Notes on Ergodic Theory

by Jeff Steif

1 Introduction

Because of its vast scope, it is difficult to give an overview of ergodic theory. Nonetheless, one of the original questions in statistical physics is the equality of so–called phase averages and time averages. Will the amount of time that some physical system spends in some region in phase space in the long run (i.e., the time average) be the same as the amount of volume occupied by this region (the phase average)? For example, if you continually mix your coffee cup, is it the case that the portion of time in the long run that a given particle spends in the top half of the cup is equal to 1/2? This is called the ergodic hypothesis and is one of the origins of ergodic theory.

Ergodic theory impinges on many areas of mathematics- most notably, probability theory and dynamical systems as well as Fourier analysis, functional analysis and group theory.

At the simplest level, ergodic theory is the study of transformations of a measure space which preserve the measure. However, with this dry description, both the interest of the subject and the wide range of its applications are lost.

The point of these notes is to give the reader some feeling of what ergodic theory is and how it can be used to shed some light on some classical problems in mathematics. As I will be concentrating more on how ergodic theory can be used, I am afraid the reader will end up knowing how ergodic theory can be used but not knowing what ergodic theory is.

In short, ergodic theory is the following. Certainly, dynamics of any kind are important. Three main areas of dynamics are differential dynamics (the study of iterates of a differentiable map on a manifold), topological dynamics (the study of iterates of a continuous map on a metric or topological space), and measurable dynamics (the study of iterates of a measure–preserving map on a measure space). Ergodic theory is the third of these. However, in these notes, we will be dealing with both topological dynamics and measurable dynamics.

In the next section, I will give what might be a prejudicial view of the history of the subject but which is easy to write since I'm just copying it (and slightly modifying it) from the introduction to my thesis. The reader is encouraged to just skim (or read as he or she wishes) this section. One will not lose anything by immediately turning to §3.

2 A Brief Overview of Ergodic Theory

Ergodic Theory began in the last decade of the nineteenth century when Poincaré studied the solutions of differential equations from a new point of view. From this perspective, one concentrated on the set of all possible solution curves instead of the individual solution curves. This naturally brought about the notion of the phase space and what came to be called the qualitative theory of differential equations. Another motivation for ergodic theory came from statistical mechanics where one of the central questions was the equality of phase (space) means and time means for certain physical systems, the so called *ergodic hypothesis*.

The mathematical beginning of ergodic theory is usually considered to have taken place in 1931 when G.D. Birkhoff proved the *pointwise ergodic theorem*. It was at this point that ergodic theory became a legitimate mathematical discipline. Moreover, ergodic theory became, in its most general form, the study of abstract dynamical systems, where an abstract dynamical system is a quadruple $(\Omega, \mathcal{A}, \mu, \pi_G)$ where Ω is a set, $\mathcal{A}$ is a σ–field of subsets of Ω, μ is a probability measure on $\mathcal{A}$ and π_G is a group action of G on Ω by bijective bimeasurable measure-preserving transformations. G is always assumed to be locally compact. Moreover, it is also assumed that the mapping from $G \times \Omega$ to Ω induced from the group action is jointly measurable where the Borel structure of G is generated by its topology. Actually, G may be only a semigroup in many contexts.

In these notes, G will mostly be $\mathbf{Z}$ but it might be $\mathbf{Z}^n$ or $\mathbf{N}$ (in which case we have a semigroup). If G is $\mathbf{N}$ or $\mathbf{Z}$, then G is generated by one transformation T. In this case, the Birkhoff pointwise ergodic theorem states that for all f in $L^1(\mu)$

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

exists a.e., and denoting this limit by f^*, f^* is also in $L^1(\mu)$ and

$$\int_\Omega f d\mu = \int_\Omega f^* d\mu.$$

Furthermore, if the dynamical system is ergodic, then f^* is constant a.e., where ergodicity means that all invariant sets (sets A such that $T^{-1}A = A$)

have measure 0 or 1. This theorem also holds when G is taken to be $\mathbf{Z}^n$ or $\mathbf{R}^n$. The proof of the Birkhoff ergodic theorem for a single transformation can be found in [Wal], while the more general version can be found in [D+S].

Once it was clear that the mathematical objects that should be studied in ergodic theory are abstract dynamical systems, it was natural to define the notion of isomorphism between two such systems providing that the two groups acting are the same. One says that $(\Omega, \mathcal{A}, \mu, \pi_G)$ and $(\Omega', \mathcal{B}, \nu, \Psi_G)$ are isomorphic if there are G–invariant measurable sets A contained in Ω and B contained in Ω' each of measure 1 such that for all g, π_g and Ψ_g are bijective when restricted to these sets and such that there exists a bimeasurable measure-preserving mapping f from A to B such that $f(\pi_g(x)) = \Psi_g(f(x))$ for all g in G and x in A. Since only sets of full measure are relevant, it is obvious that this is the correct definition. In order to distinguish between dynamical systems, a number of ergodic theoretic properties were introduced, all of which were isomorphism invariants. In addition to ergodicity, other properties that are often considered are weak-mixing, mixing, k-mixing, and Bernoulli some of which we will care about and therefore define. Dynamical systems also yield a canonical unitary representation of the respective group on the corresponding L^2 space of the underlying measure space, and it is sometimes useful to consider spectral invariants as well. The definitions of these standard notions can be found in [Wal] in the case where G is $\mathbf{Z}$. For more general groups, some of these properties are given in certain areas of the literature although it is not so easy to track down all the definitions for general groups. We give here the definitions of ergodicity, mixing and Bernoulli since these are the only concepts that we will need.

Definition 2.1: *A dynamical system* $(\Omega, \mathcal{A}, \mu, \pi_G)$ *is* **ergodic** *if whenever*

$\pi_g A = A$ for all g in G where A is measurable, then $\mu(A) = 0$ or 1.

Definition 2.2: *A dynamical system $(\Omega, \mathcal{A}, \mu, \pi_G)$ is* **mixing** *if G is not compact and if for all measurable A and B contained in Ω, $\lim_{g\to\infty} \mu(\pi_g(A) \cap B) = \mu(A)\mu(B)$.*

In the above, as $g \to \infty$ means as g leaves compact subsets of G.

Finally, to introduce the notion of a Bernoulli system, we will always assume that G is of the form $\mathbf{R}^m \times \mathbf{Z}^n$. We first define this when G is $\mathbf{Z}^n$.

Definition 2.3: *$(\Omega, \mathcal{A}, \mu, \pi_{\mathbf{Z}^n})$ is* **Bernoulli** *if it is isomorphic to $(W^{\mathbf{Z}^n}, \mathcal{B}, p, \pi_{\mathbf{Z}^n})$ for some Lebesgue space W where $\mathcal{B}$ is the canonical σ–field on the product space, p is product measure and $\pi_{\mathbf{Z}^n}$ is the canonical action of $\mathbf{Z}^n$ on $W^{\mathbf{Z}^n}$.*

Next, if G is $\mathbf{R}^m \times \mathbf{Z}^n$, then we can restrict this action to the subgroup $\mathbf{Z}^{m+n}$.

Definition 2.4: *$(\Omega, \mathcal{A}, \mu, \pi_{\mathbf{R}^m \times \mathbf{Z}^n})$ is* **Bernoulli** *if the corresponding discrete dynamical system $(\Omega, \mathcal{A}, \mu, \pi_{\mathbf{Z}^m \times \mathbf{Z}^n})$ is Bernoulli.*

If W is a finite set with a certain probability measure defined on it and n is 1, the corresponding system is also referred to as a Bernoulli shift. From a probabilistic viewpoint, these are nothing but independent and identically distributed random variables.

In [Wal], it is shown that for $\mathbf{Z}$–actions the above properties (some of which we have not defined) are in order from strongest to weakest, Bernoulli, k-mixing, mixing, weak mixing, and ergodic. Moreover, k-mixing is equivalent to the holding of the well-known 0–1 Law in probability theory for any stationary stochastic process arising from the dynamical system. In particular, a Bernoulli system satisfies the 0–1 Law.

We continue with a brief outline of the classical development of ergodic theory until the important work of Ornstein in 1970.

To motivate this work, we will consider the simplest type of Bernoulli shifts. If $(p_1, \ldots, p_k)$ are such that each p_i is non-negative and $\sum_i p_i = 1$, we let $B(p_1, \ldots, p_k)$ denote the Bernoulli system $(\{0, 1, \ldots, k-1\}^{\mathbf{Z}}, \mathcal{A}, m, \mathbf{Z})$ where $\mathcal{A}$ is the natural σ–field, m is product measure where each $\{0, 1, \ldots, k-1\}$ has probability measure $(p_1, \ldots, p_k)$, and $\mathbf{Z}$ acts canonically on this space. Here we mean that $T(\{x_n\}) = \{y_n\}$ where $y_n = x_{n+1}$.

