ERGODIC THEORY

INTRODUCTION

These notes grew out of a one semester course on Ergodic Theory in Indian Statistical Institute, Bengaluru. Several books have been used extensively. In particular, the books of Peter Walters and Karl Petersen need special mention.

Apart from a basic knowledge of Measure Theory (inlcuding complex measures and differentiation of measures) and Functional Analysis (including Banach - Alaoglu and Krein - Milman Theorems), Ergodic Theory requires many non-trivial results from many areas of Mathematics. To mention a few, isomorphism theorems of Measure Theory, existence and uniqueness of Haar measure and extension of characters on closed subgroups of locally compact Hausdorff topological groups to the whole groups are needed. We include these special topics in appendices, therby making the notes essentially self - contained. (It may be hard to cover the topics in appendices in a one semester course). The appendix on Isomorphism Theorems is based on Cohn"s Measure Theory and the one on Character Theory is based on Rudin"s Fourier analysis On Groups.

A table of contents appears on page 114

Let $(\Omega, \mathcal{F}, P)$ be a measure space and $T : \Omega \to \Omega$ be measurable. If $P \circ T^{-1} = P$ (i.e. $P(T^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$ then we say that T is a *measure preserving* (m.p.) transformation. We call $(\Omega, \mathcal{F}, P, T)$ a dynamical system (DS). For example $T(x) = a + x$ defines a m.p. transformation on $\mathbb{R}$ with the $\sigma-$ field of Lebesgue measurable sets and the Lebesgue measure. Though some of our results are true for arbitrary positive measures P we shall consider only m.p. transformations on probability spaces in these notes. Thus $P(\Omega)$ is *always* assumed to be 1.

Examples

Ex 0

Interval exchange transformations:

Consider the partition $\{[\frac{i-1}{n}, \frac{i}{n}) : 1 \leq i \leq n\}\}$ of $[0, 1)$. Define $T : [0.1) \to [0.1)$ by $T(x) = \frac{1}{n} + x$ if $x \in [\frac{i-1}{n}, \frac{i}{n})$ with $i < n$ and $T(x) = x - \frac{n-1}{n}$ if $x \in [\frac{n-1}{n}, \frac{n}{n})$. T is m.p. It is an "interval exchange transformation". We can also permute the intervals using an arbitrary permutation of $\{1, 2, ..., n\}$.

Ex 1.

Let $\alpha \in (0, 1), \Omega = (0, 1)$, $\mathcal{F} =$ Borel $\sigma-$ field and $P =$ Lebesgue measure. Let $Tx = x + \alpha (\mathrm{mod}\, 1)$. To see that this map is m.p. consider an interval $(a, b) \subset (0, \alpha)$. We have $T^{-1}((a, b)) = \{x : x + \alpha < 1 \text{ and } a < x + \alpha < b\} \cup \{x : x + \alpha \geq 1 \text{ and } a + 1 < x + \alpha < b + 1\} = (a - \alpha + 1, b - \alpha + 1)\}$ because

$b < \alpha$ and $a > 0$. Hence $P(T^{-1}(a,b)) = b - a$. Similarly if $(a,b) \subset [\alpha, 1)$ then $T^{-1}((a,b)) = (a-\alpha, b-\alpha)$ and $P(T^{-1}(a,b)) = b - a$. For an arbitrary interval $(a,b) \subset (0,1)$ we have $P(T^{-1}(a,b)) = P(T^{-1}(a,b) \cap (0,\alpha)) + P(T^{-1}(a,b) \cap [\alpha, 1)) = P((a,b) \cap (0,\alpha)) + P((a,b) \cap [\alpha,1))$ by what we already proved and hence T is m.p. We remark that this map is bijective and $T^{-1}x = x - \alpha (\text{mod}\, 1)$.

Ex.2

Let S^1 be the circle group and $a \in S^1$. Define $Tz = az$ for all $z \in S^1$. Let P be the normalized arc-length measure on the Borel sigma field of S^1. Then T is m.p. [T is rotation by angle θ if $a = e^{i\theta}$] This map is, in some sense, *equivalent* to the previous one. Equivalence of m.p. transformations will be discussed later.

In above examples T is also a bijection and T^{-1} is also m.p.. Such maps are called *invertible measure preserving* (i.m.p.)

Ex. 3

Let $Tx = 2x \,\text{mod}(1)$ on the space of Ex. 1

To show that this map is m.p. we compute $T^{-1}[\frac{i-1}{2^n}, \frac{i}{2^n})$. This set is the disjoint union of the intervals $[\frac{i-1}{2^{n+1}}, \frac{i}{2^{n+1}})$ and $[\frac{i-1}{2^{n+1}} + \frac{1}{2}, \frac{i}{2^{n+1}} + \frac{1}{2})$. Note that both these intervals are contained in $[0,1)$. Thus $PT^{-1}[\frac{i-1}{2^n}, \frac{i}{2^n}) = \frac{1}{2^{n+1}} + \frac{1}{2^{n+1}} = \frac{1}{2^n}$. Since dyadic intervals generate the Borel sigma field we have proved that T is m.p. [We can show, by the same method, that $Tx = kx$ is m.p. for any $k \in \mathbb{N}$. Note that if x is not a dyadic rational (so that it has a unique dyadic expansion) and its dyadic expansion is $\sum\limits_{k=1}^{\infty} \frac{a_k}{2^k}$ (with $a_k's \in \{0,1\}$) then $T(\sum\limits_{k=1}^{\infty} \frac{a_k}{2^k}) = \sum\limits_{k=2}^{\infty} \frac{a_k}{2^{k-1}}$ as seen easily by considering the cases $x < 1/2$ and $x \geq 1/2$. Iteration gives $T^m(\sum\limits_{k=1}^{\infty} \frac{a_k}{2^k}) = \sum\limits_{k=m+1}^{\infty} \frac{a_k}{2^{k-m}}$ from which we conclude that $a_{m+1} = [2T^m x]$. Thus the coefficients in the dyadic are given by the formula $a_m = [2T^{m-1}x], m = 1, 2, \ldots.$ Similar statements hold for $Tx = kx \,\text{mod}(1)$.

Let G be a locally compact Hausdorff topological group. Then there is a positive measure P on the Borel sigma field of G such that $P(gA) = P(A)$ for all Borel sets A and all $g \in G$ (where $gA = \{gh : h \in A\}$). Any two such measures differ only by a multiplicative constant. Such a measure is called a left-Haar measure. There is also a right-Haar measure [$P(Ag) = P(A)$]. If G is abelian then the two Haar measures coincide. If G is compact then also the two measures coincide up to a constant factor. The measures are finite in this case, so there is a unique probability measure P satisfying both of the above equations. In these notes all topological groups considered are compact we refer to the measure P as *the* Haar measure.

Ex. 4

Let G be a compact group and P the Haar measure. Let $g \in G$ and $Th = gh$ for all $h \in G$. Then $PT^{-1}(A) = P(g^{-1}A) = P(A)$ so $PT^{-1} = P$. Hence T is m.p.

Ex. 5

Let G be a compact group and P the Haar measure. Let T be a continuous automorphism of G. Then $PT^{-1}(gA) = P((T^{-1}g)(T^{-1}A)) = P(T^{-1}A)$ so PT^{-1} is also a Haar measure. Uniqueness of Haar measure implies that $PT^{-1} = P$. Hence T is m.p.

Theorem

The only continuous automorphisms of S^1 are the maps $z \to \frac{1}{z}$ and the identity map. The only continuous automorphisms of the Torus $S^1 \times S^1$ are of the type $T(a,b) = (a^j b^n, a^k b^m)$ where j, k, n, m are integers and $jm - kn = \pm 1$. All such maps are continuous automorphisms.

Remark: using the exercise below we show that any continuous homomorphism if S^1 is of the type $z \to z^n$ for some integer n. The only continuous homomorphisms of the Torus $S^1 \times S^1$ are of the type $T(a,b) = (a^j b^n, a^k b^m)$ where j, k, n, m are integers.

Proof:

Let T be a continuous automorphism of S^1. Let N be a positive integer. Define $\tau : [-N, N] \to S^1$ by $\tau(t) = T(e^{2\pi i t})$. A basic result in Complex Analysis says that this map has a continuous logarithm: there is a unique continuous map $\phi_N : [-N, N] \to S^1$ such that $\tau(t) = e^{\phi_N(t)}$ for all $t \in [-N, N]$ and $\phi_N(0) = 0$ [Note that $\tau(0) = T(1) = 1$]. Uniqueness of such a ϕ shows that $\phi_N' s$ are consistently defined and hence there is a unique continuous map $\phi_N : \mathbb{R} \to S^1$ such that $\tau(t) = e^{\phi(t)}$ for all $t \in \mathbb{R}$ and $\phi(0) = 0$. Now $e^{\phi(t+s)} = T(e^{2\pi i(t+s)}) = e^{\phi(t)} e^{\phi(s)}$ which implies $\phi(t+s) = \phi(t) + \phi(s)$. [The two sides differ by a constant of the type $2\pi i k$ by continuity and they both vanish at 0]. By an elementary result in analysis we conclude that $\operatorname{Re}\phi(t) = at$ for some $a \in \mathbb{R}$ and $\operatorname{Im}\phi(t) = bt$ for some $b \in \mathbb{R}$. We have proved that $T(e^{2\pi i t}) = e^{(a+ib)t}$. Necessarily $\left|e^{(a+ib)t}\right| = 1$ so $a = 0$. Since $e^{ib} = T(e^{2\pi i}) = T(1)$ and $e^0 = T(1)$ we get $b = 2j\pi$ for some integer j. Thus $T(z) = z^j$. Since T is one-to-one we get $j = \pm 1$.

