra generated by sets of the form $\bar{A}_n = \{\bar{x} \in \overline{X}\colon x_n \in A\}$, $\bar{\mu}(\bar{A}_n) = \mu(A)$ and $\overline{T}(x_0, x_1, \ldots) = (T(x_0), x_0, x_1, \ldots)$. The transformation $\overline{T}$ of $(\overline{X}, \overline{\mathcal{B}}, \overline{\mu})$ is called the *natural extension* of the transformation T of $(X, \mathcal{B}, \mu)$ (see the chapter on Ergodic Theory: Basic Examples and Constructions for more details). In situations where one wants to use invertibility, it is often possible to pass to the natural extension, work there and then derive conclusions about the original non-invertible transformation.

6. Ergodicity

Given a measure-preserving transformation $T\colon X \to X$, if $T^{-1}A = A$, then $T^{-1}A^c = A^c$ also. This allows us to decompose the transformation X into two pieces A and A^c and study the transformation T separately on each. In fact the same situation holds if $T^{-1}A$ and A agree up to a set of measure 0. For this reason, we call a set A *invariant* if $\mu(T^{-1}A \triangle A) = 0$

Returning to Boltzmann's ergodic hypothesis, existence of an invariant set of measure between 0 and 1 would be a bad situation as his essential idea was that the orbit of a single point would 'see' all of X, whereas if X were decomposed in this way, the most that a point in A could see would be all of A, and similarly the most that a point in A^c could see would be all of A^c.

A measure-preserving transformation will be called *ergodic* if it has no non-trivial decomposition of this form. More formally, let T be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. The transformation T is said to be ergodic if for all invariant sets, either the set or its complement has measure 0.

Unlike the remaining concepts that we discuss in this article, this definition of ergodicity applies also to infinite measure-preserving transformations and even to certain non-measure-preserving transformations. See Aaronson's book [1] for more information.

The following lemma is often useful:

Lemma 1. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T\colon X \to X$ be a measure-preserving transformation. Then T is ergodic if and only if the only measurable functions f satisfying $f \circ T = f$ (up to sets of measure 0) are constant almost everywhere.*

For the straightforward proof, we notice that if the condition in the lemma holds and A is an invariant set, then $\mathbf{1}_A \circ T = \mathbf{1}_A$ almost everywhere, so that $\mathbf{1}_A$ is an a.e. constant function and so A or A^c is of measure 0. Conversely, if f is an invariant function, we see that for each α, $\{x\colon f(x) < \alpha\}$ is an invariant set and hence of measure 0 or 1. It follows that f is constant almost everywhere. We remark for future use that it is sufficient to check that the bounded measurable invariant functions are constant.

The following corollary of the lemma shows that ergodicity is a spectral property.

Corollary 2. *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then T is ergodic if and only if 1 is a simple eigenvalue of U_T.*

The ergodic theorems mentioned earlier due to von Neumann and Birkhoff are the following (see also the chapter on Ergodic Theorems).

Theorem 3 (von Neumann Mean Ergodic Theorem [50])**.** *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. For $f \in L^2(X, \mathcal{B}, \mu)$, let $A_N f = 1/N(f + f \circ T + \ldots + f \circ T^{N-1})$. Then for all $f \in L^2(X, \mathcal{B}, \mu)$, $A_N f$ converges in L^2 to an invariant function f^*.*

Theorem 4 (Birkhoff Pointwise Ergodic Theorem [6])**.** *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Let $f \in L^1(X, \mathcal{B}, \mu)$. Let $A_N f$ be as above. Then for μ-almost every $x \in X$, $(A_N f(x))$ is a convergent sequence.*

Of these two theorems, the pointwise ergodic theorem is the deeper result, and it is straightforward to deduce the mean ergodic theorem from the pointwise ergodic theorem. The mean ergodic theorem was reproved very concisely by Riesz [70] and it is this proof that is widely known now. Riesz's proof is reproduced in Parry's book [62]. There have been many different proofs given of the pointwise ergodic theorem. Notable amongst these are the argument due to Garsia [22] and a proof due to Katznelson and Weiss [38] based on work of Kamae [33], which appears in a simplified form in work of Keane and Petersen [40].

If the measure-preserving transformation T is ergodic, then by virtue of Lemma 1, the limit functions appearing in the ergodic theorems are constant. One sees that the constant is simply the integral of f with respect to μ, so that in this situation $A_N f(x)$ converges to $\int f\, d\mu$ in norm and pointwise almost everywhere, thereby providing a justification of Boltzmann's original claim: for ergodic measure-preserving transformations, *time averages agree with spatial averages*. In the case where T is not ergodic, it is also possible to identify the limit in the ergodic theorems: we have $f^* = \mathbb{E}(f|\mathcal{I})$, where $\mathcal{I}$ is the σ-algebra of T-invariant sets.

Note that the set on which the almost everywhere convergence in Birkhoff's theorem takes place depends on the L^1 function f that one is considering. Straightforward considerations show that there is no single full measure set that works simultaneously for all L^1 functions. In the case where X is a compact metric space, it is well known that $C(X)$, the space of continuous functions on X with the uniform norm has a countable dense set, $(f_n)_{n\geq 1}$ say. If the invariant measure μ is ergodic, then for each n, there is a set B_n of measure 1 such that for all $x \in B_n$, $A_N f_n(x) \to \int f_n \, d\mu$. Letting $B = \bigcap_n B_n$, one obtains a full measure set such that for all n and all $x \in B$, $A_N f_n(x) \to \int f_n \, d\mu$. A simple approximation argument then shows that for all $x \in B$ and all $f \in C(X)$, $A_N f(x) \to \int f \, d\mu$. A point x with this property is said to be *generic* for μ. The observations above show that for an ergodic invariant measure μ, we have $\mu\{x \colon x \text{ is generic for } \mu\} = 1$.

If T is ergodic, but T^n is not ergodic for some n, then one can show that the space X splits up as $A_1, \ldots, A_d$ for some $d|n$ in such a way that $T(A_i) = A_{i+1}$ for $i < d$ and $T(A_d) = A_1$ with T^n acting ergodically on each A_i. The transformation T is *totally ergodic* if T^n is ergodic for all $n \in \mathbb{N}$. One can check that a non-invertible transformation T is ergodic if and only if its natural extension is ergodic.

The following lemma gives an alternative characterization of ergodicity, which in particular relates it to mixing.

Lemma 5 (Ergodicity as a Mixing Property)**.** *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then T is ergodic if and only if for all f and g in L^2,*

$$\frac{1}{N} \sum_{n=0}^{N} \langle f, g \circ T^n \rangle \to \langle f, 1 \rangle \langle 1, g \rangle.$$

In particular, if T is ergodic, then $(1/N) \sum_{n=0}^{N-1} \mu(A \cap T^{-n} B) \to \mu(A)\mu(B)$ for all measurable sets A and B.

Proof. Suppose that T is ergodic. Then the left-hand side of the equality is equal to $\langle f, (1/N) \sum_{n=0}^{N-1} g \circ T^n \rangle$. The mean ergodic theorem shows that the second term converges in L^2 to the constant function with value $\int g \, d\mu = \langle g, 1 \rangle$, and the equality follows.

Conversely, if the equation holds for all f and g in L^2, suppose that A is an invariant set. Let $f = g = \mathbf{1}_A$. Then since $g \circ T^n = \mathbf{1}_A$ for all n, the left-hand side is $\langle \mathbf{1}_A, \mathbf{1}_A \rangle = \mu(A)$. On the other hand, the right-hand side is $\mu(A)^2$, so that the equation yields $\mu(A) = \mu(A)^2$, and $\mu(A)$ is either 0 or 1 as required.

Taking $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$ for measurable sets A and B gives the final statement. □

We now examine the ergodicity of the examples presented above. Firstly, for the rotation of the circle, we claim that the transformation is ergodic if and only if the 'angle' α is irrational. To see this, we argue as follows. If $\alpha = p/q$, then we see that $f(x) = e^{2\pi i q x}$ is a non-constant R_α-invariant function, and hence R_α is not ergodic. On the other hand, if α is irrational, suppose f is a bounded measurable invariant function. Since f is bounded, it is an L^2 function, and so f may be expressed in L^2 as a Fourier series: $f = \sum_{n \in \mathbb{Z}} c_n e_n$ where $e_n(x) = e^{2\pi i n x}$. We then see that $f \circ R_\alpha = \sum_{n \in \mathbb{Z}} e^{2\pi i n \alpha} c_n e_n$. In order for f to be equal in L^2 to $f \circ R_\alpha$, they must have the same Fourier coefficients, so that $c_n = e^{2\pi i n \alpha} c_n$ for each n. Since α is irrational, this forces $c_n = 0$ for all $n \neq 0$, so that f is constant as required.