One of the first natural questions that arose is whether $B(1/2, 1/2)$ is isomorphic to $B(1/3, 1/3, 1/3)$. None of the properties listed above could distinguish these, both of these systems satisfying all of the above properties. In addition, the induced unitary operators were unitarily equivalent. Finally, in 1958, Kolmogorov introduced the notion of entropy, which was already being used in information theory, into ergodic theory. The definition was slightly modified in 1959 by Sinai. This notion assigns a non-negative real number to each dynamical system which is an isomorphism invariant (see [Wal]). This number then allowed one to finally distinguish between $B(1/2, 1/2)$ and $B(1/3, 1/3, 1/3)$ since it was easy to show that the entropies of the above two systems were $\log(2)$ and $\log(3)$, respectively. After this, the next natural question was asked: If two Bernoulli shifts have the same entropy, are they necessarily isomorphic? In this same year, Meshalkin ([Mesh]) obtained some positive results in this direction when certain algebraic relationships held between the two probability vectors. Finally, ten years later, in 1969, using very powerful methods, Ornstein proved this conjecture in general. The techniques developed by Ornstein not only solved the isomorphism problem but also gave certain criteria which could be more readily checked which implied that a dynamical system is isomorphic to a Bernoulli shift. References for this theory are [Sh] (which covers the case $G = \mathbf{Z}$), [O] (which covers the case $G = \mathbf{Z}$ or $\mathbf{R}$), and [Feld] or [Lind] (each covering the case $G = \mathbf{R}^m \times \mathbf{Z}^n$.) Recently, the theory has been extended to general amenable groups ([O+W1]).

Ergodic theory arises in many different contexts in mathematics, in particular in probability theory in the study of stationary stochastic processes. In fact, there is a type of correspondence between dynamical systems and stationary stochastic processes, which was alluded to earlier. This correspondence is, however, by no means one to one. In [Wal], it is shown how a dynamical system yields many stationary stochastic processes (different processes being obtained from different partitions of the underlying measure space). On the other hand, it is clear how a stationary process yields a dynamical system. If, for example, $\{X_n\}$ is a stationary process taking on only the values 0 and 1, then this induces a measure on $\{0,1\}^{\mathbf{Z}}$ which is invariant under the natural $\mathbf{Z}$ action (this is just the definition of stationarity).

3 Diophantine Approximation and Equidistribution

This section is quite long. The purpose is to see how methods from topological dynamics and measurable dynamics can be used in number theory and analysis. We will therefore recover some theorems in these areas using these tools. The two objects we want to look at are
1) diophantine approximation and 2) equidistribution.

We need to develop the study of continuous mappings on compact metric spaces and their corresponding invariant measures. For the results on diophantine approximation that we will obtain, we do not need to consider invariant measures and will therefore be working purely with topological dynamics. However for the results on equidistribution, we will need to consider invariant measures. We will start off with only topological dynamics. **Our setup in this section will always be a compact metric space X together with a continuous map T from X to itself.**

The first important concept is recurrence, the phenomenon that some points return arbitrarily close to themselves. This can be viewed as a generalization of a point being periodic.

Definition 3.1: *$x \in X$ is recurrent if there is $n_i \to \infty$ with $T^{n_i}(x) \to x$ as $i \to \infty$.*

Theorem 3.2: *There always exist recurrent points.*

Proof: Let A be a minimal (with respect to inclusion) closed nonempty $T-$invariant set (invariance means once we are in A, we stay in A or $T(A) \subseteq A$.) It is an easy consequence of Zorn's lemma that such an A exists. Now each $x \in A$ is recurrent since by minimality, for each $x \in A$, we must have that the closure of $\{T^n(x), n \geq 1\}$ is A. This gives us that x is recurrent. □

As in any mathematical object, there is a notion of equivalence and of factor.

Definition 3.3: (X, T) and (Y, S) are equivalent (or isomorphic or conjugate) if there is a homeomorphism h from X to Y such that $hT = Sh$.

The orbit structure (from the topological point of view) of two such equivalent systems must be the same.

Definition 3.4: *(Y, S) is a factor of (X, T) if there is a surjective continuous h from X to Y such that $hT = Sh$.*

As usual, a factor inherits properties of the first system. For example,

Theorem 3.5: *If (Y, S) is a factor of (X, T) with factor map h and $x \in X$ is recurrent, then $h(x)$ is recurrent.*

Proof: Trivial. $\Box$

The above discussion all of which was soft allows us to easily prove Kronecker's Theorem.

Theorem 3.6 (Kronecker's Theorem): *Let T be a rotation of the unit circle. Then every point is recurrent.*

In fact, we prove the following stronger result.

Theorem 3.7: *Let G be a compact (not necessarily abelian) group. Let $w \in G$ and consider the mapping T_w from G to itself given by left multiplication by w, so $T_w(g) = wg$. Then every point is recurrent.*

Proof: We know by Theorem 3.2 that some $x \in G$ is recurrent. Let g be arbitrary and consider the map from G to itself given by right multiplication by $x^{-1}g$. This is a homeomorphism from G to G which conjugates T_w with itself. (The fact that this conjugates is simply the associative law of the group. This is why we used right multiplication. If we had used left multiplication and the group were not abelian, it would not have conjugated.) It follows from Theorem 3.5 that $xx^{-1}g = g$ is recurrent. $\Box$

With some more work, we will be able to prove the following less trivial result which is due to Hardy and Littlewood (see [HL]).

Theorem 3.8: Let α be any real number. Then for all $\epsilon > 0$, the diophantine inequality

$$|\alpha n^2 - m| < \epsilon$$

is solvable for $n, m \in \mathbf{Z}$, $n \geq 1$.

Since so far all of our theorems have been soft, we obviously will need to do something a little harder to obtain this result but this extension is not so hard. Before proving Theorem 3.8 and an extension of this result, we will need to do further development. (Theorem 3.8 is of course trivial if α is rational.)

Before this further development, let's first relate this result to Kronecker's Theorem. Consider the map T which rotates the unit circle by θ. Then the orbit of 0 is

$$\theta, 2\theta, 3\theta, \ldots.$$

By Kronecker's Theorem, for any ϵ, there is an n such that $n\theta$ is within ϵ of 0 on the circle, i.e.,

$$|n\theta - m| < \epsilon$$

is solvable for integers n and m, $n \geq 1$. Theorem 3.8 says that the forward orbit of 0 gets close to itself even when we look only at times which are squares, i.e., $0 \in \overline{\{T^{n^2}(0)\}_{n \geq 1}}$. It turns out that such a theorem is true in general, namely, given a topological system, there always exists some x and $n_i \to \infty$ with $T^{n_i^2}(x) \to x$ (i.e., recurrence along squares) which by the previous discussion clearly gives Theorem 3.8. This general result is however more difficult and seems to require invariant measures and unitary operators which we will come back to later. As we want to stay in the context of topological dynamics, we use another approach.

Continuing with our development, we know images of recurrent points under factor maps are also recurrent (Theorem 3.5). We will need to show that in a particular case, all inverse images of a recurrent point of a factor are also recurrent (which obviously is not true in general).

The setting for this is

Group extensions or skew products

Let (Y,T) be a topological system and let $\psi : Y \to G$ be continuous where G is a compact group. (If you don't know what a compact group is, assume G is the unit circle in the complex plane with a multiplication given by usual complex multiplication. You won't lose much by doing this.) The "group extension of Y by ψ" is the topological system given by $Y \times G$ and

$$(y, g) \to (T(y), \psi(y)g)$$

where the multiplication in the second piece is in the group G. How does one think of such a skew product? Picture $Y \times G$ as a square. We move the base Y by T so that each fiber $y \times G$ goes to $T(y) \times G$. Moreover, this fiber is "rotated" by simply multiplying (on the left) by $\psi(y)$ (i.e., $g \to \psi(y)g$). (The analogy with a skew product in group theory is fairly clear.)

Theorem 3.9: *Consider a group extension of (Y, T) by (G, ψ). If y is T–recurrent, then (y, g) is recurrent for the group extension for all g.*

While the proof is not hard (but not as trivial as our previous results), we do it later and do the applications first. The applications will be the Hardy–Littlewood result and an extension of this result.

Proof of Theorem 3.8 (Hardy–Littlewood): Let T^2 denote the 2-dimensional torus (which is a nice group) and let f be the mapping from T^2

to itself given by

$$f(\theta, \phi) = (\theta + \alpha, \phi + 2\theta + \alpha).$$

It is easy to see that this is a group extension of $T : T \to T, \theta \to \theta + \alpha$ with (G, ψ) being $(T, \psi(\theta) = 2\theta + \alpha)$.

Since all points of T are recurrent (Kronecker's Theorem), Theorem 3.9 tells us that all points of the extension are also recurrent, in particular, $(0, 0)$. The orbit of $(0, 0)$ in the group extension is

$$(0, 0) \to (\alpha, \alpha) \to (2\alpha, 4\alpha) \to (3\alpha, 9\alpha)$$

and it's easy to see by induction that

$$T^n(0, 0) = (n\alpha, n^2\alpha).$$

By recurrence, this gets very close to $(0, 0)$ (mod 1) and hence the second coordinate gets close to 0 (mod 1). This means that $|\alpha n^2 - m| < \epsilon$ is solvable for $n \geq 1$ and $m \in \mathbf{Z}$. $\square$

We now go on and use this same method to prove the stronger

Theorem 3.10: Let $p(x)$ be a real polynomial with $p(0) = 0$. Then we can solve the diophantine inequality $|p(n) - m| < \epsilon$ with $n \geq 1$ and $m \in \mathbf{Z}$.

[$p(x) = \alpha x^2$ gives Hardy-Littlewood]

Proof: Assume that d is the degree of the polynomial. Let $p_d(x) = p(x)$. Let $p_{d-1}(x) = p_d(x+1) - p_d(x)$. Let $p_{d-2}(x) = p_{d-1}(x+1) - p_{d-1}(x)$. Keep going until $p_0(x) = p_1(x+1) - p_1(x)$. Note that the degree of p_i is i and we let $\alpha = p_0$.