Now let T be an automorphism of the torus $S^1 \times S^1$(which is a group under coordinatewise multiplication). Let $T_1(z)$ be the first coordinate of $T(z,1)$ and $T_2(z)$ be the second coordinate of $T(z,1)$. Let $T_3(z)$ be the first coordinate of $T(1,z)$ and $T_4(z)$ be the second coordinate of $T(1,z)$. Then T_j is a homomorphism of S^1 for $j = 1,2,3,4$. Hence there exist integers j, k, n, m such that $T_1(z) = z^j, v, T_2(z) = z^k, T_3(z) = z^n, T_4(z) = z^m$. It follows that $T(a,b) = T(a,1)T(1,b) = (a^j, a^k)(b^n, b^m) = (a^j b^n, a^k b^m)$. We have to determine when this map is an automorphism. If T is an automorphism the so is T^{-1} and so $T^{-1}(a,b) = (a^{j'} b^{n'}, a^{k'} b^{m'})$ for some integers j', n', k', m'. We now have

$(a,b) = TT^{-1}(a,b) = (a^{jj'+k'n}b^{jn'+m'n}, a^{j'k+k'm}b^{n'k+m'm})\forall a,b \in S^1$. This implies $jj'+k'n = 1, jn'+m'n = 0, j'k+k'm = 0$ and $n'k+m'm = 1$. In other words $\begin{pmatrix} n & j \\ m & k \end{pmatrix}\begin{pmatrix} k' & m' \\ j' & n' \end{pmatrix} = 1$. Taking determinants and noting that the determinants of the two matrices are integers we conclude that $nk - mj = \pm 1$. Conversely suppose $nk - mj = \pm 1$. The inverse of $\begin{pmatrix} n & j \\ m & k \end{pmatrix}$ has integer entries because the determinant is ± 1 and the adjoint has integer entries. Thus there is a transformation of the type $S(a,b) = (a^{j'}b^{n'}, a^{k'}b^{m'})$ with $TS = I = ST$. It follows that T is bijective with an inverse which is also a homomorphism. The inverse is automatically continuous.

Exercise:

Find all continuous maps $f : \mathbb{R} \to S^1 \equiv T$ such that $f(x+y) = f(x)f(y)$ for all x, y. Do the same when S^1 is replaced by $\mathbb{C}$. Also find all continuous homomorphisms of T.

Solution: first part: note that $f(0) = 1$. Fix a positive integer N. By a standard argument in Complex Analysis there exists a unique continuous function $h_N : [-N, N] \to \mathbb{R}$ such that $f(x) = e^{ih_N(x)}$ $(|x| \le N)$ and $h_N(0) = 1$. It follows easily that $h_N's$ define a continuous function $h : \mathbb{R} \to \mathbb{R}$ such that $h(0) = 0$ and $f(x) = e^{ih(x)}$ for all real numbers x. Note that $e^{i[h(a+b)-h(a)-h(b)]} = 1$ so $h(a+b) - h(a) - h(b) = 2n\pi$ for some integer n. By continuity of h we conclude that n does not depend on a and b. Since $h(0) = 0$ we conlude that h is additive.

Since h is additive and continuous there is a real number a such that $h(x) = ax$ for all x. Hence $f(x) = e^{iax}$. Any function of the type e^{iax} satisfies the given functional equation, so the first part is complete. Now consider the second part. Since $f(0) = f^2(0)$ either $f(0) = 0$ or $f(0) = 1$. If $f(x) = 0$ for some x then $f(x+y) = f(x)f(y) = 0$ for all y which gives $f \equiv 0$. If this is not the case then $f(0) = 1$ and f never vanishes. Let $g(x) = \frac{f(x)}{|f(x)|}$. The first part can be applied to g and we get $f(x) = e^{iax}|f(x)|$. Also $\log|f(x)|$ is an additive continuous function on $\mathbb{R}$, so $|f(x)| = e^{bx}$ for some real number b. We now have $f(x) = e^{(b+ia)x}$. Now let $\phi : T \to T$ be a continuous homomorphism and $f(x) = \phi(e^{2\pi ix})$. Then $f : \mathbb{R} \to T$ satisfies the equation $f(x+y) = f(x)f(y)$ and f is continuous. Hence $f(x) = e^{i2\pi ax}$ for some real number a. Thus $\phi(e^{2\pi ix}) = e^{i2\pi ax}$. Since the left side has the same value for $x = 0$ and $x = 1$ we see that a must be an integer. It follows that $\phi(z) = z^a$ $\forall z \in T$. Hence continuous homomorphisms of T are precisely the maps $z \to z^n$ where n is an integer. Note that such a map is injective if and only if $n = 1$ or $n = -1$. In other words, the only automorphisms of T that are continuous are the identity map and the map $z \to \frac{1}{z}$.

Exercise: show that continuous homomorphisms of $T^n (\equiv T \times T \times ... \times T$ (n-times)) are all of the type $(z_1, z_2, ..., z_n) \to (z_1^{a_{11}} z_2^{a_{12}} ...., z_n^{a_{1n}}, z_1^{a_{21}} z_2^{a_{2n}} ...., z_n^{a_{2n}}, ..., z_1^{a_{n1}} z_2^{a_{n2}} ...., z_n^{a_{nn}})$ where each a_{ij} is an integer.

This is an easy consequence of the previous exercise.

The next example is from Number Theory.

Ex. 6

Define $T : [0,1] \to [0,1]$ by $T(0) = 1$ and $T\omega = \frac{1}{\omega} - [\frac{1}{\omega}]$ if $\omega > 0$. If we provide $[0,1]$ with the Borel sigma field and the Lebesgue measure it turns out that T is not m.p.. However there is a Borel probability measure P such that $P \circ T^{-1} = P$ and P is equivalent to Lebesgue measure m (in the sense $P << m << P$). T is called the Gauss transformation and P is called the Gauss measure.

Let us begin by computing $T^{-1}(a,b)$ where $0 < a < b < 1$. We have $T^{-1}(a,b) = \bigcup_{n=1}^{\infty}\{\omega : \frac{1}{\omega} \in [n, n+1), \omega \in (\frac{1}{b+n}, \frac{1}{a+n})\} = \bigcup_{n=1}^{\infty}\{(\frac{1}{b+n}, \frac{1}{a+n}) \cap (\frac{1}{n+1}, \frac{1}{n}]\} = \bigcup_{n=1}^{\infty}(\frac{1}{b+n}, \frac{1}{a+n})$ since $(\frac{1}{b+n}, \frac{1}{a+n}) \subset (\frac{1}{n+1}, \frac{1}{n})$. Thus $m(T^{-1}((a,b))) = \sum_{n=1}^{\infty}\frac{b-a}{(a+n)(b+n)}$. If T is "Lebesgue measure preserving " then $\sum_{n=1}^{\infty}\frac{1}{(a+n)(b+n)} = 1$ whenever $0 < a < b < 1$. However the left side is strictly monotonic in a (and b).

Now let $P(A) = \frac{1}{\ln(2)}\int_A \frac{1}{1+x}dx$. We claim that $P \circ T^{-1} = P$. Recall that $T^{-1}(a,b) = \bigcup_{n=1}^{\infty}(\frac{1}{b+n}, \frac{1}{a+n})$ and the union here is a disjoint union. Hence it suffices to show that $\sum_{n=1}^{\infty}\frac{1}{\ln(2)}\int_{\frac{1}{b+n}}^{\frac{1}{a+n}}\frac{1}{1+x}dx = \int_a^b \frac{1}{\ln(2)}\frac{1}{1+x}dx$. It is easy to see that the (telescopic) product $\prod_{n=1}^{\infty}\frac{a+n+1}{b+n+1}\frac{b+n}{a+n}$ converges to $\frac{b+1}{a+1}$. The desired equation is obtained by taking logarithms on both sides.