The doubling map and the Bernoulli shift are both ergodic, although we defer proof of this for the time being, since they in fact have the strong-mixing property. A Markov chain with matrix P and vector π is ergodic if and only if for all i and j in A with $\pi_i > 0$ and $\pi_j > 0$, there exists an $n \geq 0$ with $P^n_{ij} > 0$. This follows from the ergodic theorem for Markov chains (which is derived from the Strong Law of

Large Numbers) (see [17] for details). In particular, if the underlying Markov chain is irreducible, then the measure is ergodic.

In the case of interval exchange transformations, there is a simple necessary condition on the permutation for irreducibility, namely for $1 \leq j < k$, we do not have $\pi\{1, \ldots, j\} = \{1, \ldots, j\}$. Under this condition, Masur [47] and Veech [88] independently showed that for almost all values of the sequence of lengths $(\ell_i)_{1 \leq i \leq k}$, the interval exchange transformation is ergodic. (In fact they showed the stronger condition of *unique ergodicity*: that the transformation has no other invariant measure than Lebesgue measure. This implies that Lebesgue measure is ergodic, because if there were a non-trivial invariant set, then the restriction of Lebesgue measure to that set would be another invariant measure).

For the hard sphere systems, there are no results on ergodicity in full generality. Important special cases have been studied by Sinai [85], Sinai and Chernov [86], Krámli, Simányi and Szász [43], Simányi and Szász [81], Simányi [79, 80] and Young [94].

7. Ergodic Decomposition

We already observed that if a transformation is not ergodic, then it may be decomposed into parts. Clearly if these parts are not ergodic, they may be further decomposed. It is natural to ask whether the transformation can be decomposed into ergodic parts, and if so what form does the decomposition take? In fact such a decomposition does exist, but rather than decompose the transformation, it is necessary to decompose the measure into ergodic pieces. This is known as ergodic decomposition.

The set of invariant measures for a measurable map T of a measurable space $(X, \mathcal{B})$ to itself forms a simplex. General functional analytic considerations (due to Choquet [14, 13] - see also Phelps' account [65] of this theory) mean that it is possible to write any member of the simplex as an integral-convex combination of the extreme points. Further, the extreme points of the simplex may be identified as precisely the ergodic invariant measures for T. It follows that any invariant probability measure μ for T may be uniquely expressed in the form

$$\mu(A) = \int_{M_{\mathrm{erg}}(X,T)} \nu(A)\, dm(\nu),$$

where $M_{\mathrm{erg}}(X, T)$ denotes the set of ergodic T-invariant measures on X and m is a measure on $M_{\mathrm{erg}}(X, T)$.

We will give a proof of this theorem in the special case of a continuous transformation of a compact space. Our proof is based on the Birkhoff ergodic theorem and the Riesz Representation Theorem identifying the dual space of the space of continuous functions on a compact space as the set of bounded signed measures on the space (see Rudin's book [74] for details). We include it here because this special case covers many cases that arise in practice, and because few of the standard ergodic theory references include a proof of ergodic decomposition. Exceptions to this are Rudolph's book [75] which gives a full proof in the case that X is a Lebesgue space based on a detailed development of the theory of these spaces and builds measures using conditional expectations. Kalikow's notes [30] give a brief outline of a proof similar to that which follows. Oxtoby [61] also wrote a survey article containing much of the following (and much more besides).

Theorem 6. *Let X be a compact metric space, $\mathcal{B}$ be the Borel σ-algebra, μ be an invariant Borel probability measure and T be a continuous measure-preserving transformation of $(X,\mathcal{B},\mu)$. Then for each $x\in X$, there exists an invariant Borel measure μ_x such that:*

1. *For $f\in L^1(X,\mathcal{B},\mu)$, $\int f\,d\mu=\int\left(\int f\,d\mu_x\right)\,d\mu(x)$;*
2. *Given $f\in L^1(X,\mathcal{B},\mu)$, for μ-almost every $x\in X$, one has $A_Nf(x)\to\int f\,d\mu_x$;*
3. *The measure μ_x is ergodic for μ-almost every $x\in X$.*

Notice that conclusion (2) shows that μ_x can be understood as the distribution on the phase space "seen" if one starts the system in an initial condition of x. This interpretation of the measures μ_x corresponds closely with the ideas of Boltzmann and the Ehrenfests in the formulation of the ergodic and quasi-ergodic hypotheses, which can be seen as demanding that μ_x is equal to μ for (almost) all x.

Proof. The proof will be divided into 3 main steps: defining the measures μ_x, proving measurability with respect to x and proving ergodicity of the measures.

Step 1: Definition of μ_x

Given a function $f\in L^1(X,\mathcal{B},\mu)$, Birkhoff's theorem states that for μ-almost every $x\in X$, $(A_Nf(x))$ is convergent. It will be convenient to denote the limit by $\tilde f(x)$. Let $f_1,f_2,\ldots$ be a sequence of continuous functions that is dense in $C(X)$. For each n, there is a set B_n of measure 1 on which $(A_nf_k(x))_{n=1}^\infty$ is a convergent sequence. Intersecting these gives a set B of full measure such that for $x\in B$, for each $k\geq 1$, $A_nf_k(x)$ is convergent. A simple approximation argument shows that for $x\in B$ and f an arbitrary continuous function, $A_nf(x)$ is convergent. Given $x\in B$, define a map $L_x\colon C(X)\to\mathbb{R}$ by $L_x(f)=\tilde f(x)$. This is a continuous linear functional on $C(X)$, and hence by the Riesz Representation Theorem there exists a Borel measure μ_x such that $\tilde f(x)=\int f\,d\mu_x$ for each $f\in C(X)$ and $x\in B$. Since $L_x(f)\geq 0$ when f is a non-negative function and $L_x(1)=1$, the measure μ_x is a probability measure. Since $L_x(f\circ T)=L_x(f)$ for $f\in C(X)$, one can check that μ_x must be an invariant probability measure. For $x\notin B$, simply define $\mu_x=\mu$. Since B^c is a set of measure 0, this will not affect any of the statements that we are trying to prove.

Now for f continuous, we have A_Nf is a bounded sequence of functions with $A_Nf(x)$ converging to $\int f\,d\mu_x$ almost everywhere and $\int A_Nf\,d\mu=\int f\,d\mu$ since T is measure-preserving. It follows from the bounded convergence theorem that for $f\in C(X)$,

$$\int f\,d\mu=\int\left(\int f\,d\mu_x\right)\,d\mu(x). \tag{1}$$

Step 2: Measurability of $x\mapsto\mu_x(A)$

Lemma 7. *Let $C\in\mathcal{B}$ satisfy $\mu(C)=0$. Then $\mu_x(C)=0$ for μ-almost every $x\in X$.*

Proof. Using regularity of Borel probability measures (see Rudin's book [74] for details), there exist open sets $U_1\supset U_2\supset\cdots\supset C$ with $\mu(U_k)<1/k$. There exist continuous functions $g_{k,m}$ with $(g_{k,m}(x))_{m=1}^\infty$ increasing to $\mathbf{1}_{U_k}$ everywhere (e.g. $g_{k,m}(x)=\min(1,m\cdot d(x,U_k^c))$). By (1), we have $\int(\int g_{k,m}\,d\mu_x)\,d\mu(x)<1/k$ for all k,m. Note that $\int g_{k,m}\,d\mu_x=\lim_{n\to\infty}A_ng_{k,m}(x)$ is a measurable function of x, so that using the monotone convergence theorem (taking the limit in m), $x \mapsto \int \mathbf{1}_{U_k}\, d\mu_x = \mu_x(U_k)$ is measurable and $\int(\int \mathbf{1}_{U_k}\, d\mu_x)\, d\mu(x) \leq 1/k$. We now see that $x \mapsto \lim_{k\to\infty} \mu_x(U_k) = \mu_x(\bigcap U_k)$ is also measurable, and by monotone convergence we see $\int \mu_x\left(\bigcap U_k\right)\, d\mu(x) = 0$. It follows that $\mu_x(\bigcap U_k) = 0$ for μ-almost every x. Since $\bigcap U_k \supset C$, the lemma follows. □

Given a set $A \in \mathcal{B}$, let f_k be a sequence of continuous functions (uniformly bounded by 1) satisfying $\|f_k - \mathbf{1}_A\|_{L^1(\mu)} < 2^{-n}$, so that in particular $f_k(x) \to \mathbf{1}_A(x)$ for μ-almost every x. For each k, $x \mapsto \int f_k\, d\mu_x = \lim_{n\to\infty} A_n f_k(x)$ is a measurable function. By Lemma 7, for μ-almost every x, $f_k \to 1_A$ μ_x-almost everywhere, so that by the bounded convergence theorem $\lim_{k\to\infty} \int f_k\, d\mu_x = \mu_x(A)$ for μ-almost every x. Since the limit of measurable functions is measurable, it follows that $x \mapsto \mu_x(A)$ is measurable for any measurable set $A \in \mathcal{B}$.