Consider the mapping from T^d (the d–dimensional torus) to itself given by

$$(\theta_1, \theta_2, \ldots, \theta_d) \to (\theta_1 + \alpha, \theta_2 + \theta_1, \theta_3 + \theta_2, \ldots, \theta_d + \theta_{d-1}).$$

It is easily seen that this is a group entension of a map of the $d-1$–torus which is a group extension of a map on the $d-2$–torus, etc. As all points of the base $\theta_1 \to \theta_1 + \alpha$ are recurrent, induction plus Theorem 3.9 tells us all points of the mapping on T^d are recurrent. Consider now the orbit of $(p_1(0), p_2(0), \ldots, p_d(0))$.

As $p_i(n) + p_{i-1}(n) = p_i(n+1)$, we see that

$$T(p_1(n), p_2(n), \ldots, p_d(n)) = (p_1(n+1), p_2(n+1), \ldots, p_d(n+1))$$

and hence

$$T^n(p_1(0), p_2(0), \ldots, p_d(0)) = (p_1(n), p_2(n), \ldots, p_d(n)).$$

Using the recurrence of $(p_1(0), p_2(0), \ldots, p_d(0))$ and the fact that $p_d(0) = 0$, by looking at the dth coordinate, the result is proved. $\Box$.

Proof of Theorem 3.9: First, if T is any topological dynamical system, we let $Q(x) = \overline{\{T^n(x), n \geq 1\}}$ be the closure of the forward orbit of x. It is clear that $x \in Q(x)$ is equivalent to x being recurrent.

Next note that for any $h \in G$, the map which multiplies the second coordinate on the right by h conjugates the group extension with itself and hence if we can show that (y, e) is recurrent, it will follow (from Theorem 3.5) that all (y, g) are recurrent. Let R_h denote this continuous map (which was also used in proving the generalization of Kronecker's Theorem).

Using compactness and the fact that y is T–recurrent, there clearly exists some $(y, k_1) \in Q(y, e)$. Applying R_{k_1} to this inclusion and noticing $QR_{k_1}(x) = R_{k_1}Q(x)$ gives us $(y, k_1^2) \in Q(y, k_1)$. Noticing that the relationship is $x' \in Q(x)$ is a transitive relationship (check this), we get

$$(y, k_1^2) \in Q(y, e).$$

By induction we get that for all m

$$(y, k_1^m) \in Q(y, e).$$

Using the generalization of Kronecker's Theorem to G together with the fact that $Q(y, e)$ is closed, we get

$$(y, e) \in Q(y, e)$$

as desired. $\Box$

THIS ENDS THE FIRST PART OF THIS SECTION.

We now move into equidistribution and this will require that we move into the subject of the set of invariant measures of a topological dynamical system.

Definition 3.11: *A measure μ is T–invariant if $\mu T^{-1}(A) = \mu(A)$ for all A.*

Theorem 3.12: *There exists a T–invariant measure.*

We give 2 proofs, the first of which uses functional analysis and the second of which is more hands on. I prefer the second. The first proof uses a result which might not be so familiar, the Markov–Kakutani Fixed Point Theorem which can be read about in Rudin's functional analysis book. (The first edition of this book contains something called the Kakutani Fixed Point Theorem but this is not it. It is only contained in the second edition on page 140). The second proof uses much less. Note that Theorem 3.12 is trivial if there exists a periodic point (Why?).

Proof 1: Let $C(X)$ be the Banach space of continuous functions on X in the sup norm. T then induces a bounded operator on $C(X)$ to itself (by $f \to fT$) which induces a map T^* on the dual space of $C(X)$ which is just (by the Riesz Representation Theorem) signed Borel measures (which are necessarily regular) on X, which we denote by $M(X)$. One checks that the probability measures $P(X)$ on X are T^*–invariant, convex and compact in the weak* topology. Then the Markov–Kakutani fixed point theorem guarantees that $P(X)$ has a fixed point for T^*. Unravelling definitions gives that this fixed point is an invariant measure for T.

Proof 2: Let $x \in X$. Let μ_n be uniform distribution on the first n points of the orbit of x, ie,

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)}.$$

(Notice that μ_n is more or less invariant if n is large.) $P(X)$ is a nice compact metric space in the weak* topology and so there is a limit μ for some subsequence of the μ_n's.

EXERCISE. Check that any such limit μ is T–invariant. $\Box$

The reader should note that Proof 2 essentially is proving the Markov–Kakutani Fixed Point Theorem in a special case. The method of the second proof gives another theorem which seems uninteresting but will be useful later for proving a uniformity in the ergodic theorem in certain cases.

Theorem 3.13: Let ν_n be any sequence in $P(X)$. Let

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} T^i \nu_n.$$

Then any weak limit of the μ_n's is T–invariant. ($T\nu(A) = \nu(T^{-1}(A))$ here).

Note: $\nu_n = \delta_x$ for all n gives Proof 2 above.

Proof: EXERCISE. $\Box$

....

ASIDE (which you will lose nothing from skipping but which is interesting.) One might wonder if one can generalize Theorem 3.12 in a different direction. For example, if T and S are two mappings acting on a compact metric space, is there a probability measure which is both T and S invariant? The answer turns out to be always if S and T commute and not necessarily otherwise. More generally, one may ask if given a group G, is it the case that whenever it acts on a compact metric space, there is then a probability measure which is invariant under all elements of G. It turns out it depends on certain "geometrical" properties of the group. Groups for which there always exist such a probability measure are called amenable. It turns out that all abelian groups are amenable. This is not obvious at all and is "essentially" the Markov–Kakutani Fixed Point Theorem. An example of a group that is not amenable is F_2, the free nonabelian group on two letters. To see why this is the case, let S be an irrational rotation of the circle and let T be any homeomorphism of the circle which does not preserve Lebesgue measure. Since there are no relations in the group F_2, S and T generate an action of F_2 on the circle. Since the only invariant measure for an irrational rotation is Lebesgue measure (a fact which we will prove later), there is no invariant measure for the group action.

How can one "see" if a group is amenable? One constructs its so-called Cayley graph as follows. Choose a finite generating set for the group (I guess we will assume the group is finitely generated). Let the vertices of the graph be the elements of the group and put an edge between x and y if x=gy for some g in the generating set. For example, if G is $\mathbf{Z}^2$ with generating set $\{(1,0),(0,1)\}$, then the Cayley graph is the usual 2–dimensional lattice. The group is amenable if and only if there exist finite sets in the graph whose "surface" to "volume" ratio goes to 0. (It turns out that this is independent of the generating set and hence a property of the group itself.) Notice that this is true for $\mathbf{Z}^2$. For F_2 (using the obvious generating set), the Cayley graph is a tree (this means that there are no loops) and it is well known that for homogeneous trees, the "surface" to "volume" ratio over all sets stays bounded away from 0.

END OF ASIDE

....

We now know (by Theorem 3.12) that there are invariant measures. But why are there any ergodic ones?

Let $\mathcal{I}$ denote the set of invariant measures which you should easily check to be compact (weakly) and convex. Since $\mathcal{I}$ is contained in the dual space of $C(X)$, the Krein-Milman theorem tells us that $Ext(\mathcal{I})$ (the set of extremal elements of $\mathcal{I}$) is nonempty. The existence of ergodic measures now follows from the following result.

Theorem 3.14: Let $\mu \in \mathcal{I}$. Then $\mu \in Ext(\mathcal{I})$ if and only if μ is ergodic.

Proof: If μ is not ergodic, let A be an invariant set which has measure in $(0, 1)$. Let $\mu_1(B) = \frac{\mu(B \cap A)}{\mu(A)}$. (This is μ restricted to A and renormalized). Let $\mu_2(B) = \frac{\mu(B \cap A^c)}{\mu(A^c)}$.

EXERCISE. Check that μ_1 and μ_2 are distinct elements of $\mathcal{I}$ and that μ is a nontrivial convex combination of these guys proving μ is not extremal.

For the other way, there is more than one proof, one which uses the pointwise ergodic theorem and one which does not. We use the pointwise ergodic theorem since this argument is very illustrative. For the statement of the ergodic theorem, see either §2 or §4.

We now assume that μ is ergodic and that it is a nontrivial convex combination of μ_1 and μ_2 with the latter two measures in $\mathcal{I}$. We need to show $\mu = \mu_1$. If not there exists B with $\mu(B) \neq \mu_1(B)$. The ergodic theorem (together with the ergodicity of μ) applied to μ tells us that

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{T^i(x)\in B} \to \mu(B)$$

for μ a.e. x. As $\mu_1 \leq \mu$ (absolutely continuous wrt), we also have

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{T^i(x)\in B} \to \mu(B)$$

for μ_1 a.e. x.

On the other hand, the ergodic theorem applied to μ_1 tells us that

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{T^i(x)\in B} \to h(x)$$

for μ_1 a.e. x for some function $h(x)$ which has integral with respect to μ_1 equal to $\mu_1(B)$. As $\mu(B) \neq \mu_1(B)$, the above gives us a contradiction. □

EXERCISE: Using a method similar to the proof that μ ergodic implies μ is extremal above, show that any two distinct ergodic measures are mutually singular.

The next concept has to do with the orbit of a point being uniformly distributed.

Definition 3.15: *If $\mu \in \mathcal{I}$, then $x \in X$ is* **generic for** μ *if*

$$\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i(x)} \to \mu.$$

In words, uniform distribution on the first n points of the orbit of x approaches μ as $n \to \infty$.