Measure preserving transformations in Hamiltonian dynamics: the state of a system at time t is $(p_1(t), p_2(t), ..., p_n(t), q_1(t), q_2(t), ..., q_n(t))$ where the $p's$ are the positions on the n particles in the system and the $q's$ are the momenta of the particles. The motion is governed by the equations

$\frac{\partial q_i}{\partial t} = \frac{\partial H}{\partial p_i}, \frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q_i}, 1 \leq i \leq n$

where the Hamiltonian H is a function from $\mathbb{R}^{2n}$ to $\mathbb{R}$. Define $T_t, t \geq 0$ on $\mathbb{R}^{2n}$ by $T_t((p_1, p_2, ..., p_n, q_1, q_2, ..., q_n)) = (p_1(t), p_2(t), ..., p_n(t), q_1(t), q_2(t), ..., q_n(t))$ where the right side is obtained by solving the system of partial differential equations above with the initial conditions $(p_1(0), p_2(0), ..., p_n(0), q_1(0), q_2(0), ..., q_n(0)) = (p_1, p_2, ..., p_n, q_1, q_2, ..., q_n)$.

Theorem [Liouville]

Each T_t preserves Lebesgue measure on $\mathbb{R}^{2n}$.

We shall not prove this theorem.

In above example the maps $\{T_t\}_{\geq 0}$ satisfy the physically intuitive properties $T_0 = I$ and $T_{t+s} = T_t \circ T_s$ for all $t, s \geq 0$. Such a collection of measure preserving transformation is called a *flow* if $(x, t) \to T_t(x)$ is measurable.

Of course the example from Hamiltonian dynamics does not satisfy our assumption that the basic measure space is a probability space.

Ex. 7

This and the next example are from Stochastic Processes.

Let $\Omega = \mathbb{R}^\infty$, the space of all sequences of real numbers with the Frechet metric $d(\{a_n\}, \{b_n\}) = \sum_{n=1}^{\infty} \frac{|a_n - b_n|}{2^n[1+|a_n - b_n|]}$.

Exercise: the Borel sigma field $\mathcal{F}$ of Ω coincides with the sigma field generated by the projection maps $p_1, p_2, ...$ defined by $p_n(\{a_1 a_2, ...\}) = a_n$.

A set of the type $\{\omega \in \Omega : (\omega_{n_1}, w_{n_2}, ..., \omega_{n_k}) \in A\}$ where k is a positive integer, $n_1, n_2, ..., n_k$ are distinct positive integers and A is a Borel set in $\mathbb{R}^k$ is called a *cylinder set*. Noting that $\{\omega \in \Omega : (\omega_{n_1}, w_{n_2}, ..., \omega_{n_k}) \in A\} = (p_{n_1}, p_{n_2}, ..., p_{n_k})^{-1}(A)$ we see that the cylinder sets generate the Borel sigma field. Note that a given cylinder set has many representations; for example $\{\omega : \omega_1 \in A\} = \{\omega : (\omega_1, \omega_2) \in A \times \mathbb{R}\}$. We claim that cylinder sets form a field. This is easily seen from the fact that the set of coordinates in a cylinder set can always be enlarged (using the ' $\times\mathbb{R}$' technique) so that any two sets can be defined in terms of the same set of coordinates.

A probability measure P on the Borel sigma field of Ω is called a product measure if $P\{\omega \in \Omega : (\omega_{n_1}, w_{n_2}, ..., \omega_{n_k}) \in A_1 \times A_2 \times ... \times A_k\} = P_{n_1}(A_1) P_{n_2}(A_2) ... P_{n_k}(A_k)$ for some probability measures $P_1, P_2, ...$ on $\mathbb{R}$.

We now assume that P is a product measure in which $P_n = P_1$ for all n.

Claim: $p_1, p_2, ...$ is an i.i.d. sequence on $(\Omega, \mathcal{F}, P)$.

Remark: this example is a typical i.i.d sequence in some sense. This will elaborated upon later.

Proof: $P\{\omega : (p_1(\omega), p_2(\omega), ..., p_k(\omega) \in A_1 \times A_2 \times ... \times A_k)\} = P_1(A_1) P_2(A_2) ... P_k(A_k)$ and $p\{\omega : p_i(\omega) \in A_i\} = P_i(A_i)$ so $P\{\omega : (p_1(\omega), p_2(\omega), ..., p_k(\omega) \in A_1 \times A_2 \times ... \times A_k) = P\{\omega : p_1(\omega) \in A_1\} P\{\omega : p_2(\omega) \in A_2\} ... P\{\omega : p_k(\omega) \in A_k\}$

Now define $T : \Omega \to \Omega$ by $T(\omega_1, \omega_2, ...) = (\omega_2, \omega_3, ...)$. We claim that this map is m.p. The fact that $PT^{-1}(A) = P(A)$ for a cylinder set A follows immediately from $P\{\omega \in \Omega : (\omega_{n_1}, w_{n_2}, ..., \omega_{n_k}) \in A_1 \times A_2 \times ... \times A_k\} = P_1(A_1) P_1(A_2) ... P_1(A_k)$ and since cylinder sets generate the Borel sigma field it follows that T is m.p..

Ex 8

This is similar to Ex. 7 but we replace $\mathbb{R}$ by a finite set. Let $S = \{1, 2, ..., N\}$ and $p_i \geq 0 (1 \leq i \leq N)$ with $p_1 + p_2 + ... + p_N = 1$. Let $\Omega = S^\infty$, the space of all sequences from S. Let P be a probability measure on the sigma field generated by cylinder sets of Ω such that $P\{\omega_1 = i_1, \omega_2 = i_2, ..., \omega_k = i_k\} = p_{i_1}p_{i_2}...p_{i_k}$.

Thus P is the 'distribution' of an i.i.d sequence of random variables taking values in S. T defined as in above example is again m.p.. We call such a T a *Bernoulli shift*.

Ex. 9
Stationary shift

A sequence $\{X_n\}$ of random variables on $(\Omega, \mathcal{F}, P)$ is said to be *stationary* if the joint distribution of $(X_{n_1}, X_{n_2}, ..., X_{n_k})$ is same as that of $(X_{n_1+m}, X_{n_2+m}, ..., X_{n_k+m})$ for any positive integer m. Let us consider the 'canonical version' of this, i.e. assume that $\Omega = \mathbb{R}^\infty, \mathcal{F} =$ cylinder sigma field and P is a p.m. such that $P\{\omega : (\omega_{n_1+m}, \omega_{n_2+m}, ..., \omega_{n_k+m}) \in A\}$ is independent of $m \in \{0, 1, 2, ...\}$. Let T be the usual shift on $\Omega : T(\omega_1, \omega_2, ...) = (\omega_2, \omega_3, ...)$. Then T is m.p. Just take $m = 1$ in the definition to show that $P(T^{-1}(A)) = P(A)$ for any cylinder set A. We remark that the shift in the case if independent $p_n' s$ need not be m.p.; it is m.p. iff $p_n' s$ are i.i.d. Conversely, if the shift corresponding to a sequence $\{X_n\}$ is m.p. then the process $\{X_n\}$ is stationary.

Ex. 10
Markov shift
If the measure P on $\mathbb{R}^\infty$ makes $\{p_n\}$ a Markov chain the shift T need not be m.p.. It is m.p. if the Markov chain has a stationary distribution and the chain starts with this distribution. If this condition holds we call T a Markov shift.

Ex. 11
Affine maps
Composition of two m.p. transformations on the same probability space is m.p.. A map on a topological group G of the type $h \to gT(h)$ where g is a fixed element and T is a continuous automorphism is called an *affine* map. By Examples 3 and 4 above this map is m.p..

We now begin with a study of m.p. transformations. If $\omega \in \Omega$ the set $\{T^n(\omega)\}$ is called the orbit of ω under T. It is useful to think of $T^n\omega$ as the position of a point at time n.

We now show how to construct an i.m.p. transformation from a m.p. transformation.

Let $(\Omega, \mathcal{F}, P, T)$ be a DS. Let $\Omega' = \prod_{i=0}^{\infty} \Omega_i$ where $\Omega_i = \Omega$ for each i. Let $\Omega_0 = \{(\omega_i) \in \Omega' : T\omega_i = \omega_{i-1} \ \forall i \geq 1\}$. Let $\mathcal{F}_0$ be the trace of the cylinder sigma field of Ω' on Ω_0. Let Q be the probability measure on Ω' which makes the projection maps i.i.d. with common distribution P. Let $S(\omega_0, \omega_1, ...) = (T\omega_0, \omega_0, \omega_1, ...)$. Then S is i.m.p and it is ergodic iff T is.

Theorem [Poincare's Recurrence Theorem]

Let $(\Omega, \mathcal{F}, P, T)$ be a DS and $P(A) > 0$. Then $P\{\omega \in A : T^n(\omega) \in A$ for infinitely many $n\} = P(A)$.

Proof: let $A_n = \bigcup_{k=n}^{\infty} T^{-k}A, n = 0, 1, 2, ...$ $(T^0 = I)$. Then A_n is decreasing and $T^{-1}(A_n) = A_{n+1}$. Since T is m.p. we get $P(A_n) = P(A_0)$ for all n. This and the monotonicity of $A_n's$ shows that $P(A_n \Delta A_0) = 0$ for all n. Let $B = \bigcap_{n=0}^{\infty} A_n$. Then Then $B \subset A_0$ and $P(B \Delta A_0) = 0$. Therefore $P((A \cap B)\Delta(A \cap A_0)) = 0$. Hence $P(A \cap B) = P(A \cap A_0) = P(A)$ (because $A \subset A_0$).