This allows us to define a measure ν by $\nu(A) = \int \mu_x(A)\, d\mu(x)$. For a bounded measurable function f, we have $\int f\, d\nu = \int(\int f\, d\mu_x)\, d\mu(x)$. Since this agrees with $\int f\, d\mu$ for continuous functions by (1), it follows that $\mu = \nu$. Conclusion (1) of the theorem now follows easily.

Given $f \in L^1(X)$, we let (f_k) be a sequence of continuous functions such that $\|f_k - f\|_{L^1(\mu)}$ is summable. This implies that $\|f_k - f\|_{L^1(\mu_x)}$ is summable for μ-almost every x and in particular, $\int f_k\, d\mu_x \to \int f\, d\mu_x$ for almost every x. On the other hand, by the remark following the statement of Birkhoff's theorem, we have $\tilde{f}_k = \mathbb{E}(f_k|\mathcal{I})$ so that $\|\tilde{f} - \tilde{f}_k\|_{L^1(\mu)}$ is summable and $\tilde{f}_k(x) \to \tilde{f}(x)$ for μ-almost every x. Combining these two statements, we see that for μ-almost every x, we have

$$\tilde{f}(x) = \lim_{k\to\infty} \tilde{f}_k(x) = \lim_{k\to\infty} \int f_k\, d\mu_x = \int f\, d\mu_x.$$

This establishes conclusion (2) of the theorem.

Step 3: Ergodicity of μ_x

We have shown how to disintegrate the invariant measure μ as an integral combination of μ_x's, and we have interpreted the μ_x's as describing the average behaviour starting from x. It remains to show that the μ_x's are ergodic measures.

Fix for now a continuous function f and a number $0 < \epsilon < 1$. Since $A_n f(x) \to \tilde{f}(x)$ μ-almost everywhere, there exists an N such that $\mu\{x\colon |A_N f(x) - \tilde{f}(x)| > \epsilon/2\} < \epsilon^3/8$.

We now claim the following:

$$\mu\{x\colon \mu_x\{y\colon |\tilde{f}(y) - \textstyle\int f\, d\mu_x| > \epsilon\} > \epsilon\} < \epsilon. \tag{2}$$

To see this, note that $\{y\colon |\tilde{f}(y) - \int f\, d\mu_x| > \epsilon\} \subset \{y\colon |\tilde{f}(y) - A_N f(y)| > \epsilon/2\} \cup \{y\colon |A_N f(y) - \tilde{f}(x)| > \epsilon/2\}$, so that if $\mu_x\{y\colon |\tilde{f}(y) - \int f\, d\mu_x| > \epsilon\} > \epsilon$, then either $\mu_x\{y\colon |\tilde{f}(y) - A_N f(y)| > \epsilon/2\} > \epsilon/2$ or $\mu_x\{y\colon |A_N f(y) - \tilde{f}(x)| > \epsilon/2\} > \epsilon/2$. We show that the set of x's satisfying each condition is small.

Firstly, we have $\epsilon^3/8 > \mu\{y\colon |\tilde{f}(y) - A_N f(y)| > \epsilon/2\} = \int \mu_x\{y\colon |\tilde{f}(y) - A_N f(y)| > \epsilon/2\}\, d\mu(x)$, so that $\mu\{x\colon \mu_x\{y\colon |\tilde{f}(y) - A_N f(y)| > \epsilon/2\} > \epsilon/2\} < \epsilon^2/4 < \epsilon/2$.

For the second term, given $c \in \mathbb{R}$, let $F_c(x) = |A_N f(x) - c|$ and $G(x) = F_{\tilde{f}(x)}(x)$. Note that

$$\int F_{\tilde{f}(x)}(y)\, d\mu_x(y) = \lim_{n\to\infty} A_n F_{\tilde{f}(x)}(x) = \lim_{n\to\infty} A_n G(x)$$

(using the facts that $y \mapsto F_{\tilde{f}(x)}(y)$ is a continuous function and that since $\tilde{f}(x)$ is an invariant function, $F_{\tilde{f}(x)}(T^k x) = G(T^k x))$. Since $\int G(x)\,d\mu(x) < \epsilon^3/8$, it follows that $\int F_{\tilde{f}(x)}(y)\,d\mu_x(y) \le \epsilon^2/4$ except on a set of x's of measure less than $\epsilon/2$. Outside this bad set, we have $\mu_x\{y\colon |A_N f(y) - \tilde{f}(x)| > \epsilon/2\} < \epsilon/2$ so that $\mu\{x\colon \mu_x\{y\colon |A_N f(y) - \tilde{f}(x)| > \epsilon/2\} > \epsilon/2\} < \epsilon/2$ as required.

This establishes our claim (2) above. Since $\epsilon > 0$ is arbitrary, it follows that for each $f \in C(X)$, for μ-almost every x, μ_x-almost every y satisfies $\tilde{f}(y) = \int f\,d\mu_x$. As usual, taking a countable dense sequence (f_k) in $C(X)$, it is the case that for all k and μ-almost every x, $\tilde{f}_k(y) = \int f_k\,d\mu_x$ μ_x-almost everywhere. Let the set of x's with this property be D. We claim that for $x \in D$, μ_x is ergodic. Suppose not. Then let $x \in D$ and let J be an invariant set of μ_x measure between δ and $1-\delta$ for some $\delta > 0$. Then by density of $C(X)$ in $L^1(\mu_x)$, there exists an f_k with $\|f_k - \mathbf{1}_J\|_{L^1(\mu_x)} < \delta$. Since $\mathbf{1}_J$ is an invariant function, we have $\tilde{\mathbf{1}}_J = \mathbf{1}_J$. On the other hand, $\tilde{f}_k$ is a constant function. It follows that $\|\tilde{f}_k - \tilde{\mathbf{1}}_J\|_{L^1(\mu_x)} \ge \delta > \|f_k - \mathbf{1}_J\|_{L_1(\mu_x)}$. This contradicts the identification of the limit as a conditional expectation and concludes the proof of the theorem.

□

8. Mixing

As mentioned above, ergodicity may be seen as an independence on average property. More specifically, one wants to know whether in some sense $\mu(A \cap T^{-n}B)$ converges to $\mu(A)\mu(B)$ as $n \to \infty$. Ergodicity is the property that there is convergence in the Césaro sense. Weak-mixing is the property that there is convergence in the strong Césaro sense. That is, a measure-preserving transformation T is *weak-mixing* if

$$\frac{1}{N}\sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \to 0 \text{ as } N \to \infty.$$

In order for T to be *strong-mixing*, we require simply $\mu(A \cap T^{-N}B) \to \mu(A)\mu(B)$ as $N \to \infty$. It is clear that strong-mixing implies weak-mixing and weak-mixing implies ergodicity.

If T^d is not ergodic (so that $T^{-d}A = A$ for some A of measure strictly between 0 and 1), then $|\mu(T^{-nd}A \cap A) - \mu(A)^2| = \mu(A)(1-\mu(A))$, so that T is not weak-mixing.

An alternative characterization of weak-mixing is as follows:

Lemma 8. *The measure-preserving transformation T is weak-mixing if and only if for every pair of measurable sets A and B, there exists a subset J of $\mathbb{N}$ of density 1 (i.e. $\#(J \cap \{1,\ldots,N\})/N \to 1$) such that*

$$\lim_{n\to\infty\, n\notin J} \mu(A \cap T^{-n}B) = \mu(A)\mu(B). \tag{3}$$

By taking a countable family of measurable sets that are dense (with respect to the metric $d(A,B) = \mu(A \triangle B)$) and taking a suitable intersection of the corresponding J sets, one shows that for a given weak-mixing measure-preserving transformation, there is a single set $J \subset \mathbb{N}$ such that (3) holds for *all* measurable sets A and B (see Petersen [63] or Walters [91] for a proof).