The following tells us that for ergodic measures, there are many generic points.

Theorem 3.16: *If μ is ergodic, then μ a.e. x is generic for μ.*

Proof: The definition of x being μ–generic is equivalent (check this) to

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \to \int f d\mu$$

for all continuous f. For a fixed f, this certainly holds for μ a.e. x by the ergodic theorem. One cannot simply put together all these sets of measure 0 (for various f) since there are uncountable many f. But you do the next most obvious thing. Choose a countable set of f which are dense in $C(X)$. Certainly there is a set of full measure such that the above holds for this countable set.
EXERCISE: Finish the proof with a simple approximation argument. □

EXERCISE: Show that for nonergodic μ's, it is possible that there are no generic points for μ.

We now want to obtain Weyl's theorem (and other stuff also). One of the key concepts in topological dynamics was recurrence. Here a key idea will be unique ergodicity (THIS IS A CONDITION OF THE TOPOLOGICAL DYNAMICAL SYSTEM.)

Definition 3.17: *Let (X, d) be a compact metric space and T map X to itself continuously. We say (X, T) is* **uniquely ergodic** *(u.e.) if $|\mathcal{I}| = 1$, i.e., if there is only one T–invariant measure.*

Note that by Theorem 3.14, it is necessarily ergodic. Before discussing the implications of this important notion, we give an example which will also be used later in the proof of Weyl's Theorem.

Theorem 3.18: *An irrational rotation of the circle is u.e..*

Proof: We assume from Fourier analysis that a measure is determined by its Fourier coefficients. First let μ be any measure, not necessarily invariant and let $T\mu$ be defined by $T\mu(A) = \mu(T^{-1}(A))$. A trivial computation gives us that the nth Fourier coefficient of $T\mu$ is $e^{-2\pi in\theta}$ times the nth Fourier coefficient of μ where θ is the rotation angle.
EXERCISE: Show this fact.

Now, if $\mu \in \mathcal{I}$, then all the nonzero coefficients must be 0 since $e^{-2\pi in\theta}$ is never 1 for $n \neq 0$ as θ is irrational. It follows that μ must be Lebesgue measure. □

In the u.e. case, Theorem 3.16 can be strengthened very much.

Theorem 3.19: *If (X, T) is uniquely ergodic with unique invariant measure μ, then every x is generic for μ and the ergodic theorem holds uniformly in x in the sense that for every continous f,*

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \to \int f d\mu$$

uniformly in x.

Proof: We could first prove every point is generic for μ and then the uniformity in the ergodic theorem but clearly the latter implies the former.

If this uniformity failed for some f, then for this f, there would exist $\epsilon > 0$ such that for all n, there would exist $x(n)$ such that

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x(n))) - \int f d\mu| \geq \epsilon.$$

Let ν_n be $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{T^i(x(n))}$. Clearly $|\int f d\nu_n - \int f d\mu| \geq \epsilon$. Let ν be a weak limit of some subsequence of the ν_n. By Theorem 3.13, ν is invariant but clearly we also have $|\int f d\nu - \int f d\mu| \geq \epsilon$. Hence there is another invariant measure contradicting unique ergodicity. □

(We mention that the above method of proof can be used in other contexts as well, for example, in showing a uniform (in the initial configuration) of convergence of Feller Markov chains with compact state space when there is a unique stationary distribution and for similar results in Markov Random Fields).

The main theorem we are after is the following result due to Weyl.

Theorem 3.20: *If $p(x)$ is a polynomial with at least one coefficient other than the constant term irrational, then the sequence $p(n)$ is equidistributed (mod 1).*

Equidistributed means what you think it should, namely that

$$\frac{1}{n}\sum_{i=0}^{n-1}\delta_{p(n)} \to dx$$

where dx denotes Lebesgue measure.

This result will take a fair amount of development. We first need a general theorem which is an analogue of Theorem 3.9 in our measure setting. Before this, we need a lemma. But even before this, we need to quickly discuss the important concept of Haar measure. (If you prefer, you can skip Haar measure discussed below and assume all groups are either the circle or a torus where Haar measure is Lebesgue measure (e.g. arc length on the circle)).

....

HAAR MEASURE

It is a fact that every locally compact second countable group has a $\sigma-$ finite (possibly infinite) measure which is invariant under left translations. It is called a Left Haar Measure. Here are some facts, none of which we will prove.

1. There always exists a Left Haar Measure.
2. There always exists a Right Haar Measure.
3. These measures are (EACH) unique up to scaling.
4. These measures are finite if and only if the group is compact.
5. The Right and Left Haar Measures may be distinct (eg. $SL(2, R)$ which is the set of 2×2 real matrices with determinant 1).
6. If the group is compact, the Right and Left Haar Measures are the same. (If the group is abelian, they are obviously the same).

If the group is compact (as will always be the case for us), we will always assume that the Haar measure is normalized to be a probability measure.

....

Lemma 3.21: *Let T be a continuous map from the compact metric space X to itself and let μ be T–invariant. Let ψ take X to G continuously where G is a compact group with Haar measure dg. Then $\mu \times dg$ is an invariant measure for the group extension by ψ.*

Proof: This is left to the reader. Since each fiber is rotated and Haar measure is rotation invariant, this is reasonable- use Fubini's Theorem. □

Our main result is

Theorem 3.22: *Let T be a continuous map from the compact metric space X to itself which is uniquely ergodic with unique invariant measure μ. Let ψ take X to G continuously where G is a compact group with Haar measure*

dg. If $\mu \times dg$ is ergodic for the group extension by ψ (it is invariant by the previous lemma), then the group extension is uniquely ergodic.

....

At this point, I use an insert from Furstenberg's book. This will be 8 (but 4 for you) xeroxed pages.

HARD EXERCISE: Using what we have seen, show that if $A \subseteq \mathbf{N}$ has positive upper density which means that

$$\limsup_{n\to\infty} \frac{|A \cap [1,n]|}{n} > 0,$$

then there is a and b in A such that $a - b = n^2$ for some integer n.

4 The mean and pointwise ergodic theorems

In this section, by a dynamical system, we will mean a probabilty space with a measure preserving transformation to itself ($\mu(T^{-1}(A)) = \mu(A)$).

We first give the proof of von Neumann of the so-called mean ergodic theorem. This actually has nothing to do with ergodic theory but rather is a theorem about unitary (and in fact normal) operators.

Theorem 4.1: *Let T be a contraction on a complex Hilbert space H which is normal (this means T commutes with its adjoint, e.g., unitary and self–adjoint operators are normal. Normality is what one needs for the spectral theorem to hold although we won't need that now). Let P be the orthogonal projection onto the space of T–invariant functions (i.e., the kernel of $T - I$). Then for all $v \in H$,*

$$\frac{1}{n}\sum_{i=0}^{n-1} T^i(v) \to P(v)$$

as $n \to \infty$.

Lemma 4.2: *Let T be a bounded operator on a complex Hilbert space. Then $N(T^*)^{\perp} = \overline{R(T)}$ where N denotes the kernel of an operator and R denotes the image.*

Proof: This is left to the reader. There is no idea involved, just trivial manipulation. □

Lemma 4.3: *Let T be a bounded normal operator on a complex Hilbert space. Then $N(T) = N(T^*)$.*

Proof: $\|Tx\|^2 =< Tx, Tx >=< x, T^*Tx >=< x, TT^*x >=< T^*x, T^*x >= \|T^*x\|^2$. The result follows. □

Corollary 4.4: *Let T be a bounded normal operator on a complex Hilbert space H. Then $N(T) = R(T)^{\perp}$ (or equivalently $N(T)^{\perp} = \overline{R(T)}$).*

Proof of 4.1: Applying Corollary to the normal operator $T - I$, we have that $H = N(T - I) \oplus \overline{R(T)}$. It suffices to prove the result on each of these direct summands. This is an easy exercise left to the reader. □

What does this result have to do with ergodic theory? It is fairly obvious. Given a dynamical system, one obtains an induced operator on all the L_p spaces which are norm–preserving, $\|T(f)\| = \|f\|$ for all f where $\|\|$ denotes any L_p space you want. T is defined by $T(f) = f(T)$ of course. If the dynamical system is invertible, the above operators are also bijective and hence are isometries.
EXERCISE: Show these facts.

Ergodic theory is concerned with the asymptotic behavior of the averages

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

in various senses. This is just the average of f along orbits.

EXERCISE: Show that the mean-ergodic theorem implies that given a dynamical system and an $f \in L_2$, $A_n(f)$ converges in L_2 to the projection of f onto the subspace of T–invariant L_2 functions. (This is really just an observation.)

Ideally, one would want to obtain some type of pointwise convergence in the above. This is of course a much more delicate question just as is the case with L_2 convergence and pointwise convergence of Fourier series for L_2 functions.

This question of pointwise convergence was open for a while until Birkhoff finally solved it. One might wonder why the publication of this proof actually preceded slightly the publication of von Neumann's simpler mean ergodic theorem. It is also interesting to note that Birkhoff was the editor of the journal where von Neumann published his result. But I'm sure the above had nothing to do with each other.