The result follows if we show that $A \cap B \subset \{\omega \in A : T^n(\omega) \in A$ for infinitely many $n\}$. If $\omega \in A \cap B$ then for each n $\omega \in A_n$ and hence $\omega \in T^{-k}A$ (or $T^k\omega \in A$) for at least one $k \geq n$.

Poincare also proved a topological property of an open set of points that return to it infinitely often.

Theorem

Suppose Ω is a metric space and $\mathcal{F}$ its Borel sigma field. Let P be a Borel probability measure with full support (i.e. $P(U) > 0$ for every non-empty open set U). Let T be a continuous m.p. transformation on Ω. Then for any open set U the set of points of U which return to it infinitely many times is the complement of a set of first category in U.

[Hence the points of U which return only finitely many times if both measure theoretically and topologically small].

Proof: for any fixed open set U consider $A_k = \{x \in U : T^i x \notin U \ \forall i > k\}$. This set is clearly closed and so $A = \bigcup_{k=1}^{\infty} A_k$ is an F_σ. Note that $E_n = U \backslash A$ is precisely the set of points of U which return to U infinitely many times. E_n is a G_δ because A is an F_σ. Also, since every non-empty open set has positive measure previous theorem implies $P(E_n) = P(U)$. This implies that E_n is dense in U. Hence E_n is a dense G_δ in U and this implies that its complement in U is of first category.

Remark: more topological results will be proved later in the section on 'Topological Dynamics'. In particular we prove another recurrence theorem there called The Birkhoff's Recurrence Theorem.

Definition: let $(\Omega, \mathcal{F}, P, T)$ be a DS. We say T is ergodic if $T^{-1}(A) = A, (A \in \mathcal{F})$ implies $P(A) = 0$ or 1.

A set $A \in \mathcal{F}$ with $T^{-1}(A) = A$ is called an *invariant set*. The collection of all invariant sets is a sigma field called the *invariant sigma field*.

Examples:

Examples 1 and 2: we claim that the transformation in the first example is ergodic iff α is irrational and the one in Example 2 is ergodic iff a is not a root of unity.

Suppose $\alpha \in (0,1)$ be irrational. Let A be an invariant set for the transformation in Ex. 1 and define $f \in L^2([0,2\pi])$ by $f = I_{2\pi A}$. Consider the Fourier coefficients $\hat{f}(n)$ of f. We have $\hat{f}(n) = \frac{1}{2\pi}\int\limits_0^{2\pi} e^{-inx} I_{2\pi A}(x)dx = \int\limits_0^1 e^{-2\pi iny} I_A(y)dy$. Hence $\hat{f}(n) = \int\limits_0^1 e^{-2\pi iny} I_{T^{-1}(A)}(y)dy = \int\limits_0^1 e^{-2\pi in(z-\alpha)} I_A(z)dz = e^{2\pi in\alpha}\hat{f}(n)$. Since $e^{2\pi in\alpha} = 1$ iff $n = 0$ we see that $\hat{f}(n) = 0$ for $n \neq 0$ which means f is a.e. constant. Thus $I_{2\pi A} = 1$ a.e. or $I_{2\pi A} = 0$ a.e. which implies that $P(A) = 0$ or 1.

Now suppose α is a rational number $\frac{p}{q}$. Let $A = \{\omega : e^{2\pi iq\omega} \in C\}$ where C is a Borel set in S^1. This set is invariant. If T is ergodic then $P(A) = 0$ or 1 for any C. The measure Q induced by the function $\omega \to e^{2\pi iq\omega}$ takes only the values 0 and 1.

Exercise

Show that such a measure is degenerate, i.e. $Q = \delta_a$ for some $a \in S^1$.

Hint: use a compactness argument.

We now conclude that $e^{2\pi iq\omega}$ is a.e. constant. Being continuous, it must be a constant everywhere and so $e^{2\pi iq\omega} = 1$ for all $\omega \in (0,1)$. This is a contradiction.

The proof of the corresponding result for a rotation on S^1 is very similar and we omit the details.

Ergodicity of the transformation in Ex. 7

This is an easy consequence of Kolmogorov $0-1$ Law: if $T^{-1}(A) = A$ then $A \in \sigma\{p_n, p_{n+1}, ...\}$ for each n and hence $P(A) = 0$ or 1. [Indeed $A \in \sigma\{p_1, p_2, ...\}$ and so $T^{-1}(A) \in \sigma\{p_2, p_3, ...\}$: this is easy to verify for a cylinder set A and the collections of all Borel sets A which satisfies this property is a $\sigma-$field.

We now prove a basic theorem on ergodicity:

Theorem

With above notations, FAE:

1. T is ergodic
2. $P(A\Delta T^{-1}(A)) = 0$ implies $P(A) = 0$ or 1.
3. $P(T^{-n}(A) \cap B) = 0$ for all $n \in \mathbb{N}$ implies $P(A) = 0$ or $P(B) = 0$.
4. If $f : \Omega \to \mathbb{R}$ is measurable and $f(T(\omega)) = f(\omega)$ for all ω then there is a constant c such that $f = c$ a.e.

[Def: An *invariant function* for T is a measurable function f such that $f(T(\omega) = f(\omega)$ for all ω].

5. If $f : \Omega \to \mathbb{R}$ is measurable and $f(T(\omega)) = f(\omega)$ for almost all ω then there is a constant c such that $f = c$ a.e.

6. If $f \in L^2$ and $f(T(\omega)) = f(\omega)$ for all ω then there is a constant c such that $f = c$ a.e.

7. If $f \in L^2$ and $f(T(\omega)) = f(\omega)$ for almost all ω then there is a constant c such that $f = c$ a.e.

8. If $f \in L^1$ and $f(T(\omega)) = f(\omega)$ for all ω then there is a constant c such that $f = c$ a.e.

9. If $f \in L^1$ and $f(T(\omega)) = f(\omega)$ for almost all ω then there is a constant c such that $f = c$ a.e.

Proof: Suppose $P(A\Delta T^{-1}(A)) = 0$. Let B be the set of points ω such that $T^n(\omega) \in A$ for infinitely many positive integers n. It is clear that B is an invariant set. Hence $P(B) = 0$ or 1. Note that $I_A = I_{T^{-1}(A)}$ a.e.. By iteration this gives $I_A = I_{T^{-n}(A)}$ a.e. for each n. Hence $I_A = \limsup_n I_{T^{-n}(A)}$ a.e. which means $I_A = I_B$ a.e. Hence $P(A) = P(B) = 0$ or 1.

Since 2) obviously implies 1) we conclude that 1) and 2) are equivalent. 3) implies 1) follows by taking $B = A^c$. We now prove 2) implies 3). We have $P((\bigcup_{n=1}^{\infty} T^{-n}A) \cap B) = 0$. Let $C = \bigcup_{n=1}^{\infty} T^{-n}A$. Then $T^{-1}(C) \subset C$. Also $P(T^{-1}(C)) = P(C)$ (because T is m.p.) and hence $P((T^{-1}C)\Delta C) = 0$. By 2) $P(C) = 0$ or 1. If $P(A) > 0$ then $P(C) \geq P(T^{-1}(A)) = P(A) > 0$ and $P(C) = 1$. But $P(C \cap B) = 0$ and hence $P(B) = 0$. We have proved the equivalence of 1), 2) and 3).

1) implies 4): let f be as in 4). For every $a \in \mathbb{R}$ the set $A = \{\omega : f(\omega) < a\}$ is an invariant set. Thus $P\{\omega : f(\omega) < a\} = 0$ or 1 for each a. Since this probability is a monotonic function of a is easy to see that there is a_0 such that $P\{\omega : f(\omega) < a\} = 0$ if $a < a_0$ and $P\{\omega : f(\omega) < a\} = 1$ if $a > a_0$. Thus $P\{\omega : f(\omega) \notin [a_0 - \frac{1}{n}, a_0 + \frac{1}{n})\} \leq P\{\omega : f(\omega) < a_0 - \frac{1}{n}\} + P\{\omega : f(\omega) \geq a_0 + \frac{1}{n}\} = 0 + 0$ for each n. Letting $n \to \infty$ we get $P\{\omega : f(\omega) = a_0\} = 1$

4) implies 1): just take $f = I_A$.

The equivalence of 2) and 5) is similar to that of 1) and 4): if $f(T(\omega) = f(\omega)$ for almost all x and $A = \{\omega : f(\omega) < a\}$ then $P(A\Delta T^{-1}(A)) = 0$ since $I_{T^{-1}(A)}(\omega) = I_A(T\omega) = \{\omega : f(T(\omega)) < a\}$ which differs from $\{\omega : f(\omega) < a\} = A$ only by a null set. Conversely $P(A\Delta T^{-1}(A)) = 0$ and $f = I_A$ imply $f(T(\omega) = f(\omega)$ for almost all ω.