We show that an irrational rotation of the circle is not weak-mixing as follows: let $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ and let A be the interval $[\frac{1}{4}, \frac{3}{4})$. There is a positive proportion of n's in the natural numbers (in fact proportion 1/3) with the property that $|T^n(\frac{1}{2}) - \frac{1}{2}| < \frac{1}{6}$. For these n's $\mu(A \cap T^{-n}A) > \frac{1}{3}$, so that in particular $|\mu(A \cap T^{-n}A) - \mu(A)\mu(A)| > \frac{1}{12}$. Clearly this precludes the required convergence to 0 in the definition of weak-mixing, so that an irrational rotation is ergodic but not weak-mixing. Since $R_\alpha^n = R_{n\alpha}$, the earlier argument shows that R_α^n is ergodic, so that R_α is totally ergodic.

On the other hand, we show that any Bernoulli shift is strong-mixing. To see this, let A and B be arbitrary measurable sets. By standard measure-theoretic arguments, A and B may each be approximated arbitrarily closely by a finite union of cylinder sets. Since if A' and B' are finite unions of cylinder sets, we have that $\mu(A' \cap T^{-n}B')$ is *equal* to $\mu(A')\mu(B')$ for large n, it is easy to deduce that $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ as required. Since the doubling map is measure-theoretically isomorphic to a one-sided Bernoulli shift, it follows that the doubling map is also strong-mixing.

Similarly, if a Markov Chain is irreducible (i.e. for any states i and j, there exists an $n \geq 0$ such that $P_{ij}^n > 0$) and aperiodic (there is a state i such that $\gcd\{n \colon P_{ii}^n > 0\} = 1$), then given any pair of cylinder sets A' and B', we have by standard theorems of Markov chains $\mu(A' \cap T^{-n}B') \to \mu(A')\mu(B')$. The same argument as above then shows that an aperiodic irreducible Markov Chain is strong-mixing. On the other hand, if a Markov chain is periodic ($d = \gcd\{n \colon P_{ii}^n > 0\} > 0$), then letting $A = B = \{x \colon x_0 = i\}$, we have that $\mu(A \cap T^{-n}B) = 0$ whenever $d \nmid n$. It follows T^d is not ergodic, so that T is not weak-mixing.

Both weak- and strong-mixing have formulations in terms of functions:

Lemma 9. *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$.*

(1) *T is weak-mixing if and only if for every $f, g \in L^2$ one has*

$$\frac{1}{N} \sum_{n=0}^{N-1} |\langle f, g \circ T^n \rangle - \langle f, 1 \rangle \langle 1, g \rangle| \to 0 \text{ as } N \to \infty.$$

(2) *T is strong-mixing if and only if for every $f, g \in L^2$, one has*

$$\langle f, g \circ T^N \rangle \to \langle f, 1 \rangle \langle 1, g \rangle \text{ as } N \to \infty.$$

Using this, one can see that both mixing conditions are spectral properties.

Lemma 10. *Weak- and strong-mixing are spectral properties.*

Proof. Suppose S is a weak-mixing transformation of $(Y, \mathcal{F}, \nu)$ and the transformation T of $(X, \mathcal{B}, \mu)$ is spectrally isomorphic to S by the Hilbert space isomorphism Θ. Then for $f, g \in L^2(X, \mathcal{B}, \mu)$, $\langle f, g \circ T^n \rangle_X - \langle f, 1 \rangle_X \langle 1, g \rangle_X = \langle \Theta(f), \Theta(g) \circ S^n \rangle_Y - \langle \Theta(f), \Theta(1) \rangle_Y \langle \Theta(1), \Theta(g) \rangle_Y$. Since 1 is an eigenfunction of U_T with eigenvalue 1, $\Theta(1)$ is an eigenfunction of U_S with an eigenvalue 1, so since S is ergodic, $\Theta(1)$ must be a constant function. Since Θ preserves norms, $\Theta(1)$ must have a constant value of absolute value 1 and hence $\langle f, g \circ T^n \rangle_X - \langle f, 1 \rangle_X \langle 1, g \rangle_X = \langle \Theta(f), \Theta(g) \circ S^n \rangle_Y - \langle \Theta(f), 1 \rangle_Y \langle 1, \Theta(g) \rangle_Y$. It follows from Lemma 9 that T is weak-mixing.

A similar proof shows that strong-mixing is a spectral property. □

Both weak- and strong-mixing properties are preserved by taking natural extensions.

Recent work of Avila and Forni [4] shows that for interval exchange transformations of $k \geq 3$ intervals with the underlying permutation satisfying the non-degeneracy condition above, almost all divisions of the interval (with respect to Lebesgue measure on the $k-1$-dimensional simplex) lead to weak-mixing transformations. On the other hand, work of Katok [34] shows that no interval exchange transformation is strong-mixing.

It is of interest to understand the behaviour of the 'typical' measure-preserving transformation. There are a number of Baire category results addressing this. In order to state them, one needs a set of measure-preserving transformations and a topology on them. As mentioned earlier, it is effectively no restriction to assume that a transformation is a Lebesgue-measurable map on the unit interval preserving Lebesgue measure. The classical category results are then on the collection of invertible Lebesgue-measure preserving transformations of the unit interval. One topology on these is the 'weak' topology, where a sub-base is given by sets of the form $N(T, A, \epsilon) = \{S \colon \lambda(S(A) \triangle T(A)) < \epsilon\}$. With respect to this topology, Halmos [25] showed that a residual set (i.e. a dense G_δ set) of invertible measure-preserving transformations is weak-mixing (see also work of Alpern [3]), while Rokhlin [71] showed that the set of strong-mixing transformations is meagre (i.e. a nowhere dense F_σ set), allowing one to conclude that with respect to this topology, the typical transformation is weak- but not strong-mixing.

As often happens in these cases, even when a certain kind of behaviour is typical, it may not be simple to exhibit concrete examples. In this case, a well-known example of a transformation that is weak-mixing but not strong-mixing was given by Chacon [12].

While on the face of it the formulation of weak-mixing is considerably less natural than that of strong-mixing, the notion of weak-mixing turns out to be extremely natural from a spectral point of view. Given a measure-preserving transformation T, let U_T be the Koopman operator described above. Since this operator is an isometry, any eigenvalue must lie on the unit circle. The constant function 1 is always an eigenfunction with eigenvalue 1. If T is ergodic and g and h are eigenfunctions of U_T with eigenvalue λ, then $g\bar{h}$ is an eigenfunction with eigenvalue 1, hence invariant, so that $g = Kh$ for some constant K. We see that for ergodic transformations, up to rescaling, there is at most one eigenfunction with any given eigenvalue.

If U_T has a non-constant eigenfunction f, then one has $|\langle U_T^n f, f\rangle| = \|f\|^2$ for each n, whereas by Cauchy-Schwartz, $|\langle f, 1\rangle|^2 < \|f\|^2$. It follows that $|\langle U_T^n f, f\rangle - \langle f, 1\rangle\langle 1, f\rangle| \geq c$ for some positive constant c, so that using Lemma 9, T is not weak-mixing.

Using the spectral theorem, the converse is shown to hold.

Theorem 11. *The measure-preserving transformation T is weak-mixing if and only U_T has no non-constant eigenfunctions.*

Of course this also shows that weak-mixing is a spectral property. Equivalently, this says that the transformation T is weak-mixing if and only if the apart from the constant eigenfunction, the operator U_T has only continuous spectrum (that is, the operator has no other eigenfunctions). For a very nice and concise development of the part of spectral theory relevant to ergodic theory, the reader is referred to the Appendix in Parry's book [62]. See also the chapter on Spectral Properties.

Using this theory, one can establish the following:

Theorem 12.

(1) *T is weak-mixing if and only if $T \times T$ is ergodic;*
(2) *If T and S are ergodic, then $T \times S$ is ergodic if and only if U_S and U_T have no common eigenvalues other than 1.*

Proof. The main factor in the proof is that the eigenvalues of $U_{T\times S}$ are precisely the set of $\alpha\beta$, where α is an eigenvalue of U_T and β is an eigenvalue of U_S. Further, the eigenfunctions of $U_{T\times S}$ with eigenvalue γ are spanned by eigenfunctions of the form $f \otimes g$, where f is an eigenfunction of U_T, g is an eigenfunction of U_S, and the product of the eigenvalues is γ.

Suppose that T is weak-mixing. Then the only eigenfunction is the constant function, so that the only eigenfunction of $U_{T\times T}$ is the constant function, proving that $T \times T$ is ergodic. Conversely, if U_T has an eigenvalue (so that $f \circ T = \alpha T$ for some non-constant f) then $f \otimes \bar{f}$ is a non-constant invariant function of $T \times T$ so that $T \times T$ is not ergodic.