Most of the standard proofs of the pointwise ergodic theorem require the development of certain maximal inequalities. Fortunately, in the last few years, a MUCH SIMPLER proof of this result has developed, which is quite surprising. Some feel that the original idea of this proof comes from nonstandard analysis in work by Kamae. Others however feel that this idea was already visible in some earlier work of Ornstein. Since the pointwise ergodic theorem implies the strong law of large numbers (SLLN) in probability theory, one might even argue that this proof is presently the simplest proof of the SLLN. The only things one needs to know is that $f \leq g$ and $\int f \geq \int g$ imply that $f = g$ a.e. (which of course is trivial) and the monotone convergence theorem (MCT). If one deals exclusively with bounded functions, one doesn't even need the MCT. The proof below really could be given in a page–the only reason it's longer it that I like to talk my way through a proof to give the ideas.

Theorem 4.5: *Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system and let $f \in L_1$. Then $A_n(f)(x)$ converges to a limit $f'(x)$ (do NOT think derivative) for* *a.e. x. Moreover, $f \in L_1$ and $\int f = \int f'$. If the system is ergodic, then $f' = \int f$a.e..*

Proof: For warm–up purposes, We first prove this in the case where $f = I_A$ is the indicator of some set A, ($f \equiv 1$ on A and $\equiv 0$ off A). In many instances, one can then pass to linear combinations and then use denseness arguments to then pass to all functions. One must be careful since while the above philosophy tends to work in functions spaces and with norm convergence, we want a pointwise result here. (Of course, one could never be able to do such a thing generally since one would then have a simple proof of pointwise convergence of Fourier Series for L_2 functions since this result is easy for differentiable functions.)

We want to show that

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

converges a.e.. Let

$$\overline{I_A} = \limsup \frac{1}{n}\sum_{i=0}^{n-1} I_A(T^i(x))$$

and

$$\underline{I_A} = \liminf \frac{1}{n}\sum_{i=0}^{n-1} I_A(T^i(x)).$$

Obviously, we want to show that $\overline{I_A} = \underline{I_A}$ a.e.. We will show that

$$\int \overline{I_A} \le \int I_A.$$

The reader should easily check for herself that applying the above to I_{A^c} gives

$$\int \underline{I_A} \ge \int I_A.$$

This would then give

$$\int \underline{I_A} \geq \int \overline{I_A}$$

which together with $\underline{I_A} \leq \overline{I_A}$ gives the desired result and also the fact that $\int I'_A = \int I_A$ where I'_A is the common limsup and liminf above.

We now show that

$$\int \overline{I_A} \leq \int I_A.$$

First note that $\overline{I_A}$ is a T–invariant function, i.e. $T(\overline{I_A}) = \overline{I_A}$ (this is obvious and easy but important). Let $\epsilon > 0$. Clearly, for every x, there exists $N(x)$ such that

$$A_{N(x)}(I_A)(x) \geq \overline{I_A}(x) - \epsilon.$$

Choose M such that

$$\mu(x : N(x) \leq M) \geq 1 - \epsilon.$$

So M is such that for all but ϵ portion of the points x in X, the average along the orbit of x gets within ϵ of its $\limsup$ by time M (but not necessarily at time M). Let $B = \{x : N(x) > M\}$ be the bad set where this fails. The key step is to prove that for all $n \geq M$,

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{A\cup B}(T^i(x)) \geq \frac{n-M}{n}(\overline{I_A}(x) - \epsilon).$$

There is no measure here. This is simply a combinatorial fact which is true FOR ALL x. We first finish the proof and then come back to this. Integrating both sides and using the fact that T is measure preserving (so that $\int f(T) = \int f$), we get

$$\mu(A \cup B) \geq \frac{n-M}{n}\int(\overline{I_A}(x) - \epsilon).$$

Letting $n \to \infty$, gives

$$\mu(A \cup B) \geq \int (\overline{I_A}(x) - \epsilon).$$

Then we get

$$\int \overline{I_A} \leq \mu(A \cup B) + \epsilon \leq \mu(A) + 2\epsilon = \int I_A + 2\epsilon.$$

As ϵ is arbitrary, we are done. The key inequality above is a picture proof which is very simple if you draw a picture and think about it.

To see this combinatorial fact, we break the orbit $\{x, T(x), \ldots, T^{n-1}(x)\}$ into segments as follows. First for any w, we let $N'(w)$ be $N(w)$ if $w \notin B$ and 1 if $w \in B$. We then take the first segment to be $\{x, T(x), \ldots, T^{N'(x)-1}(x)\}$. It is trivial to see that whether x is in B or not, the average of $I_{A\cup B}$ along this segment is larger than $(\overline{I_A}(x) - \epsilon)$. The second segment starts from $y = T^{N'(x)}(x)$ and is just $\{y, T(y), \ldots, T^{N'(y)-1}(y)\}$. The average of $I_{A\cup B}$ along this second segment is $\geq (\overline{I_A}(y) - \epsilon)$ and by noting that $\overline{I_A}(x) = \overline{I_A}(y)$, it is $\geq (\overline{I_A}(x) - \epsilon)$. As long as the segments begin below $n - M$, we're ok. However, when we enter $\{n - M, \ldots, n\}$, our point z might not be in B but the sequence $\{z, T(z), \ldots, T^{N(z)-1}(z)\}$ might bring us past n. This problem is taken care of however by the term $\frac{n-M}{n}$. This proves the result when $f = I_A$.

We now want to do general f. You can't really do any approximation but rather must repeat the above argument and modify it.

EXERCISE: The reader should now modify the above proof and prove the theorem in the case of BOUNDED f. This is a good exercise in making sure you understand the above proof. If you can go to all f, great, but I'll do that now. The basic problem that arises in the general case which does not arise in the case of bounded f is that when you enter the bad set B, you have no a priori bounds on how bad your average can get messed up.

GENERAL f

By breaking f into its positive and negative parts, it suffices to show the a.e. convergence for the case $f \geq 0$. Let

$$\overline{f} = \limsup \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$$

and

$$\underline{f} = \liminf \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)).$$

We first show that

$$\int \overline{f} \leq \int f.$$

We do this by showing that for all K ($a \wedge b = \min(a, b)$ here),

$$\int \overline{f} \wedge K \leq \int f$$

and then simply applying the Monotone Convergence Theorem with $K \to \infty$. Letting $\epsilon > 0$, it suffices to show that

$$\int \overline{f} \wedge K \leq \int f + \epsilon + K\epsilon$$

as K is fixed. Let $N(x)$ be such that

$$A_{N(x)}(f)(x) \geq \overline{f}(x) - \epsilon.$$

Choose M such that

$$\mu(x : N(x) \leq M) \geq 1 - \epsilon.$$

Let $B = \{x : N(x) > M\}$ be the bad set where this fails. The key combinatorial fact this time is is to prove that for all $n \geq M$,

$$\frac{1}{n}\sum_{i=0}^{n-1}\max\{f(T^i(x)), KI_{T^i(x)\in B}\} \geq \frac{n-M}{n}((\overline{f}(x)\wedge K)-\epsilon).$$

This fact is proved like the previous one and is therefore omitted. Similar to the above, replacing the max by a sum, integrating, and using $\mu(B) < \epsilon$, gives

$$\int f + K\epsilon \geq (\frac{n-M}{n})(\int(\overline{f}\wedge K)-\epsilon).$$

Letting $n \to \infty$ gives us the desired inequality.

We have now shown that

$$\int \overline{f} \leq \int f.$$

We might now say that similarly

$$\int \underline{f} \geq \int f$$

but the careful reader should object to this. If f (which we are assuming to be positive) were bounded by say L, then there is no problem since we could apply what we did to $L - f$ to obtain

$$\int \overline{L-f} \leq \int (L-f)$$

which gives with a little thought

$$L - \int \underline{f} \leq L - \int f$$

which implies

$$\int \underline{f} \geq \int f$$

as we wanted. However, for general f, one needs to say the following. First for any M (this is a different M than before),

$$\int \underline{f} \geq \int \underline{f \wedge M}.$$

BE CAREFUL!! $\underline{f \wedge M} \neq \underline{f} \wedge M$ so you need to think what everything means. By what we proved above for bounded functions,

$$\int \underline{f} \geq \int \underline{f \wedge M} \geq \int f \wedge M.$$

Now let $M \to \infty$ using MCT and conclude that $\int \underline{f} \geq \int f$ as desired.

Note that letting f' denote the limit we now proved exists, obviously $\int f' = \int f$. This shows that f' is in L_1 if f is positive. For general f, writing $f = f_+ - f_-$ with the latter two functions positive, clearly $f' = f'_+ - f'_-$. This shows that f' is in L_1 and that $\int f = \int f'$.

Clearly, if the dynamical system is ergodic, then f' being an invariant function must be constant a.e. and therefore must be $\equiv \int f$ a.e.. $\square$

We now prove that the convergence in the ergodic theorem holds in L_1 in addition to a.e.. The proof is very soft and quite easy.

Theorem 4.6: *$A_n(f) \to f'$ in L_1.*

Proof: Letting $\mathcal{A}$ be the set of L_1 functions for which we have L_1 convergence in the ergodic theorem, it suffices to show that $\mathcal{A}$ is closed in L_1 since L_∞ is trivially contained in $\mathcal{A}$ (by bounded convergence) and L_∞ is of course dense in L_1.

Letting F denote the map which takes f to f' (which is obviously linear), we first show that F is a contraction (also called a nonexpansive map) of L_1 (this means that $\|F(f)\| \leq \|f\|$ for all f). For any $f \in L_1$,

$$\int |F(f)| \leq \int F(|f|) = \int |f|,$$

the equality following from part of the ergodic theorem.