Since indicator functions belong to both L^1 and L^2 it is clear that we can restrict f to functions in either of these spaces without any change in proof. This completes the proof.

In the course of the above proof we have made an elementary but very useful observation:

if T is ergodic and C is a measurable set with $T^{-1}(C) \subset C$ then $P(C) = 0$ or 1. Indeed $P(T^{-1}(C)) = P(C)$ (because T is m.p.) and hence $P((T^{-1}C)\Delta C) = 0$ and we can apply equivalence of 1) and 2).

The same conclusion holds if $C \subset T^{-1}(C)$; if T is i.m.p and ergodic we can replace the hypothesis by either of the two conditions $T(C) \subset C, C \subset T(C)$.

Theorem

Let X be a compact metric space and P a Borel probability measure on X with full support (i.e. $P(U) > 0$ for every non-empty open set U). If $T : X \to X$ is ergodic then almost all orbits are dense (i.e. $P\{x : \{T^n(x) : n = 0, 1, 2, ...\}$ is dense in $X\} = 1$.

Proof: let $\{U_n\}$ be a countable base for the topology. Then $x \in \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{\infty} T^{-k}U_n$ iff the orbit of x intersects each U_n iff the orbit of x intersects each open set iff the orbit of x is dense. It remains to show $P\{\bigcap_{n=1}^{\infty} \bigcup_{k=0}^{\infty} T^{-k}U_n\} = 1$. Let $C_n = \bigcup_{k=0}^{\infty} T^{-k}U_n$. Then $T^{-1}(C_n) \subset C_n$ and hence $P(C_n) = 0$ or 1. However, $C_n \supset U_n$ and $P(U_n) > 0$ so $P(C_n) = 1$. This is true for each n so $P(\bigcap_{n=1}^{\infty} C_n) = 1$ as required.

Remark: this theorem proves that if a rotation $z \to az$ on S^1 is ergodic then the orbit of some point is dense which means $\{a^n\}_{n \geq 0}$ is dense. [The orbit of some point is dense implies orbit of every point is dense!].

For the next theorem we need the following facts from the theory of topological groups:

Let G be a compact metric topological group and $\hat{G}$ its dual group. $\hat{G}$ is the collection of all *characters* on γ, i.e. continuous homomorphisms $\gamma : G \to S^1$ which is a group under pointwise multiplication. Let P be the Haar measure on G. [Recall that on any compact group there is a unique probability measure P on the Borel sigma field such that $P(gA) = P(Ag) = P(A)$ for all Borel sets A and all g. This measure is the Haar measure of the group]. Each character γ belongs to $L^2(P)$ and any two distinct characters are orthogonal: if $\gamma_1 \neq \gamma_2$ then $\gamma = \frac{\gamma_1}{\gamma_2}$ is also a character and $\int \gamma dP = \int \gamma(ag)dP(g)$ (by definition of Haar measure)

$= \gamma(a)\int \gamma dP$ for each a. Since γ is not the constant character it follows that $\int \gamma dP = 0$. This means $\int \gamma_1 \bar{\gamma_2} dP = 0$ as stated. Of course each character has norm 1 in L^2. Since G is a compact metric space the space $C(G)$ is separable and $C(G)$ is dense in L^2. Thus L^2 is separable too (since uniform approximation implies approximation in L^2) and hence $\hat{G}$ is a countable set $\{\gamma_i\}$. It can be shown that the orthonormal set $\{\gamma_i\}$ is complete. [If we assume that for each

$g \neq 1$ there is a character γ such that $\gamma(g) \neq 1$ the we can use Stone-Weierstrass Theorem to conclude that the vector space spanned by the characters, which is clearly an algebra, is dense in $C(G)$ and this implies that $\{\gamma_i\}$ is complete. We

refer the reader to Theorem (22.17), p. 345 of Abstract Harmonic Analysis Vol I by E. Hewitt And K. A. Ross]. The expansion of an L^2 function w.r.t. the orthonormal basis $\{\gamma_i\}$ is called the *Fourier series* expansion of f.

Theorem

Let G be a compact metric topological group and $Tg = ag$. Then T is ergodic iff $\{a^n\}_{n\geq 0}$ is dense. In this case G is necessarily abelian.

Proof: T is ergodic implies $\{a^n\}_{n\geq 0}$ is dense. This follows from above corollary and the fact that Haar measure has full support. Now suppose $\{a^n\}_{n\geq 0}$ is dense. It is clear that G is abelian. Let $f \in L^2$ and $f \circ T = f$. Recalling that the dual group $\hat{G}$ is countable, say, $\{\gamma_1, \gamma_2, ...\}$, and that the characters $\{\gamma_i\}$ form an orthonormal basis for L^2 we can write $f = \sum\limits_i < f, \gamma_i > \gamma_i$. Hence $f(g) = f(Tg) = f(ag) = \sum\limits_i < f, \gamma_i > \gamma_i(ag) = \sum\limits_i < f, \gamma_i > \gamma_i(a)\gamma_i(g)$. Thus $\sum\limits_i < f, \gamma_i > \gamma_i = \sum\limits_i < f, \gamma_i > \gamma_i(a)\gamma_i$. Orthonormality of $\gamma_i's$ implies that $< f, \gamma_i >=< f, \gamma_i > \gamma_i(a)$ for each i. If $\gamma_i(a) = 1$ then $\gamma_i(a^n) = 1$ for each n and the hypothesis now implies that $\gamma_i = 1$. Thus $< f, \gamma_i >= 0$ except when $\gamma_i = 1$ which implies $f = \sum\limits_i < f, \gamma_i > \gamma_i$ is a constant. This completes the proof.

Theorem

The map $T : S^1 \to S^1$ defined by $Tz = az$ where $a \in S^1$ is ergodic iff a is not a root of unity.

Proof: we have to show that $\{a^n\}_{n\geq 0}$ is dense in S^1 iff a is not a root of unity. If a is a root of unity then $\{a^n\}_{n\geq 0}$ is a finite set and hence it is not dense. Suppose now that a is not a root of unity. Then $\{a, a^2, ...\}$ is an infinite set and hence it has a limit point. Hence, if $\epsilon > 0$ we can find positive integers n and k such that $\left|a^n - a^{n+k}\right| < \epsilon/2$. For some m the points $a^n, a^{n+k}, ..., a^{n+mk}$ form an ϵ net for S^1. This of course implies that $\{a^n\}_{n\geq 0}$ is dense in S^1.

APPENDIX
SEPARABILITY OF C(X)

Theorem

Let X be a compact Hausdorff space. Then X is metrizable iff $C(X)$ is separable.

One way is easy. if $\{f_n\}$ is dense in $C(X)$ then $d(x,y) = \sum_{n=1}^{\infty} \frac{|f_n(x)-f_n(y)|}{2^n[1+|f_n(x)-f_n(y)|]}$ defines a metric on X such that the identity map from X with the original topology to X with the metric d is continuous (in view of continuity of the functions f_n) and the inverse map is automatically continuous by compactness.

Now let (X,d) be a compact metric space. For each n we can find a finite set $\{x_{n,1}, x_{n,2}, ..., x_{n,k_n}\}$ such that $X = B(x_{n,1}\frac{1}{n}) \cup B(x_{n,2}\frac{1}{n}) \cup ... \cup B(x_{n,k_n}\frac{1}{n})$. Let $f_{n,i}$ be a continuous function : $X \to [0,1]$ such that $f_{n,i}(x) = 1$ if $d(x, x_{n,i}) < \frac{1}{n}$ and $f_{n,i}(x) = 0$ if $d(x, x_{n,i}) \geq \frac{2}{n}$. If $x \neq y$ choose n such that $\frac{3}{n} < d(x,y)$. Then $x \in B(x_{n,i}\frac{1}{n})$ for some i and $f_{n,i}(x) = 1$. Also $d(y, x_{n,i}) \geq \frac{2}{n}$ because, otherwise, $d(x,y) \leq d(x, x_{n,i}) + d(x_{n,i}, y) < \frac{1}{n} + \frac{2}{n} = \frac{3}{n}$ a contradiction. Thus $f_{n,i}(y) = 0 \neq 1 = f_{n,i}(x)$. We have proved that the set $\{f_{n,i} : 1 \leq i \leq k_n, n = 1, 2, ...\}$ separates points. By Stone-Wierestrass Theorem finite linear combinations of finite product of these functions form a dense subalgebra of $C(X)$. Hence finite rational linear combinations of finite product of these functions form a countable dense subset of $C(X)$.

End of Appendix

Ergodicity of a continuous homomorphism on the torus $S^1 \times S^1$ (coordinatewise multiplication).