For the second part, if U_S and U_T have a common eigenvalue other than 1 (say $f \circ T = \alpha f$ and $g \circ T = \alpha g$), then $f \otimes \bar{g}$ is a non-constant invariant function. Conversely, if $T \times S$ has a non-constant invariant function h, then h can be decomposed into functions of the form $f \otimes g$, where f and g are eigenfunctions of U_T and U_S respectively with eigenvalues α and β satisfying $\alpha\beta = 1$. Since the eigenvalues of S are closed under complex conjugation, we see that U_T and U_S have a common eigenvalue other than 1 as required. □

For a measure-preserving transformation T, we let K be the subspace of L^2 spanned by the eigenfunctions of U_T. It is a remarkable fact that K may be identified as $L^2(X, \mathcal{B}', \mu)$ where $\mathcal{B}'$ is a sub-σ-algebra of $\mathcal{B}$. The space K is called the *Kronecker factor* of T. The terminology comes from the fact that any sub-σ-algebra $\mathcal{F}$ of $\mathcal{B}$ gives rise to a factor mapping $\pi\colon (X, \mathcal{B}, \mu) \to (X, \mathcal{F}, \mu)$ with $\pi(x) = x$. By construction $L^2(X, \mathcal{B}', \mu)$ is the closed linear span of the eigenfunctions of T considered as a measure-preserving transformation of $(X, \mathcal{B}', \mu)$. By the Discrete Spectrum Theorem of Halmos and von Neumann [24], T acting on $(X, \mathcal{B}', \mu)$ is measure-theoretically isomorphic to a rotation on a compact group. This allows one to split $L^2(X, \mathcal{B}, \mu)$ as $L^2(X, \mathcal{B}', \mu) \oplus L^2_c(X, \mathcal{B}, \mu)$, where, as mentioned above the first part is the discrete spectrum part, spanned by eigenfunctions, and the second part is the continuous spectrum part, consisting of functions whose spectral measure is continuous. Since we have split L^2 into a discrete part and a continuous part, it is natural to ask whether the underlying transformation T can be split up in some way into a weak-mixing part and a discrete spectrum (compact group rotation) part, somewhat analogously to the ergodic decomposition. Unfortunately, there is no-such decomposition available. However for some applications, for example to multiple recurrence (starting with the work of Furstenberg [19, 20]), the decomposition of L^2 (possibly into more complicated parts) plays a crucial role (see the chapters on Ergodic Theory: Recurrence and Ergodic Theory: Interactions with Combinatorics and Number Theory).

For non-invertible measure-preserving transformations, the transformation is weak- or strong-mixing if and only if its natural extension has that property.

The understanding of weak-mixing in terms of the discrete part of the spectrum of the operator also extends to total ergodicity. T^n is ergodic if and only if T has no eigenvalues of the form $e^{2\pi ip/n}$ other than 1. From this it follows that an ergodic measure-preserving transformation T is totally ergodic if and only if it has no *rational spectrum* (i.e. no eigenvalues of the form $e^{2\pi ip/q}$ other than the simple eigenvalue 1).

An intermediate mixing condition between strong- and weak- mixing is that a measure-preserving transformation is *mild-mixing* if whenever $f \circ T^{n_i} \to f$ for an L^2 function f and a sequence $n_i \to \infty$, then f is a.e. constant. Clearly mild-mixing is a spectral property. If a transformation has an eigenfunction f, then it is straightforward to find a sequence n_i such that $f \circ T^{n_i} \to f$, so we see that mild-mixing implies weak-mixing. To see that strong-mixing implies mild-mixing, suppose that T is strong-mixing and that $f \circ T^{n_i} \to f$. Then we have $\int f \circ T^{n_i} \bar{f} \to \|f\|^2$. On the other hand, the strong mixing property implies that $\int f \circ T^{n_i} \bar{f} \to |\langle f, 1\rangle|^2$. The equality of these implies that f is a.e. constant. Mild-mixing has a useful reformulation in terms of ergodicity of general (not necessarily probability) measure-preserving transformations: A transformation T is mild-mixing if and only if for every conservative ergodic measure-preserving transformation S, $T \times S$ is ergodic. See Furstenberg and Weiss' article [21] for further information on mild-mixing.

The strongest spectral property that we consider is that of having countable Lebesgue spectrum. While we will avoid a detailed discussion of spectral theory in this article, this is a special case that can be described simply. Specifically, let T be an invertible measure-preserving transformation. Then T has countable Lebesgue spectrum if there is a sequence of functions $f_1, f_2, \ldots$ such that $\{1\} \cup \{U_T^n f_j \colon n \in \mathbb{Z}, j \in \mathbb{N}\}$ forms an orthonormal basis for $L^2(X)$.

To see that this property is stronger than strong-mixing, we simply observe that it implies that $\langle U_T^t U_T^n f_j, U_T^m f_k\rangle \to 0$ as $t \to \infty$. Then by approximating f and g by their expansions with respect to a finite part of the basis, we deduce that $\langle U_T^n f, g\rangle \to \langle f, 1\rangle\langle 1, g\rangle$ as required. Since already strong-mixing is atypical from the topological point of view, it follows that countable Lebesgue spectrum has to be atypical. In fact, Yuzvinskii [95] showed that the typical invertible measure-preserving transformation has simple singular spectrum.

The property of countable Lebesgue spectrum is by definition a spectral property. Since it completely describes the transformation up to spectral isomorphism, there can be no stronger spectral properties. The remaining properties that we shall examine are invariant under measure-theoretic isomorphisms only.

An invertible measure-preserving transformation T of $(X, \mathcal{B}, \mu)$ is said to be K (for Kolmogorov) if there is a sub-σ-algebra $\mathcal{F}$ of $\mathcal{B}$ such that

(1) $\bigcap_{n=1}^{\infty} T^{-n}\mathcal{F}$ is the trivial σ-algebra up to sets of measure 0 (i.e. the intersection consists only of null sets and sets of full measure).
(2) $\bigvee_{n=1}^{\infty} T^n\mathcal{F} = \mathcal{B}$ (i.e. the smallest σ-algebra containing $T^n\mathcal{F}$ for all $n > 0$ is $\mathcal{B}$.

The K property has a useful reformulation in terms of entropy as follows: T is K if and only if for every non-trivial partition $\mathcal{P}$ of X, the entropy of T with respect to the partition $\mathcal{P}$ is positive: T has *completely positive entropy*. See the chapter on Entropy in Ergodic Theory for the relevant definitions. The equivalence of the K property and completely positive entropy was shown by Rokhlin and Sinai [73]. For a general transformation T, one can consider the collection of all subsets B of X such that with respect to the partition $\mathcal{P}_B = \{B, B^c\}$, $h(\mathcal{P}_B) = 0$. One can show that this is a σ-algebra. This σ-algebra is known as the *Pinkser* σ-algebra.

The above reformulation allows us to say that a transformation is K if and only if it has a trivial Pinsker *σ-algebra*.

The K property implies countable Lebesgue spectrum (see Parry's book [62] for a proof). To see that K is not implied by countable Lebesgue spectrum, we point out that certain transformations derived from Gaussian systems (see for example the paper of Newton and Parry [51]) have countable Lebesgue spectrum but zero entropy.

The fact that (two-sided) Bernoulli shifts have the K property follows from Kolmogorov's 0–1 law by taking $\mathcal{F} = \bigvee_{n=0}^{\infty} T^{-n}\mathcal{P}$, where $\mathcal{P}$ is the partition into cylinder sets (see Williams's book [92] for details of the 0–1 law).

Although the K property is explicitly an invertible property, it has a non-invertible counterpart, namely exactness. A transformation T of $(X, \mathcal{B}, \mu)$ is *exact* if $\bigcap_{n=0}^{\infty} T^{-n}\mathcal{B}$ consists entirely of null sets and sets of measure 1. It is not hard to see that a non-invertible transformation is exact if and only if its natural extension is K.

The final and strongest property in our list is that of being measure-theoretically isomorphic to a Bernoulli shift. If T is measure-theoretically isomorphic to a Bernoulli shift, we say that T has the *Bernoulli property.* While in principle this could apply to both invertible and non-invertible transformations, in practice the definition applies to a large class of invertible transformations, but occurs comparatively seldom for non-invertible transformations. For this reason, we will restrict ourselves to a discussion of the Bernoulli property for invertible transformations (see however work of Hoffman and Rudolph [27] and Heicklen and Hoffman [26] for work on the one-sided Bernoulli property).