Now let g be in L_1 and f in $\mathcal{A}$ close to g (in the L_1 sense of course). Then

$$\int |A_n(g)-F(g)| \leq \int |A_n(g)-A_n(f)|+\int |A_n(f)-F(f)|+\int |F(f)-F(g)|.$$

The first term is small for all n (this is just the triangle inequality together with the measure preserving property), the second term is small for large n by assumption, and the third term is small as F is a contraction. Hence $g \in \mathcal{A}$ and we're done. $\square$

Note the while F is a contraction, it is VERY FAR from being norm–preserving. For example, if the dynamical system is ergodic, then the kernel of T has codimension 1. As an aside, we mention that f' is also the conditional expectation given the invariant σ–field. If you don't know what this means, it doesn't matter.

EXERCISE: Show that F is a contraction on each L_p space for $1 \leq p \leq \infty$ and that $A_n(f) \to F(f)$ in L_p for all $1 \leq p < \infty$. Why does the latter fail for L_∞? (Remember, our measure space is finite and so all the L_p spaces are contained in L_1.)

EXERCISE: Give an alternative proof of the L_1 convergence in the ergodic theorem by demonstating uniform integrability (u.i.). (You know in fact that we must have u.i. since it is equivalent to L_1 convergence.) Hint: Show that all the fT^i have the same distribution (the distribution of a function is the measure induced on the real line by pushing forward the measure on the measure space via fT^i).

The rest of this section is a digression which you should feel free to skip but does allow one to see how one puts (some of) probability theory into the context of ergodic theory.

STRONG LAW OF LARGE NUMBERS (SLLN)

How does one prove the SLLN which has to do with chance from a theorem concerning the evolution of a deterministic transformation? It is all fairly simple and soft, the only real content is proving the ergodicity assumption when we construct our dynamical system.

Theorem 4.7 (STRONG LAW OF LARGE NUMBERS): *Let $\{X_i\}_{i=0}^{\infty}$ be independent and identically distributed random variables with common finite mean m. Then $\frac{1}{n}\sum_{i=1}^{n} X_i \to m$ as $n \to \infty$ a.s..*

[Here is what this means to people who don't know probability theory:

We have a probability space (a measure space with total measure 1) and a family of integrable functions X_i defined on it. Clearly such a function (which we (the probabilists) call a random variable) can be used to push the measure from the probability space to the real line (the measure of the Borel set A in the real line is simply the measure in the probability space of the inverse image under the function in question of the set A). We call this probability measure the "distribution of the random variable". "Identically distributed" means all the measures obtained from different X_i's are the same. Independence means for any finite set of n of these X_i's, the measure on R^n obtained by this n–vector (again by pushing forward the measure) is a product of the 1–dimensional measures. The above is all mathematics and doesn't say what this has to do with chance but we leave it here.]

Let N be the positive integers and let ν be the measure on R^N corresponding to the distribution of the X_i's. (The X_i's map our probability space to R^N in the obvious way ($\omega \to (X_1(\omega), X_2(\omega), \ldots)$) and ν is just the measure on the probability space pushed forward to R^N. By assumption, ν is a product measure all of whose 1-dimensional marginals are the same. Let T be the transformation on R^N given by

$$T(w_1, w_2, \ldots) = (w_2, w_3, \ldots)$$

which is clearly noninvertible. Then (R^N, ν, T) (the σ–algebra here is obvious) is a dynamical system.

EXERCISE: Show T preserves ν.

Now let g be the map from R^N to R given by

$$g(w_1, w_2, \ldots,) = w_1.$$

EXERCISE: $\int g d\nu = m$.

The ergodic theorem tells us

$$\frac{1}{n}\sum_0^{n-1} g(T^i(w_1, w_2, \ldots)) \to g'$$

for ν a.e. w with $\int g' d\nu = m$. This means that

$$\frac{1}{n}\sum_1^n w_i \to g'$$

for ν a.e. w. If T were ergodic, this would give us

$$\frac{1}{n}\sum_1^n w_i \to m$$

for ν a.e. w.

EXERCISE: Show that this then gives us that $\frac{1}{n}\sum_{i=1}^n X_i \to m$ as $n \to \infty$ a.s. as desired.

The above was somewhat formal but the meat (although it's not hard) is to prove the ergodicity assumption. The proof is a wonderful picture proof.

Let B be invariant but have ν measure different from 0 and 1. Find A such that $\nu(A \triangle B)$ is small and A depends only on finitely many coordinates. (This can be done from general measure theory and does not need a product measure to be true. One shows that the set of such B which can be so approximated is a σ–algebra. Since it clearly contains the sets which only depend on finitely many coordinates, it's everything.)

Now choose n so large that A and $T^{-n}(A)$ depend on different coordinates. As T preserves measure, $\nu(T^{-n}(A) \triangle T^{-n}(B))$ is small and so $\nu(T^{-n}(A) \triangle B)$ is small as B is invariant. This gives $\nu(T^{-n}(A) \triangle A)$ is also small. But (now we use the fact that the measure is product measure) A and $T^{-n}(A)$ are independent sets and independent sets don't have $\nu(T^{-n}(A) \triangle A)$ being small.

EXERCISE: Draw pictures to convince yourself of the above and if unconvinced, fill in the ϵ's and δ's yourself. $\Box$

5 Continued Fraction Expansions and Gauss measure

The ergodic theorem (or the strong law of large numbers) tells us that a.e. number in $[0, 1]$ (wrt Lebesgue measure) has a binary expansion where the portion of 1's is $1/2$ in the sense that the portion of 1's in the first n bits tends to $1/2$ as $n \to \infty$. (Check this.)

This might seem nice but one could criticize this by saying that it is unnatural to look at binary expansions in the first place since they are not intrinsic as they completely depend on the base of expansion you happen to be looking at.

The continued fraction expansion on the other hand is completely natural as it is algebraically defined and does not depend on some arbitrary base of expansion. Every number has its so–called continued fraction expansion.

....

A QUICK ASIDE.

The following has nothing to do with what we'll be doing but it's an interesting fact that someone pointed out to me a few years ago.

First, it's a general fact from point set topology that if you have a complete metric space and take a subset which is a countable intersection of open sets (such a set is called a G_δ set), then this set is what is called topologically complete. This means that although it might not be complete in the initial metric, there is another metric WHICH GENERATES THE SAME TOPOLOGY in which the subset is complete (i.e., it means that you are homeomorphic to a complete metric space). This is a useful thing to know since one can then apply Baire category arguments since clearly such arguments are topological statements and hence only require the space in question to be topologically complete.

As a special case, this tells us that the irrationals in $[0, 1]$ are topologially complete (how about the rationals?). What does this have to do with continued fractions? The point is that the continued fraction expansion allows us to CONCRETELY find this wierd metric on the irrationals which is complete. How do we do this?

The continued fraction expansion gives us a bijection from the irrationals to sequences of POSITIVE integers. It is easy to see that when one places the product topology on this sequence space (with the discrete topology on $\mathbf{N}$ of course), this bijection becomes a homeomorphism. But there is an easy way to put a complete metric on the product space (How?) and now one just pulls back this metric via the bijection to the irrationals, giving the desired complete metric.

....

What we want to investigate is the following type of question. How many 7's are there in the continued fraction expansion of a "typical" number in $[0, 1]$? We will answer this question using ergodic theory and also discuss a result that Gauss wrote in a letter to Laplace with no proof–it is unclear what proof he had in mind.

We will use the notation

$$w = [a_1(w), a_2(w), \ldots]$$

to mean the usual continued fraction expansion (see later xeroxed pages to see what this means). This means in particular that $a_1(w) = \lfloor \frac{1}{w} \rfloor$. Now letting,

$$w' = [a_2(w), a_3(w), \ldots]$$

we have $w = \frac{1}{a_1+w'}$ or $w' = 1/w - a_1$ which we denote by $\{\frac{1}{w}\}$ which is the fractional part of $1/w$. This implies that

$$T(w) = \{\frac{1}{w}\}$$

brings the shift map back to the unit interval. This is expressed in the following lemma which follows by induction. For $w = 0$, we take $T(w) = 0$.

Lemma 5.1: If $a(w) = \lfloor \frac{1}{w} \rfloor$ (with $a(0) = \infty$), then $a_n(w) = a(T^{n-1}(w))$ for $n = 1, 2, \ldots$.

We now introduce integer valued functions $p_n(w)$ and $q_n(w)$ as follows.

$$p_{-1}(w) = 1, p_0(w) = 0, p_n(w) = a_n(w)p_{n-1}(w) + p_{n-2}(w)$$

$$q_{-1}(w) = 0, q_0(w) = 1, q_n(w) = a_n(w)q_{n-1}(w) + q_{n-2}(w).$$

Lemma 5.2:

1. $w = [a_1(w), a_2(w), \ldots, a_{n-1}(w), a_n(w) + T^n(w), 0, 0, \ldots]$.

2. $p_{n-1}(w)q_n(w) - p_n(w)q_{n-1}(w) = (-1)^n, n \geq 0$.

3. $[a_1(w), a_2(w), \ldots, a_{n-1}(w), a_n(w) + t, 0, 0, \ldots] = \frac{p_n(w)+tp_{n-1}(w)}{q_n(w)+tq_{n-1}(w)}$, $n \geq 1, 0 \leq t \leq 1$.