Let $T : S^1 \times S^1 \to S^1 \times S^1$ be a continuous homomorphism. Then there exist integers a, b, c, d such that $T(u,v) = (u^a v^b, u^c v^d)$ $\forall (u,v) \in S^1 \times S^1$. To see this note that the two components of $T(u,1)$ and $T(1,v)$ are homomorphisms of S^1. Since any homomorphism of S^1 is of the type $u \to u^a$ for some integer a and since $T(u,v) = T(u,1)T(1,v)$ we are done. Let us prove that T is surjective if and only if $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$. Indeed, if T is not surjective then its range is a proper (compact) subgroup and hence there is a non-trivial character which has the value 1 at every point of the range. Hence there exists characters γ_1, γ_2 of S^1, no both 1, such that $\gamma_1(u^a v^b)\gamma_2(u^c v^d) = 1$ $\forall u, v \in S^1$. There exists integers λ_1, λ_2 such that $\gamma_1(u) = u^{\lambda_1}$ and $\gamma_2(v) = v^{\lambda_2}$; thus . $u^{a\lambda_1 + c\lambda_2} v^{b\lambda_1 + d\lambda_2} \equiv 1$. which implies $a\lambda_1 + c\lambda_2$ and $b\lambda_1 + d\lambda_2 = 0$. Note that λ_1 and λ_2 cannot both be 0. It follows . that the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is singular and its determinant is 0. Conversely, if the determinant is 0 then there exists integers λ_1, λ_2 not both 0 such that $a\lambda_1 + c\lambda_2$ and $b\lambda_1 + d\lambda_2 = 0$. [There exists a real number λ such that $\lambda(a,b) = (c,d)$ and λ is obviously rational. If $\lambda = \frac{p}{q}$ we can take $\lambda_1 = p, \lambda_2 = -q$]. For any point $(\xi, \zeta) = T(u,v)$ in the range of T we have $\xi^{\lambda_1}\zeta^{\lambda_2} = u^{a\lambda_1} v^{b\lambda_1} u^{c\lambda_2} v^{d\lambda_2} = u^0 v^0 = 1$. This implies that T is not surjective. [

For example $(\xi, 1)$ is not in the range of T if ξ is not a root of unity if $\lambda_1 \neq 0$ and $(1, \zeta)$ is not in the range of T if ζ is not a root of unity if $\lambda_2 \neq 0$].

We now use the following characterization of ergodicity of onto homomorphisms of compact metric groups:

if G is a compact metric group and $T : G \to G$ is a surjective homomorphism then T is ergodic if and only if the condition $\gamma \circ T^n = \gamma$ for some character γ of G and some non-zero integer n implies $\gamma = 1$.

[Proof: suppose $f \in L^1$ and $f \circ T = f$. Let $f = \sum a_n \gamma_n$ be the Fourier series of f. Then $\sum a_n \gamma_n \circ T^j = \sum a_n \gamma_n \ \forall j$. Fix n and suppose $\gamma_n \circ T, \gamma_n \circ T^2, \gamma_n \circ T^3, ...$ are all distinct. Then $a_n =$ coefficient of $\gamma_n \circ T^j$ on the right side and these coefficients are $a_{n_1}, a_{n_2}, ...$ for distinct $n_i's$. Hence $\sum_i |a_{ni}|^2 = \sum_i |a_n|^2 = \infty$, a contradiction unless $a_n = 0$. Hence $\gamma_n \circ T, \gamma_n \circ T^2, \gamma_n \circ T^3, ...$ cannot all be distinct except when $a_n = 0$. However, $\gamma_n \circ T^k = \gamma_n \circ T^l, k < l$ implies $\gamma_n \circ T^{l-k} = \gamma_n$. If we assume that $\gamma \circ T^n = \gamma$ for some character γ of G and some non-zero integer n implies $\gamma = 1$ it follows that the only non-zero coefficient in the series $\sum a_n \gamma_n$ is the one corresponding to $\gamma = 1$ and so f is a constant. This proves the *if* part. Now suppose T is ergodic. Suppose $\gamma \circ T^n = \gamma$ for some character γ of G and some non-zero integer n. We have to show $\gamma = 1$. Let m be the least positive integer such that $\gamma \circ T^m = \gamma$. Let $f = \gamma + \gamma \circ T + \gamma \circ T^2 + ... + \gamma \circ T^{m-1}$. Then f is an invarant L^1 function, hence constant. Since $\gamma, \gamma \circ T, \gamma \circ T^2, ..., \gamma \circ T^{m-1}$ are distince characters they are orthogonal and this is possible only when $m = 1$ and $f = \gamma$ is a constant].

Theorem

With above notations T is ergodic if and only if no root of unity is an eigen value of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Proof: if T is not ergodic then there is a non-trivial character γ and a non-zero integer n such that $\gamma \circ T^n = \gamma$. Since $\gamma(z, \zeta) = (z^k, \zeta^l)$ for some integers k, l not both 0 we have $(u^A v^B)^k (u^C v^D)^l = (u^k, v^l)$ for all $u, v \in S^1$ where $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^n$. This implies $Ak + Cl = k$ and $Bk + Dl = l$. This proves that transpose of $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, hence $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ itself has 1 as an eigen value. Since eigen values of $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ are the $n-th$ powers of those of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have proved that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an $n-th$ root of unity as an eigen value. Conversely suppose 1 is an eigen value of $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^n$,hence an eigen vaue of its transpose, for some positive integer n.

Then there exist integers λ_1, λ_2 not both zero such that $\begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$. Let $\gamma_1(z) = z^{\lambda_1}$ and $\gamma_2(z) = z^{\lambda_2}$. Then $\gamma = (\gamma_1, \gamma_2)$ is a non-trivial character of $S^1 \times S^1$. We claim that $\gamma \circ T^n = \gamma$. For this we have to show that $(u^A v^B)^{\lambda_1} (u^C v^D)^{\lambda_2} = u^{\lambda_1} v^{\lambda_2}$. This is true because $A\lambda_1 + C\lambda_{2=\lambda_1}$ and $B\lambda_1 + D\lambda_2 = \lambda_2$.

Remark: the result extends to the $k-$ torus $S^1 \times S^1 \times ... \times S^1$.and the proof is similar.

Necessary and sufficient conditions for ergodicity of affine maps will be given later.

The Guass transformation (Ex. 6) is ergodic. Proof will be given later.

THE ERGODIC THEOREM

The Ergodic Theorem is really a set of theorems about the convergence of certain 'time' averages. We prove a number of versions of this important theorem.

Theorem [von Neuamnn Ergodic Theorem (alias Mean Ergodic Theorem)]

Let H be a Hilbert space and $U : H \to H$ be an isometry. If $x \in H$ then $\{\frac{1}{n}(x + Ux + U^2x + ... + U^{n-1}x)\}$ converges to some point $z \in H$ with $Uz = z$.

Proof: let M be the closure of the range of $(I - U)$ and $N = \{x : Ux = x\}$. Clearly M and N are closed subspaces of L^2. We claim that $M^\perp = N$. Note that $x \in M^\perp \Leftrightarrow < x, y - Uy >= 0 \forall y \Leftrightarrow < x - U^*x, y >= 0 \forall y \Leftrightarrow U^*x = x$. We have to show that the conditions $Ux = x$ and $U^*x = x$ are equivalent for isometries. We have $\|U^*x - x\|^2 =< U^*x - x, U^*x - x >$
$= \|U^*x\|^2 + \|x\|^2 - 2\operatorname{Re} < Ux, x >$. Hence $Ux = x$ implies $\|U^*x - x\|^2 = \|U^*x\|^2 - \|x\|^2 \leq \|U^*\|^2 \|x\|^2 - \|x\|^2$
$= \|U\|^2 \|x\|^2 - \|x\|^2 = 0$ and so $U^*x = x$. The reverse implication follows by changing U to U^*. This proves the claim. Now let $x \in L^2(P)$. We can write $x = y + z$ where $y \in M$ and $z \in N$. Hence $\frac{1}{n}(x + Ux + U^2x + ... + U^{n-1}x) = \frac{1}{n}(y + Uy + U^2y + ... + U^{n-1}y) + \frac{1}{n}(z + Uz + U^2z + ... + U^{n-1}z)$. The second term is z (independent of n). If $\epsilon > 0$ then there exists u such that $\|y - (I - U)u\| < \epsilon$. Let $v = u - Uu$ so that $\|v - y\| < \epsilon$. Now $\left\| \frac{1}{n}(y + Uy + U^2y + ... + U^{n-1}y) - \frac{1}{n}(v + Uv + U^2v + ... + U^{n-1}v) \right\| < \epsilon$ and $\frac{1}{n}(v+Uv+U^2v+...+U^{n-1}v) = \frac{1}{n}((u-Uu)+(Uu-U^2u)+...+(U^{n-1}u-U^nu)) = \frac{1}{n}(u-U^nu) \to 0$ as $n \to \infty$. It is now clear that $\frac{1}{n}(y+Uy+U^2y+...+U^{n-1}y) \to 0$ as $n \to \infty$ and hence $\frac{1}{n}(x+Ux+U^2x+...+U^{n-1}x) \to z$. The proof is complete.

Remark: the limit here is obviously the projection of x on $N = \{z : Uz = z\}$.