In the case of invertible Bernoulli shifts, Ornstein [52, 56] developed in the early 1970s a powerful isomorphism theory, showing that two Bernoulli shifts are measure-theoretically isomorphic if and only if they have the same entropy. Entropy had already been identified as an invariant by Kolmogorov and Sinai [41, 83], so this established that it was a complete invariant for Bernoulli shifts. Keane and Smorodinsky [39] gave a proof which showed that two Bernoulli shifts of the same entropy are isomorphic using a conjugating map that is continuous almost everywhere. With other authors, this theory was extended to show that the property of being isomorphic to a Bernoulli shift applied to a surprisingly large class of measure-preserving transformations (e.g. geodesic flows on manifolds of constant negative curvature (Ornstein and Weiss [58]), aperiodic irreducible Markov chains (Friedman and Ornstein [18]), toral automorphisms (Katznelson [37]) and more generally many Gibbs measures for hyperbolic dynamical systems (see the book of Bowen [10])).

Initially, it was conjectured that the properties of being K and Bernoulli were the same, but since then a number of measure-preserving transformations that are K but not Bernoulli have been identified. The earliest was due to Ornstein [53]. Ornstein and Shields [57] then provided an uncountable family of non-isomorphic K automorphisms. Katok [35] gave an example of a smooth diffeomorphism that is K but not Bernoulli; and Kalikow [31] gave a very natural probabilistic example of a transformation that has this property (the T, T^{-1} process).

While in systems that one regularly encounters there is a correlation between positive entropy and the stronger mixing properties that we have discussed, these properties are logically independent (for example taking the product of a Bernoulli shift and the identity transformation gives a positive entropy transformation that fails to be ergodic; also, the zero entropy Gaussian systems with countable Lebesgue spectrum mentioned above have relatively strong mixing properties but zero entropy).

In many of the mixing criteria discussed above we have considered a pair of sets A and B and asked for asymptotic independence of A and B (so that for large n, A and $T^{-n}B$ become independent). It is natural to ask, given a finite collection of sets $A_0, A_1, \ldots A_k$, under what conditions $\mu(A_0 \cap T^{-n_1}A_1 \cap \ldots \cap T^{-n_k}A_k)$ converges to $\prod_{j=0}^k \mu(A_j)$.

A measure-preserving transformation is said to be *mixing of order* $k+1$ if for all measurable sets $A_0, \ldots, A_k$,

$$\lim_{n_1\to\infty, n_{j+1}-n_j\to\infty} \mu(A_0 \cap T^{-n_1}A_1 \cap \ldots \cap T^{-n_k}A_k) = \prod_{j=0}^k \mu(A_j).$$

An outstanding open question asked by Rokhlin [72] appearing already in Halmos' 1956 book [24] is to determine whether mixing (i.e. mixing of order 2) implies mixing of all orders. Kalikow [32] showed that mixing implies mixing of all orders for rank 1 transformations (existence of rank one mixing transformations having been previously established by Ornstein in [59]). Later Ryzhikov [77] used joining methods to establish the result for transformations with finite rank, and Host [28] also used joining methods to establish the result for measure-preserving transformations with singular spectrum, but the general question remains open.

It is not hard to show using martingale arguments that K automorphisms and hence all Bernoulli measure-preserving transformations are mixing of all orders.

For weak-mixing transformations, Furstenberg [20] has established the following *weak-mixing of all orders* statement: if a measure-preserving transformation T is weak-mixing, then given sets $A_0, \ldots, A_k$, there is a subsequence J of the integers of density 0 such that

$$\lim_{n\to\infty\, n\notin J} \mu(A_0 \cap T^{-n}A_1 \cap \ldots \cap T^{-kn}A_k) = \prod_{i=0}^k \mu(A_i).$$

Bergelson [5] generalized this by showing that

$$\lim_{n\to\infty, n\notin J} \mu(A_0 \cap T^{-p_1(n)}A_1 \cap \ldots \cap T^{-p_k(n)}A_k) = \prod_{i=0}^k \mu(A_i)$$

whenever $p_1(n), \ldots, p_k(n)$ are non-constant integer-valued polynomials such that $p_i(n) - p_j(n)$ is unbounded for $i \neq j$. The method of proof of both of these results was a Hilbert space version of the van der Corput inequality of analytic number theory. Furstenberg's proof played a key role in his ergodic proof [19] of Szemerédi's theorem on the existence of arbitrarily long arithmetic progressions in a subset of the integers of positive density (see the chapter on Ergodic Theory: Interactions with Combinatorics and Number Theory for more information about this direction of study)

The conclusions that one draws here are much weaker than the requirement for mixing of all orders. For mixing of all orders, it was required that provided the gaps between $0, n_1, \ldots, n_k$ diverge to infinity, one achieves asymptotic independence, whereas for these weak-mixing results, the gaps are increasing along prescribed sequences with regular growth properties.

It is interesting to note that the analogous question of whether mixing implies mixing of all orders is known to fail in higher-dimensional actions. Here, rather than a $\mathbb{Z}$ action, in which there is a single measure-preserving transformation (so that the integer n acts on a point $x \in X$ by mapping it to $T^n x$), one takes a $\mathbb{Z}^d$ action. For such an action, one has d commuting transformations $T_1, \ldots, T_d$ and a vector $(n_1, \ldots, n_d)$ acts on a point x by sending it to $T_1^{n_1} \cdots T_d^{n_d} x$. Ledrappier [44] studied the following two-dimensional action. Let $X = \{x \in \{0,1\}^{\mathbb{Z}^2} : x_{\mathbf{v}} + x_{\mathbf{v}+\mathbf{e}_1} + x_{\mathbf{v}+\mathbf{e}_2} = 0 \pmod 2\}$ and let $T_i(x)_{\mathbf{v}} = x_{\mathbf{v}+\mathbf{e}_i}$. Since X is a compact Abelian group, it has a natural measure μ invariant under the group operations (the Haar measure). It is not hard to show that this system is mixing (i.e. given any measurable sets A and B, $\mu(A \cap T_1^{-n_1} T_2^{-n_2} B) \to \mu(A)\mu(B)$ as $\|(n_1, n_2)\| \to \infty$). Ledrappier showed that the system fails to be 3-mixing. Subsequently Masser [46] established necessary and sufficient conditions for similar higher-dimensional algebraic actions to be mixing of order k but not order $k+1$ for any given k.

9. Hyperbolicity and Decay of Correlations

One class of systems in which the stronger mixing properties are often found is the class of smooth systems possessing uniform hyperbolicity (i.e. the tangent space to the manifold at each point splits into stable and unstable subspaces $E^s(x)$ and $E^u(x)$ such that the $\|DT|_{E_s(x)}\| \leq a < 1$ for all x and $\|DT^{-1}|_{E_u(x)}\| \leq a$ and $DT(E_s(x)) = E_s(T(x))$ and $DT(E_u(x)) = E_u(T(x))$). In some cases similar conclusions are found in systems possessing non-uniform hyperbolicity. See Katok and Hasselblatt's book [36] for an overview of hyperbolic dynamical systems, as well as the chapter in this volume on Smooth Ergodic Theory.

In the simple case of expanding piecewise continuous maps of the interval (that is, maps for which the absolute value of the derivative is uniformly bounded below by a constant greater than 1), it is known that if they are totally ergodic and topologically transitive (i.e. the forward images of any interval cover the entire interval), then provided that the map has sufficient smoothness (e.g. the map is C^1 and the derivative satisfies a certain additional summability condition), the map has a unique absolutely continuous invariant measure which is exact and whose natural extension is Bernoulli (see the paper of Góra [23] for results of this type proved under some of the mildest hypotheses). These results were originally established for maps that were twice continuously differentiable, and the hypotheses were progressively weakened, approaching, but never meeting, C^1. Subsequent work of Quas [68, 67] provided examples of C^1 expanding maps of the interval for which Lebesgue measure was invariant, but respectively not ergodic and not weak-mixing. Some of the key tools in controlling mixing in one-dimensional expanding maps that are absent in the C^1 case are bounded distortion estimates. Here, there is a constant $1 \leq C < \infty$ such that given any interval I on which some power T^n of T acts injectively and any sub-interval J of I, one has $1/C \leq (|T^n J|/|T^n I|)/(|J|/|I|) \leq C$. An early place in which bounded distortion estimates appear is the work of Rényi [69].

One important class of results for expanding maps establishes an exponential decay of correlations. Here, one starts with a pair of smooth functions f and g and one estimates $\int f \cdot g \circ T^n \, d\mu - \int f \, d\mu \int g \, d\mu$, where μ is an absolutely continuous invariant measure. If μ is mixing, we expect this to converge to 0. In fact though, in good cases this converges to 0 at an exponential rate for each pair of functions f and g belonging to a sufficiently smooth class. In this case, the measure-preserving transformation T is said to have exponential decay of correlations. See Liverani's article [45] for an introduction to a method of establishing this based on cones. Exponential decay of correlations implies in particular that the natural extension is Bernoulli.