Proof: Easy induction which is left to the reader. □

Corollary 5.3: $[a_1, a_2, \ldots, a_{n-1}, a_n, 0, 0, \ldots] = \frac{p_n(w)}{q_n(w)}, n \geq 1, 0 \leq t \leq 1$.

Corollary 5.4: $w = \frac{p_n(w)+T^n(w)p_{n-1}(w)}{q_n(w)+T^n(w)q_{n-1}(w)}, n \geq 1$.

Proposition 5.5:

$$\frac{1}{q_n(w)(q_n(w) + q_{n+1}(w))} \leq |w - \frac{p_n(w)}{q_n(w)}| \leq \frac{1}{q_n(w)q_{n+1}(w)}.$$

Proof:

$$|w - \frac{p_n(w)}{q_n(w)}| = |\frac{p_n(w) + T^n(w)p_{n-1}(w)}{q_n(w) + T^n(w)q_{n-1}(w)} - \frac{p_n(w)}{q_n(w)}| =$$
$$\text{(by 2 above and algebra)}\ \frac{1}{q_n(w)((T^n(w))^{-1}q_n(w) + q_{n-1}(w))}.$$

Now, as $a_{n+1}(w) = \lfloor \frac{1}{T^n(w)} \rfloor$, we have that $a_{n+1}(w) \leq \frac{1}{T^n(w)} \leq a_{n+1}(w) + 1$. Using the recurrence of $q_n(w)$ gives the result. □

Given positive integers $a_1, a_2, \ldots, a_n$, let $\Delta_{a_1,\ldots,a_n} = \{w : a_1(w) = a_1, \ldots, a_n(w) = a_n\}$. This is called a fundamental interval of rank n (think of the analogy with binary expansions).

Letting $\psi_{a_1,\ldots,a_n}(t)$ be $[a_1, a_2, \ldots, a_{n-1}, a_n + t, 0, 0, \ldots]$, this function is increasing in t for even n and decreasing in t for odd n. This is nothing but $\frac{p_n+tp_{n-1}}{q_n+tq_{n-1}}$ where p_n and q_n are given by the a_n's using the same recursion formulas that defined the $p_n(w)$'s and $q_n(w)$'s.

Note that $\Delta_{a_1,\ldots,a_n}$ is just the image of $\psi_{a_1,\ldots,a_n}(t)$ which is $[\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}]$ (if n is even) and $[\frac{p_n+p_{n-1}}{q_n+q_{n-1}}, \frac{p_n}{q_n}]$ (if n is odd).

The Lebesgue measure of such an interval is (doing the exact same computation as in the previous proof) $\frac{1}{q_n(q_n+q_{n-1})}$.

Clearly the q_n's increase and so these intervals have small length and so every Borel set can be approximated (in the sense of Lebesgue measure) by finite unions of fundamental intervals.

Now, to do ergodic theory, we want an invariant measure for T. Moreover, if we want to make statements about (Lebesgue or dx) typical points, we would need to have an invariant measure which is equivalent to dx. dx is not invariant but the Gauss measure dG is T–invariant where the Gauss measure has Radon–Nikodym derivative (wrt dx)

$$\frac{1}{\ln 2}\frac{1}{1+x}.$$

Note that this derivative is uniformly bounded away from 0 and ∞.

Our goal is to show that $([0,1], T, dG)$ has strong mixing properties. Ergodicity (which we will prove) yields interesting facts (see the xeroxed notes) but the stronger property called "mixing" which is defined by

$$dG(A \cap T^{-n}(B)) \to dG(A)dG(B) \text{ as } n \to \infty$$

for all sets A and B will prove the result that Gauss mentioned without proof in a letter to Laplace.

EXERCISE: Show that mixing always implies ergodic.

We need to do some general things first. If a general dynamical system (X, μ, T) is ergodic, we know that

$$\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i(x)} \to \mu$$

for μ a.e. point. (Here we are assuming that T is a continuous mapping on the compact metric space X which is of course not the situation we actually are in with our continued fraction expansions). This gives easily (check this) that for all $\nu \leq \mu$,

$$\frac{1}{n}\sum_{i=0}^{n-1} T^n \nu \to \mu$$

where $T^n\nu(A) = \nu(T^{-n}(A))$.

Can one conclude the stronger fact that $T^n\nu \to \mu$ as $n \to \infty$ for ν which are absolutely continuous with respect to μ?

EXERCISE: Not necessarily (irrational rotation of the circle).

The next proposition tells us that if rather than just having ergodicity we have the stronger property of mixing, then we can obtain this stronger result.

Proposition 5.6: If (X, μ, T) is mixing and $\nu \leq \mu$ with $0 < c_1 < \frac{d\nu}{d\mu} < c_2$, then

$$T^n(\nu) \to \mu$$

as $n \to \infty$ in the sense that for all bounded functions f

$$\int f dT^n(\nu) \to \int f d\mu.$$

[There is a stronger version of this result with a weaker assumption and a stronger conclusion. For simplicity however, we do the above instead.]

Proof: First check (via approximation) that mixing is equivalent to

$$\int (T^n f) g d\mu \to \int f d\mu \int g d\mu$$

for all bounded f and g. Then for all bounded f

$$\int f dT^n(\nu) = \int T^n(f) d\nu = \int (T^n(f))(\frac{d\nu}{d\mu}) d\mu$$

which by above

$$\to \int f d\mu \int (\frac{d\nu}{d\mu}) d\mu = \int f d\mu.$$

□

We shall later show that our c.f.e. system is mixing. Letting $\mu = dG$, $\nu = dx$ and $f = I_{[0,x]}$ in Proposition 5.6 gives us the following result which was in Gauss's letter to Laplace.

Corollary 5.7: $\lim_{n\to\infty} dx(\{w : T^n(w) \le x\}) \to dG([0,x]) = \frac{1}{\ln 2}\int_0^x \frac{1}{1+t} dt$.

It suffices to demonstrate the desired mixing conditions. Before doing this, we introduce a mixing condition which is even stronger than mixing. Let $\mathcal{G}_n$ be the σ–algebra generated by the functions $a_n(w), a_{n+1}(w), \ldots$ and let $\mathcal{G}_\infty = \cap_{n=1}^{\infty} \mathcal{G}_n$. It turns out (as we will show later) however that $\mathcal{G}_\infty$ is trivial in the sense that any set in this σ–algebra has measure 0 or 1. [Things with this property are called Kolmogorov automorphisms. Looking at so–called "tail σ–fields" is very common in probability theory.] While it is not obvious that this condition of $\mathcal{G}_\infty$ being trivial implies mixing, we will later say a word explaining why it does (and it is in fact strictly stronger although it takes some work to get an example to show this).

EXERCISE: Show that $\mathcal{G}_\infty$ being trivial implies ergodicity.

Theorem 5.8: $([0,1], dG, T)$ **is ergodic and** $\mathcal{G}_\infty$ **is trivial (and hence the system is mixing).**

The key lemma for this theorem is the following.

Lemma 5.9: For all Borel sets A **and for all integers** $a_1, a_2, \ldots, a_n$**,**

$$\frac{1}{2}dx(A) \le dx(T^{-n}(A)|\Delta_{a_1,\ldots,a_n}) \le 2dx(A).$$

Proof of Theorem 5.8: As $c_1 \le \frac{dG}{dx} \le c_2$, Lemma 5.9 gives us that

$$\frac{1}{c}dG(A) \le dG(T^{-n}(A)|\Delta_{a_1,\ldots,a_n}) \le c dG(A)$$

for some universal constant c.

Ergodicity: Assume that $T^{-1}(A) = A$ and $0 < dG(A) < 1$. Then

$$\frac{1}{c}dG(A) \le dG(A|\Delta_{a_1,\ldots,a_n})$$

which implies

$$\frac{1}{c}dG(\Delta_{a_1,\ldots,a_n}) \le dG(\Delta_{a_1,\ldots,a_n}|A)$$

which implies by approximation

$$\frac{1}{c}dG(B) \le dG(B|A)$$

for all sets B. Taking $B = A^c$ gives a contradiction.

$\mathcal{G}_\infty$ is trivial: Let $A \in \mathcal{G}_\infty$ with $0 < dG(A) < 1$. For all n, there exists C such that $A = T^{-n}(C)$. Then

$$\frac{1}{c} dG(A) = \frac{1}{c} dG(C) \leq dG(A|\Delta_{a_1,\ldots,a_n}).$$

Now follow the proof of ergodicity. $\square$

Proof of Theorem 5.9: It suffices to prove this for $A = [x, y)$ So fix $A = [x, y)$ and $a_1, a_2, \ldots, a_n$ and let $\psi = \psi_{a_1,\ldots,a_n}(t)$. We know that $\Delta_{a_1,\ldots,a_n}$ has length $\pm\psi(1) - \psi(0)$. Similarly, using the first statement of Lemma 5.2, the interval $\{w : x \leq T^n(w) \leq y\} \cap \Delta_{a_1,\ldots,a_n}$ has length $\pm\psi(y) - \psi(x)$. This implies

$$dx(T^{-n}([x, y))|\Delta_{a_1,\ldots,a_n}) = \frac{\psi(y) - \psi(x)}{\psi(1) - \psi(0)} =$$

$$\frac{\frac{p_n + yp_{n-1}}{q_n + yq_{n-1}} - \frac{p_n + xp_{n-1}}{q_n + xq_{n-1}}}{\frac{p_n + p_{n-1}}{q_n + q_{n-1}} - \frac{p_n}{q_n}} = (y - x) \frac{q_n(q_n + q_{n-1})}{(q_n + xq_{n-1})(q_n + yq_{n-1})}.$$

This last step is obtained by easy algebra similar to what we have done before. Since the q_n's are increasing, the right hand side is always between 1/2 and 2 which proves the lemma. $\square$

The fact that $\mathcal{G}_\infty$ trivial implies mixing is a standard fact in probability theory and ergodic theory. The usual proof of this uses something called the Backwards Martingale Convergence Theorem. I won't do this but there are many references.