What has this theorem to do with m.p. transformations? Well, if $(\Omega, \mathcal{F}, P, T)$ is a DS then $H = L^2(P)$ is a Hilbert space and $Uf = f \circ T$ defines an isometry on it. Thus we have:

Corollary

Let $(\Omega, \mathcal{F}, P, T)$ be a DS and $f \in L^2(P)$. Then $\{\frac{1}{n}(f + f \circ T + f \circ T^2 + ... + f \circ T^{n-1})\}$ converges in L^2 norm to an invariant function g with $\int f dP = \int g dP$. If T is ergodic then g is constant a.e..

Proof: Only the last part needs a proof. Since L^2 convergence implies L^1 convergence we get $\int g dP = \lim\{\frac{1}{n}(\int f dP + \int f \circ T dP + ... + \int f \circ T^{n-1} dP) = \int f dP$. The limit g satisfies $g \circ T = g$ a.e. and hence it is a constant if T is ergodic.

Corollary [The L^1 Ergodic Theorem]

Let $(\Omega, \mathcal{F}, P, T)$ be a DS and $f \in L^1(P)$. Then $\{\frac{1}{n}(f + f \circ T + f \circ T^2 + ... + f \circ T^{n-1})\}$ converges in L^1 norm to an invariant function g with $\int f dP = \int g dP$. If T is ergodic then g is constant.

Proof: if $f \in L^1(P)$ and $\epsilon > 0$ then there exists $g \in L^2(P)$ such that $\|f - g\|_2 < \epsilon$. Let $\phi_n = \frac{1}{n}(f + f \circ T + f \circ T^2 + ... + f \circ T^{n-1})$ and $\xi_n = \frac{1}{n}(g + g \circ T + g \circ T^2 + ... + g \circ T^{n-1})$. Then $\|\phi_n - \xi_n\|_1 < \epsilon$ because $\|f - g\|_1 < \epsilon$ and T is m.p.. It follows that $\|\phi_n - \phi_m\|_1 < 2\epsilon + \|\xi_n - \xi_m\|_1 \leq 2\epsilon + \|\xi_n - \xi_m\|_2 < 3\epsilon$ if n and m are sufficiently large. This proves convergence of $\{\frac{1}{n}(y + f \circ T + f \circ T^2 + ... + f \circ T^{n-1})\}$ in L^1 norm. The equation $\int f dP = \int g dP$ and the fact that g is a constant when T is ergodic are proved exactly as in previous theorem.

We now prove a more powerful version of the theorem.

Theorem [Birkhoff Ergodic Theorem alias Pointwise Ergodic Theorem]

Let $(\Omega, \mathcal{F}, P, T)$ be a DS and $f \in L^1(P)$. Then $\{\frac{1}{n}(f + f \circ T + f \circ T^2 + ... + f \circ T^{n-1})\}$ converges almost everywhere and in L^1 to an almost invariant function g with $\int f dP = \int g dP$. If T is ergodic then g is constant.

We first prove the following:

Theorem [Maximal Ergodic Theorem]

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $U : L^1(P) \to L^1(P)$ be a positive contraction (i.e. a linear map such that $Uf \geq 0$ whenever $f \geq 0$ and $\|U\| \leq 1$). Let $N \in \mathbb{N}, f \in L^1(P), s_0(f) = 0, s_k(f) = f + Uf + U^2 f + ... U^{k-1} f$ and $F_N = \max\{s_k(f) : 0 \leq k \leq N\}$. Then $\int_{\{x: F_N(x) > 0\}} f dP \geq 0$.

Proof: (due to Garcia)

Note that if $0 \leq n < N$ then $F_N \geq s_n$ and the hypothesis implies $UF_N \geq Us_n = s_{n+1} - f$. Hence $UF_N + f \geq \max\{s_{n+1} : 0 \leq n < N\} = \max\{s_n : 1 \leq n \leq N\}$. On the set $\{x : F_N(x) > 0\}$ we have $\max\{s_n : 1 \leq n \leq N\} = \max\{s_n : 0 \leq n \leq N\} = F_N$. It follows that $UF_N + f \geq F_N$ on $\{x : F_N(x) > 0\}$. Hence $\int\limits_{\{x:F_N(x)>0\}} f dP \geq \int\limits_{\{x:F_N(x)>0\}} F_N dP - \int\limits_{\{x:F_N(x)>0\}} UF_N dP = \int F_N dP - \int\limits_{\{x:F_N(x)>0\}} UF_N dP$

$\geq \int F_N dP - \int UF_N dP$ (because $UF_N \geq 0$). But $\int UF_N dP = \|UF_N\|_1 \leq \|F_N\|_1 = \int F_N dP$ and hence $\int F_N dP - \int UF_N dP \geq 0$.

In particular if $(\Omega, \mathcal{F}, P, T)$ a dynamical system then $Uf = f \circ T$ defines an operator on $L^1(P)$ satisfying the conditions of the theorem.

Proof of Birkhoff's theorem: let $g(x) = \limsup\limits_{n\to\infty} \frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})(x)$ and $h(x) = \liminf\limits_{n\to\infty} \frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})(x)$. We claim that g and h are invariant. This follows easily from the identity $\frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})(Tx) = \frac{1}{n}(f \circ T + f \circ T^2 + \ldots + f \circ T^n)(x) = \frac{n+1}{n}\frac{1}{n+1}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^n)(x) - \frac{1}{n}f(x)$. Let $a < b$ with $a, b \in \mathbb{Q}$. Consider $E_{a,b} = \{x : h(x) < a < b < g(x)\}$. We have $T^{-1}(E_{a,b}) = E_{a,b}$ and $E_{a,b} \subset \{x : \sup\limits_n \frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})(x) > b\} = E_b$ (say). We apply the Maximal Ergodic Theorem with $(\Omega, \mathcal{F}, P)$ changed to $(E_{a,b}, \mathcal{F} \cap E_{a,b}, \frac{P(.\cap E_{a,b})}{P(E_{a,b})}, T|E_{a,b})$ and f replaced by $f - b$ and . We get $\int\limits_{\{x:F_N(x)>b\}\cap E_{a,b}} (f-b)dP \geq 0$. Clearly, the sets $\{x : F_N(x) > b\}$ increase to E_b(and $E_{a,b} \subset E_b$). Thus, letting $N \to \infty$ we get $\int\limits_{E_{a,b}} f dP \geq bP(E_{a,b})$. Replace f by $-f$ and (a, b) by $(-b, -a)$ to get $\int\limits_{E_{a,b}} f dP \leq aP(E_{a,b})$. These two inequalities give $bP(E_{a,b}) \leq aP(E_{a,b})$ which implies that $P(E_{a,b}) = 0$. Varying a and b we conclude that $g = h$ a.e. which means $\{\frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})\}$ converges almost everywhere.

To prove L^1 converges we observe that $M = \{f \in L^1 : \{\frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1})\}$ converges in $L^1\}$ is a closed subset of L^1. This follows easily from the fact $\left\|\frac{1}{n}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{n-1}) - \frac{1}{m}(f + f \circ T + f \circ T^2 + \ldots + f \circ T^{m-1})\right\|_1 \leq$

$\left\|\frac{1}{n}(g + g \circ T + g \circ T^2 + \ldots + g \circ T^{n-1}) - \frac{1}{m}(g + g \circ T + g \circ T^2 + \ldots + g \circ T^{m-1})\right\|_1 +$

$2\|f - g\|_1$ in view of the fact that T is m.p.. Thus, if $f \in \bar{M}$ and $\epsilon > 0$ we can choose $g \in M$ with $\|f - g\|_1 < \epsilon$ and the inequality just derived shows that

$\{\frac{1}{n}(f+f\circ T+f\circ T^2+...+f\circ T^{n-1})\}$ is Cauchy, hence convergent in L^1. Since M contains L^∞ by Dominated Convergence Theorem we conclude that M is dense and closed, hence equals L^1.

Let g be the pointwise and L^1 limit of $\{\frac{1}{n}(f+f\circ T+f\circ T^2+...+f\circ T^{n-1})\}$. We have $g\circ T=\lim\frac{1}{n}(f\circ T+f\circ T^2+...+f\circ T^n)=\lim\frac{1}{n}(f\circ T+f\circ T^2+...+f\circ T^n)=\lim\frac{1}{n+1}(f\circ T+f\circ T^2+...+f\circ T^n)(x)=g$, so $g\circ T=g$ a.e.. By L^1 convergence and the fact that T is m.p. it follows that $\int gdP=\lim\int\frac{1}{n}(f\circ T+f\circ T^2+...+f\circ T^n)dP=\int fdP$ and the proof is complete.

Remark: with a little extra effort we can show that the theorem is true if we replace the probability measure P by an arbitrary positive measure.

Identifying the limiting function:

let $\mathcal{G}$ consist of those sets $A\in\mathcal{F}$ for which $I_{T^{-1}(A)}=I_A$ a.e. (i.e. $P(A\Delta T^{-1}(A))=0$). $\mathcal{G}$ is a sigma field. We have:

Theorem

Let $(\Omega,\mathcal{F},P,T)$ be a DS and $f\in L^1(P)$. Then

$\{\frac{1}{n}(f+f\circ T+f\circ T^2+...+f\circ T^{n-1})\}$ converges almost everywhere and in L^1 to $E(f|\mathcal{G})$.