Hu [29] has studied the situation of maps of the interval for which the derivative is bigger than 1 everywhere except at a fixed point, where the local behaviour is of the form $x \mapsto x + x^{1+\alpha}$ for $0 < \alpha < 1$. In this case, rather than exhibiting exponential decay of correlations, the map has polynomial decay of correlations with a rate depending on α.

In Young's survey [93], a variety of techniques are outlined for understanding the strong ergodic properties of non-uniformly hyperbolic diffeomorphisms. In her article [94], methods are introduced for studying many classes of non-uniformly hyperbolic systems by looking at suitably high powers of the map, for which the power has strong hyperbolic behaviour. The article shows how to understand the ergodic behaviour of these systems. These methods are applied (for example) to billiards, one-dimensional quadratic maps and Hénon maps.

10. Future Directions

Problem 1 (Mixing of all orders)**.** Does mixing imply mixing of all orders? Can the results of Kalikow, Ryzhikov and Host be extended to larger classes of measure-preserving transformations? Thouvenot observed that it is sufficient to establish the result for measure-preserving transformations of entropy 0. This observation (whose proof is based on the Pinsker σ-algebra) was stated in Kalikow's paper [32] and is reproduced as Proposition 3.2 in recent work of de la Rue [76] on the mixing of all orders problem.

Problem 2 (Multiple weak-mixing)**.** As mentioned above, Bergelson [5] showed that if T is a weak-mixing transformation, then there is a subset J of the integers of density 0 such that

$$\lim_{n\to\infty, n\notin J} \mu(A_0 \cap T^{-p_1(n)} A_1 \cap \ldots \cap T^{-p_k(n)} A_k) = \prod_{i=0}^{k} \mu(A_i)$$

whenever $p_1(n), \ldots, p_k(n)$ are non-constant integer-valued polynomials such that $p_i(n) - p_j(n)$ is unbounded for $i \neq j$. It is natural to ask what is the most general class of times that can replace the sequences $(p_1(n)), \ldots, (p_k(n))$. In unpublished notes, Bergelson and Håland considered as times the values taken by a family of integer-valued generalized polynomials (those functions of an integer variable that can be obtained by the operations of addition, multiplication, addition of or multiplication by a real constant and taking integer parts (e.g. $g(n) = \lfloor \sqrt{2} \lfloor \pi n \rfloor + \lfloor \sqrt{3} n \rfloor^2 \rfloor$)). They conjectured necessary and sufficient conditions for the analogue of Bergelson's weak-mixing polynomial ergodic theorem to hold, and proved the conjecture in certain cases.

In a recent paper of McCutcheon and Quas [48], the analogous question was addressed in the case where T is a mild-mixing transformation.

Problem 3 (Pascal adic transformation)**.** Vershik [89, 90] introduced a family of transformations known as the adic transformations. The underlying spaces for these transformations are certain spaces of paths on infinite graphs, and the transformations act by taking a path to its lexicographic neighbour. Amongst the adic transformations, the so-called Pascal adic transformation (so-called because the underlying graph resembles Pascal's triangle) has been singled out for attention in work of Petersen and others [64, 49, 15, 2]. In particular, it is unresolved whether this transformation is weak-mixing with respect to any of its ergodic measures. Weak-mixing has been shown by Petersen and Schmidt to follow from a number-theoretic condition on the binomial coefficients [15, 2].

Problem 4 (Weak Pinsker Conjecture)**.** Pinsker [66] conjectured that in a measure-preserving transformation with positive entropy, one could express the transformation as a product of a Bernoulli shift with a system with zero entropy. This conjecture (now known as the *Strong Pinsker Conjecture*) was shown to be false by Ornstein [54, 55]. Shields and Thouvenot [78] showed that the collection of transformations that can be written as a product of a zero entropy transformation with a Bernoulli shift is closed in the so-called $\bar{d}$-metric that lies at the heart of Ornstein's theory.

It is, however, the case that if $T\colon X \to X$ has entropy $h > 0$, then for all $h' \leq h$, T has a factor S with entropy h' (this was originally proved by Sinai [84] and reproved using the Ornstein machinery by Ornstein and Weiss in [60]). The *Weak Pinsker Conjecture* states that if a measure-preserving transformation T has entropy $h > 0$, then for all $\epsilon > 0$, T may be expressed as a product of a Bernoulli shift and a measure-preserving transformation with entropy less than ϵ.

References

[1] J. Aaronson. *An Introduction to Infinite Ergodic Theory.* American Mathematical Society, 1997.

[2] T. M. Adams and K. Petersen. Binomial coefficient multiples of irrationals. *Monatsh. Math.*, 125:269–278, 1998.

[3] S. Alpern. New proofs that weak mixing is generic. *Invent. Math.*, 32:263–278, 1976.

[4] A. Avila and G. Forni. Weak-mixing for interval exchange transformations and translation flows. *Ann. of Math.*, 165:637–664, 2007.

[5] V. Bergelson. Weakly mixing pet. *Ergodic Theory Dynam. Systems*, 7:337–349, 1987.

[6] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17:656–60, 1931.

[7] G. D. Birkhoff and P. A. Smith. Structural analysis of surface transformations. *J. Math*, 7:345–379, 1924.

[8] L. Boltzmann. Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte*, 63:679–711, 1871.

[9] L. Boltzmann. *Wissenschaftliche Abhandlungen.* Akad. Wissenschaften Berlin, 1909.

[10] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms.* Springer, 1975.

[11] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144, 2005.

[12] R. V. Chacon. Weakly mixing transformations which are not strongly mixing. *Proc. Amer. Math. Soc.*, 22:559–562, 1969.

[13] G. Choquet. Existence des représentations intégrales au moyen des points extrémaux dans les cônes convexes. *C. R. Acad. Sci. Paris*, 243:699–702, 1956.

[14] G. Choquet. Unicité des représentations intégrales au moyen de points extrémaux dans les cônes convexes réticulés. *C. R. Acad. Sci. Paris*, 243:555–557, 1956.

[15] M. Boshernitzan D. Berend and G. Kolesnik. Irrational dilations of Pascal's triangle. *Mathematika*, 48:159–168, 2001.

[16] P. Ehrenfest and T. Ehrenfest. *Begriffliche Grundlage der statistischen Auffassung in der Mechanik.* Number 4 in Encyclopädie der mathematischen Wissenschaften. Teubner, 1911.

[17] W. Feller. *An Introduction to Probability and its Applications.* Wiley, 1950.

[18] N. A. Friedman and D. S. Ornstein. On isomorphism of weak Bernoulli transformations. *Adv. Math.*, 5:365–394, 1970.

[19] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[20] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton, 1981.

[21] H. Furstenberg and B. Weiss. The finite multipliers of infinite ergodic transformations. In *The Structure of Attractors in Dynamical Systems (Proc. Conf., North Dakota State Univ., Fargo, N.D., 1977).* Springer, 1978.

[22] A. M. Garsia. A simple proof of E. Hopf's maximal ergodic theory. *J. Math. Mech.*, 14:381–382, 1965.

[23] P. Góra. Properties of invariant measures for piecewise expanding one-dimensional transformations with summable oscillations of derivative. *Ergodic Theory Dynam. Systems*, 14:475–492, 1994.

[24] P. Halmos. *Lectures on Ergodic Theory.* Chelsea, 1956.

[25] P. A. Halmos. In general a measure-preserving transformation is mixing. *Ann. of Math.*, 45:786–792, 1944.

[26] C. Hoffman and D. Heicklen. Rational maps are d-adic bernoulli. *Ann. of Math.*, 156:103–114, 2002.

[27] C. Hoffman and D. J. Rudolph. Uniform endomorphisms which are isomorphic to a Bernoulli shift. *Ann. of Math.*, 76:79–101, 2002.

[28] B. Host. Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Israel J. Math.*, 76:289–298, 1991.

[29] H. Hu. Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergodic Theory Dynam. Systems*, 24:495–524, 2004.

[30] S. Kalikow. Outline of ergodic theory. notes freely available for download - see `http://www.math.uvic.ca/faculty/aquas/kalikow/kalikow.html`.

[31] S. Kalikow. T, T^{-1} transformation is not loosely Bernoulli. *Ann. of Math.*, 115:393–409, 1982.

[32] S. Kalikow. Twofold mixing implies threefold mixing for rank one transformations. *Ergodic Theory Dynamical Systems*, 2:237–259, 1984.