....

At this point, we again use another insert, this time from Billingsley's book

"Ergodic Theory and Information". This insert is only given to provide further reading to someone who is interested.

6 Halmos-von Neumann Theorem

The isomorphism problem in general is very difficult. There are two cases in which it has been solved. The first case is when the induced unitary operators have pure point spectrum which means that the set of eigenvectors (necessarily an orthonormal set) span the entire L_2 space. This is the subject of the present section. The second case which we will not do and is much more difficult will be briefly described in the next section.

There are two results in this section. First, we show that for dynamical systems with pure point spectrum, having equivalent unitary operators implies that the dynamical systems are isomorphic. (The converse is always true. Why?) Secondly, we show that any such system is isomorphic to a rotation on some compact abelian group (the group actually being the dual group of the set of eigenvalues). One sometimes calls the latter quasiperiodic or almost periodic motion. This latter result gives us a representation theorem for dynamical systems with pure point spectrum.

At this point, we only deal with ergodic dynamical systems. Perhaps, here is a good place to look more closely at the operators induced on the different L_p spaces by our transformation.

EXERCISE: Check that the induced operator is always norm preserving.

Unfortunately, for infinite dimensional spaces, one can have norm preserving maps which are not surjective (although they are trivially injective).

EXERCISE: Consider the set of all (1–sided) sequences of 1's and 0's where each bit is independent 0 or 1 with probability 1/2. Consider the so-called shift map which moves the sequence to the left and cuts off the first bit. Check the measure is preserved and that the induced operator on L_2 is not surjective. You should recognize this dynamical system in a completely different way. This is the same as the unit interval with Lebesgue measure and the transformation being multiplication by 2 (mod 1). (To see this, just think binary expansion.)

Theorem 6.1: A dynamical system is invertible if and only if the induced unitary operator (or any of the isometries on the different L_p spaces) is surjective.

This proof will be left to the student. It really is not hard but one has to be careful with what one means. Certainly, we all know what it means for the operator to be surjective. But invertibility of the dynamical system is a more delicate matter and one can quickly run into measure–theoretic difficulties. Since sets of measure 0 are of no importance, one should not require that T as a map be invertible. The right thing to do is technical but not hard. T induces a mapping on the σ–algebra by $A \to T^{-1}(A)$. We say T is surjective if this induced map is surjective in that for all sets B there is an A such that B and $T^{-1}(A)$ are the same set up to measure 0. You can read all about this and more in [Wal] OR YOU CAN SKIP THIS LAST PARAGRAPH AND GO ON WITH NO PROBLEM.

We want to understand the discrete spectrum of the induced unitary operator. In simpler terms, what can we say about the eigenvalues of our operator?

EXERCISE. Check that ergodicity of a dynamical system is equivalent to the eigenspace of 1 being 1-dimensional (in which case this space is obviously the constants).

Lemma 6.2: Consider an ergodic dynamical system.

1. If $U(f) = \lambda f$ (i.e., λ is an eigenvalue with eigenvector f), then $|\lambda| = 1$ and $|f|$ is constant.

2. The eigenspace for any eigenvalue λ is 1-dimensional.

3. The set of eigenvalues form a subgroup of the circle group.

Proof: 1. Eigenvalues (in fact the entire spectrum if you know what that is) for any unitary operator always sit on the unit circle in the complex plane. The proof is easy and is as follows. $\|f\| = \|U(f)\| = \|\lambda f\| = |\lambda|\|f\|$ which implies $|\lambda| = 1$ as f is not 0. (This of course needs no ergodicity). Next, $U(|f|) = |U(f)| = |\lambda f| = |f|$ and so $|f|$ is constant by ergodicity since it is an invariant function.

2. Note by definition of U, it preserves products (real products, not the inner product). If f and g are in the same eigenspace, then $U(f/g) = f/g$ and so by ergodicity f/g is constant and so the eigenspace is 1-dimensional. (Note that by Part 1, $|g|$ is constant and so we can divide by it).

3. Let a and b be eigenvalues with eigenfunctions f and g. Then it is trivial to check that fg' (g' is the complex conjugate of g) is an eigenfunction for ab^{-1} and so they are a group. $\square$

The next result is from group theory but I would guess has not been seen by most people. It has the flavor of topology and if anyone knows an analogous result in topology and knows what plays the role of divisibility below, please tell me.

The ? is given a subgoup H of G, when does there exist a mapping from G to H which is the identity on H. This is called a retract which is the right word from topology to describe this situation.

Note that for the finite groups Z_2 and Z_4, this is not possible. There is a nice class however where this is possible.

Definition 6.3: *A group H is divisible if every element as an nth root for every n.*

For instance, the rationals are but remember an nth root here means dividing by n.

Theorem 6.4: *Let G be abelian and H a subgroup of G which is divisible. Then there is a retract from G to H as above.*

It is easy to see that the existence of a retract is equivalent to saying that H is a direct summand, i.e., there is another subgroup K of G such that G is the internal direct sum of H and K. If you remember your algebra, Theorem 6.4 says that a divisible group is an injective **Z**–module. The converse is also true.

....

At this point, we again use an insert, this time from Walter's book "An Introduction to Ergodic Theory".

7 Entropy and Bernoulli Shifts

The following will be a quick description of another class of systems in which the isomorphism problem has been solved. In 1959, Kolmogorov introduced an important invariant, this invariant being a specific number assigned to a dynamical system which one calls the entropy. It measures the amount of "randomness" or "chaos" and is related to Lyaponov exponents if you have heard of these. This notion of entropy also plays a large role in information theory and data compression. In fact, Kolmogorov took the idea of entropy from information theory and introduced it into ergodic theory.

It turns out there is an important large class of systems for which entropy is a complete invariant–these are the so–called Bernoulli shifts which were discussed in §2. They are from the point of view of probability simply independent and identically distributed random processes. This result is very deep and due to Donald Ornstein.

There are a number of "concrete" or "physical" systems which have been proven to be in this class- for example, the geodesic flow on a manifold of constant negative curvature. Here the space is the set of points on the manifold together with a direction tangential to the manifold (the unit tangent bundle in fancier language). The group action is $\mathbf{R}$ and the transformation is to "follow a given direction along the geodesic".

This section would be a course in itself and so I will say no more.

REFERENCES

[Arv] Arveson, William, *An Invitation to C^*–Algebras,* Springer-Verlag, New York 1976.

[Bill] Billingsley, Patrick, *Convergence of Probability Measures,* John Wiley & Sons, New York 1968.

[D+S] Dunford, N. and Schwartz, J., *Linear Operators Part I: General Theory,* Interscience, New York 1957.

[E+K] Ethier, Stewart N. and Kurtz, Thomas G., *Markov Processes–Characterization and Convergence,* John Wiley & Sons, New York 1986.

[Feld] Feldman, J., *r–Entropy, Equipartition, and Ornstein's Isomorphism Theorem in* $\mathbf{R}^n$, Israel Journal of Mathematics **36** (1980), 321–343.

[Furst] Furstenburg, H., *Recurrence in Ergodic Theory and Combinatorial Number Theory,* Princeton University Press, Princeton 1981.

[HL] Hardy, G. H. and Littlewood, J. E., The fractional part of $n^k\theta$, Acta Math. **37** (1914), 155–191.

[Lam] Lamperti, John, *Stochastic Processes: A Survey of the Mathematical Theory,* Springer-Verlag, New York 1977.

[Lig] Liggett, Thomas M., *Interacting Particle Systems,* Springer-Verlag, 1985.

[Lind] Lind, D. A., *The Isomorphism Theorem for Multidimensional Bernoulli Flows*, Unpublished preprint.

[Mesh] Meshalkin, L. D., *A Case of Isomorphism of Bernoulli Schemes,* Dokl. Akad. Nauk. SSSR **128** (1959), 41–44.

[Nach] Nachbin, Leopold, *Topology and Order,* D. Van Nostrand, Princeton 1965.

[Nat] Natarajan, S., *An Exposition of Ornstein's Isomorphism Theorem,* Macmillan India, New Delhi 1982.

[O] Ornstein, Donald S., *Ergodic Theory, Randomness and Dynamical Systems,* Yale University Press, New Haven 1974.

[O+W1] Ornstein, Donald S. and Weiss, B., *Entropy and Isomorphism Theorems for Actions of Amenable Groups,* Journal d'Analyse Mathématique **48** (1987), 1–141.

[O+W2] Ornstein, Donald S. and Weiss, B., *Finitely Determined Implies Very Weak Bernoulli,* Israel Journal of Mathematics, **17** (1974), 94–104.

[Rosen] Rosenblatt, Murray, *Markov Processes. Structure and Asymptotic Behavior,* Springer-Verlag, New York 1971.

[Sh] Shields, Paul, *The Theory of Bernoulli Shifts,* University of Chicago Press, Chicago 1973.

[Wal] Walters, Peter, *An Introduction to Ergodic Theory,* Springer-Verlag, New York 1975.