[33] T. Kamae. A simple proof of the ergodic theorem using non-standard analysis. *Israel J. Math.*, 42:284–290, 1982.

[34] A. Katok. Interval exchange transformations and some special flows are not mixing. *Israel J. Math.*, 35:301–310, 1980.

[35] A. Katok. Smooth non-Bernoulli K-automorphisms. *Invent. Math.*, 61:291–299, 1980.

[36] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems.* Cambridge, 1995.

[37] Y. Katznelson. Ergodic automorphisms of $\mathbb{T}^n$ are Bernoulli shifts. *Israel J. Math.*, 10:186–195, 1971.

[38] Y. Katznelson and B. Weiss. A simple proof of some ergodic theorems. *Israel J. Math.*, 42:291–296, 1982.

[39] M. Keane and M. Smorodinsky. Bernoulli schemes of the same entropy are finitarily isomorphic. *Ann. of Math.*, 109:397–406, 1979.

[40] M. S. Keane and K. E. Petersen. Nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. In *Dynamics and Stochastics: Festschrift in Honor of M. S. Keane*, pages 248–251. Institute of Mathematical Statistics, 2006.

[41] A. N. Kolmogorov. New metric invariant of transitive dynamical systems and endomorphisms of Lebesgue spaces. *Dokl. Russ. Acad. Sci.*, 119:861–864, 1958.

[42] B. O. Koopman. Hamiltonian systems and Hilbert space. *Proc. Nat. Acad. Sci.*, 17:315–218, 1931.

[43] A. Krámli, N. Simányi, and D. Szász. The K-property of three billiard balls. *Ann. of Math.*, 133:37–72, 1991.

[44] F. Ledrappier. Un champ Markovien peut être d'entropie nulle et mélangeant. *C. R. Acad. Sci. Paris, Sér A-B*, 287:561–563, 1978.

[45] C. Liverani. Decay of correlations. *Ann. of Math.*, 159:1275–1312, 2004.

[46] D. W. Masser. Mixing and linear equations over groups in positive characteristic. *Israel J. Math.*, 142:189–204, 2004.

[47] H. Masur. Interval exchange transformations and measured foliations. *Ann. of Math.*, 115:169–200, 1982.
[48] R. McCutcheon and A. Quas. Generalized polynomials and mild mixing systems. *Canad. J. Math.*, 2007. to appear.
[49] X. Méla and K. Petersen. Dynamical properties of the Pascal adic transformation. *Ergodic Theory Dynam. Systems*, 25:227–256, 2005.
[50] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci. USA*, 18:70–82, 1932.
[51] D. Newton and W. Parry. On a factor automorphism of a normal dynamical system. *Ann. Math. Statist.*, 37:1528–1533, 1966.
[52] D. S. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Adv. Math*, 4:337–352, 1970.
[53] D. S. Ornstein. An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Adv. Math*, 10:49–62, 1973.
[54] D. S. Ornstein. A K-automorphism with no square root and Pinsker's conjecture. *Adv. Math*, 10:89–102, 1973.
[55] D. S. Ornstein. A mixing transformation for which Pinsker's conjecture fails. *Adv. Math*, 10:103–123, 1973.
[56] D. S. Ornstein. *Ergodic Theory, Randomness, and Dynamical Systems*. Yale University Press, 1974.
[57] D. S. Ornstein and P. C. Shields. An uncountable family of K-automorphisms. *Adv. Math*, 10:89–102, 1973.
[58] D. S. Ornstein and B. Weiss. Geodesic flows are Bernoullian. *Israel J. Math*, 14:184–198, 1973.
[59] Donald S. Ornstein. On the root problem in ergodic theory. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 347–356. Univ. California Press, 1972.
[60] Donald S. Ornstein and Benjamin Weiss. Unilateral codings of Bernoulli systems. *Israel J. Math.*, 21:159–166, 1975.
[61] J. C. Oxtoby. Ergodic sets. *Bull. Amer. Math. Soc.*, 58:116–136, 1952.
[62] W. Parry. *Topics in Ergodic Theory*. Cambridge, 1981.
[63] K. Petersen. *Ergodic Theory*. Cambridge, 1983.
[64] K. Petersen and K. Schmidt. Symmetric Gibbs measures. *Trans. Amer. Math. Soc.*, 349:2775–2811, 1997.
[65] R. Phelps. *Lectures on Choquet's Theorem*. Van Nostrand, 1966.
[66] M. S. Pinsker. Dynamical systems with completely positive or zero entropy. *Soviet Math. Dokl.*, 1:937–938, 1960.
[67] A. Quas. A C^1 expanding map of the circle which is not weak-mixing. *Israel J. Math.*, 93:359–372, 1996.
[68] A. Quas. Non-ergodicity for C^1 expanding maps and g-measures. *Ergodic Theory Dynam. Systems*, 16:531–543, 1996.
[69] A. Rényi. Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hungar.*, 8:477–493, 1957.
[70] F. Riesz. Some mean ergodic theorems. *J. London Math. Soc.*, 13:274–278, 1938.
[71] V. A. Rokhlin. A 'general' measure-preserving transformation is not mixing. *Dokl. Akad. Nauk SSSR. Ser. Mat.*, 60:349–351, 1948.
[72] V. A. Rokhlin. On endomorphisms of compact commutative groups. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 13:329–340, 1949.
[73] V. A. Rokhlin and Ya. Sinai. Construction and properties of invariant measurable partitions. *Dokl. Akad. Nauk SSSR*, 141:1038–1041, 1961.
[74] W. Rudin. *Real and Complex Analysis*. McGraw Hill, 1966.
[75] D. J. Rudolph. *Fundamentals of Measurable Dynamics*. Oxford, 1990.
[76] T. de la Rue. 2-fold and 3-fold mixing: why 3-dot-type counterexamples are impossible in one dimension. *Bull. Braz. Math. Soc. (N.S.)*, 37(4):503–521, 2006.
[77] V. V. Ryzhikov. Joinings and multiple mixing of the actions of finite rank. *Funct. Anal. Appl.*, 27:128–140, 1993.
[78] P. Shields and J.-P. Thouvenot. Entropy zero $\times$ Bernoulli processes are closed in the $\bar{d}$-metric. *Ann. Probab.*, 3:732–736, 1975.

[79] N. Simányi. Proof of the ergodic hypothesis for typical hard ball systems. *Ann. Henri Poincaré*, 5:203–233, 2004.
[80] Nándor Simányi. Proof of the Boltzmann-Sinai ergodic hypothesis for typical hard disk systems. *Invent. Math.*, 154:123–178, 2003.
[81] Nándor Simányi and Domokos Szász. Hard ball systems are completely hyperbolic. *Ann. of Math.*, 149:35–96, 1999.
[82] Ya. Sinai. *Introduction to Ergodic Theory.* Princeton, 1976. translation of the 1973 Russian original.
[83] Ya. G. Sinai. On the notion of entropy of a dynamical system. *Dokl. Russ. Acad. Sci.*, 124:768–771, 1959.
[84] Ya. G. Sinai. On a weak isomorphism of transformations with invariant measure. *Mat. Sb. (N.S.)*, 63:23–42, 1964.
[85] Ya. G. Sinai. Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat. Nauk*, 25:141–192, 1970.
[86] Ya. G. Sinai and N. I. Chernov. Ergodic properties of some systems of two-dimensional disks and three-dimensional balls. *Uspekhi Mat. Nauk*, 42:153–174, 256, 1987.
[87] M. Smorodinsky. A partition on a Bernoulli shift which is not weakly Bernoulli. *Math. Systems. Th.*, 5:201–203, 1971.
[88] W. Veech. Gauss measures for transformations on the space on interval exchange maps. *Ann. of Math.*, 115:201–242, 1982.
[89] A. Vershik. A description of invariant measures for actions of certain infinite-dimensional groups. *Soviet Math. Dokl.*, 15:1396–1400, 1974.
[90] A. Vershik. Uniform algebraic approximation of shift and multiplication operators. *Soviet Math. Dokl.*, 24:97–100, 1981.
[91] P. Walters. *An Introduction to Ergodic Theory.* Springer, 1982.
[92] D. Williams. *Probability with Martingales.* Cambridge, 1991.
[93] L.-S. Young. Ergodic theory of differentiable dynamical systems. In *Real and Complex Dynamical Systems (Hillerød, 1993)*, pages 293–336. Kluwer, 1995.
[94] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.*, 147:585–650, 1998.
[95] S. A. Yuzvinskii. Metric automorphisms with a simple spectrum. *Soviet Math. Dokl.*, 8:243–245, 1967.

Department of Mathematics and Statistics, University of Victoria, Canada