Proof: let g be the limit of $\{\frac{1}{n}(f+f\circ T+f\circ T^2+...+f\circ T^{n-1})\}$ in Birkhoff's Theorem. We claim that g is measurable w.r.t $\mathcal{G}$. Indeed $g=g\circ T$ a.e.. Hence, for any Borel set C in $\mathbb{R}$ $I_{T^{-1}g^{-1}(C)}=I_{(g\circ T)^{-1}(C)}=I_{g^{-1}(C)}$ a.e. so $g^{-1}(C)\in\mathcal{G}$. It remains only to show that $\int\limits_A fdP=\int\limits_A gdP$ for all $g\in\mathcal{G}$. But this follows from L^1 convergence in Birkhoff's theorem and the fact that $\int\limits_A\frac{1}{n}(f+f\circ T+f\circ T^2+...+f\circ T^{n-1})dP=\int fdP$.

Theorem

FAE

a) The limiting function in Birkhoff's theorem is a.s. constant for every $f\in L^1$

b) the sigma field $\mathcal{G}$ is trivial in the sense every set in this sigma field has probability 0 or 1

c) $E(f|\mathcal{G})=\int fdP$ a.s. for every $f\in L^1$.

d) T is ergodic.

Proof: for a) implies d) take $f=I_A$ where A is invariant; for c) implies b) take $f=I_A$ where $A\in\mathcal{G}$. Rest of the proof is straightforward.

Theorem [Strong Law Of Large Numbers]

Let $\{X_n\}$ be an i.i.d. sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. If $E|X_1| < \infty$ then $\frac{X_1+X_2+...+X_n}{n} \to EX_1$ a.s. and in L^1.

Proof: we use the ideas from Ex. 7 above. Let $\Omega_1 = \mathbb{R}^\infty, \mathcal{F}_1 =$ Borel sigma field of Ω_1 and $Q = P \circ (X_1, X_2, ...)^{-1}$. Let Y_n be the projection to the $n-th$ coordinate. Then $\{Y_n\}$ is i.i.d on $(\Omega_1, \mathcal{F}_1, Q)$. Also, $Y_n = Y_1 \circ T^n$ where $T(a_1, a_2, ...) = (a_2, a_3, ...)$. Let $E = \{\{a_n\} \in \mathbb{R}^\infty : \lim \frac{a_1+a_2+...+a_n}{n}$ exists$\}$. We leave it as an exercise to show that this set is indeed a Borel set in $\mathbb{R}^\infty$. It follows that $Q(E) = P(X_1, X_2, ...)^{-1}(E)$. Birkhoff's theorem implies that $Q(E) = 1$ and we conclude that $P\{\omega : \{\frac{X_1(\omega)+X_2(\omega)+...+X_n(\omega)}{n}\}$ converges$\} = 1$. Also $\left\|\frac{X_1+X_2+...+X_n}{n} - \frac{X_1+X_2+...+X_m}{m}\right\|_1 = \left\|\frac{X_1+X_2+...+X_n}{n} - \frac{Y_1+Y_2+...+Y_m}{m}\right\|_1 \to 0$ so $\frac{X_1+X_2+...+X_n}{n}$ converges in L^1 too. The limiting random variable Z is invariant and T is ergodic and hence Z is a.s. constant. By L^1 convergence the constant is $\displaystyle\int X_1 dP$. This completes the proof.

Theorem [L^p Ergodic Theorem]

Let $(\Omega, \mathcal{F}, P, T)$ be a DS. and $f \in L^p$ where $1 < p < \infty$. Then $\{\frac{1}{n}(f + f\circ T + f\circ T^2 + ... + f\circ T^{n-1})\}$ converges in L^p. More generally let $U : L^p \to L^p$ be a linear map such that $\sup\{\|U^n\| : n \geq 0\} < \infty$. Then $\{\frac{1}{n}(f + Uf + U^2 f + ... + U^{n-1} f)\}$ converges in L^p. The limit function g satisfies $Ug = g$.

Proof: $\{\frac{1}{n}(f + Uf + U^2 f + ... + U^{n-1} f)\}$ is bounded sequence in L^p. Hence there is a weakly convegent subsequence. Let $\frac{1}{n_k}(f + Uf + U^2 f + ... + U^{n_k-1} f) \to g$ weakly in L^p. Since U is (weak-weak) continuous we have $\int hU\frac{1}{n_k}(f + Uf + U^2 f + ... + U^{n_k-1} f)dP \to \int hg$ $\forall h \in L^q$ where $q = \frac{p}{p-1}$. Now $f - U^n f = (I - U)\sum_{j=1}^{n-1} U^j f$. Averaging over $0 \leq n < n_k$ we see that $f - \frac{1}{n_k}(f + Uf + U^2 f + ... + U^{n_k-1} f) = (I-U)h$ for some $h \in L^p$. Let $f_k = \frac{1}{n_k}(f + Uf + U^2 f + ... + U^{n_k-1} f)$. We claim the following:

i): $\{\frac{1}{n}(f + Uf + U^2 f + ... + U^{n-1} f)\}$ converges in L^p if $f \in (I - U)(L^p)$

ii) $\{f \in L^p : \{\frac{1}{n}(f + Uf + U^2 f + ... + U^{n-1} f)\}$ converges in $L^p\}$ is a closed subspace of L^p

iii) $f - g$ belongs to the closed subspace in ii)

iv) g belongs to the closed subspace in ii)

These facts imply that f belongs to the closed subspace in ii) which finishes the proof.

Proof of i): if $f = h - Uh$ then $\frac{1}{n}(f + Uf + U^2 f + ... + U^{n-1} f) = \frac{1}{n}(h - U^n h) \to 0$ by hypothesis.

Proof of ii): this is proved along the same lines as the case $p = 2$. The hypothesis: $\sup\{\|U^n\| : n \geq 0\} < \infty$ is needed.

Proof of iii): $f - g$ is the weak limit of $f - \frac{1}{n_k}(f + Uf + U^2 f + ... + U^{n_k-1} f) = f - f_k$ and we have seen that the function belongs to the range of $I - U$. By i)

$f-f_k$ is in the subspace in ii). Since the weak closure of a convex set (hence that of a linear subspace) coincides with the norm closure we see that $f-g$ belongs to the closed subspace in ii). Finally we prove iv) by showing that $Ug = g$; this would imply that $\frac{1}{n}(g+Ug+U^2g+...+U^{n-1}g) = g\ \forall n$ thus completing the proof. We have $U(\frac{1}{n_k}(f+Uf+U^2f+...+U^{n_k-1}f)) - \frac{1}{n_k}(f+Uf+U^2f+...+U^{n_k-1}f) = \frac{1}{n_k}(f-U^{n_k}f) \to 0$ in the norm. Since the first term tends to Ug weakly and the second term tends to g weakly we must have $Ug = g$.

APPENDIX: Connections with Number Theory

Let $0 < x < 1, x$ irrational. Define $a_1 = [\frac{1}{x}]$ and $T(x) = \frac{1}{x} - [\frac{1}{x}]$. T is the Gauss transformation defined earlier. Note that $0 < T(x) < 1$. Now let $a_2 = [\frac{1}{Tx}]$ since $T(x) \neq 0$. Then $x = \frac{1}{a_1+T(x)}$ and $x = \frac{1}{a_1+\frac{1}{a_2+T^2(x)}}$. also $T^2(x) \neq 0$ since x is irrational. Proceeding this way we get an infinite sequence of positive integers $\{a_n\}$. [If x is rational then $T^k x = 0$ for some k and we get a finite sequence of integers $\{a_1, a_2, ..., a_k\}$].

For any finite or infinite sequence of positive numbers $\{t_n\}$ we define $[t_1, t_2, ..., t_n] = [t_1, t_2, ..., t_{n-2}, t_{n-1} + \frac{1}{t_n}]$ with $[t_1] = \frac{1}{t_1}$. [Thus $[t_1, t_2] = (t_1 + \frac{1}{t_2})^{-1} = \frac{1}{t_1+\frac{1}{t_2}}$ and so on].

Theorem
$x = [a_1, a_2, ..., a_n + T^n(x)]$ for each n for any $x \in (0,1)$ where $a_n = [\frac{1}{T^{n-1}x}]$.

Lemma
$[t_1, t_2, ..., t_n] = [t_1, [t_2, ..., t_n]^{-1}], n = 2, 3, ...$

Proof of the lemma: For $n = 2$ $[t_2]^{-1} = t_2$ and the result follows. Suppose the result holds for n for all choices of $t_i's$. Then $[t_1, t_2, ..., t_{n+1}] = [t_1, t_2, ..., t_{n-1}, t_n + \frac{1}{t_{n+1}}] = [t_1, [t_2, ..., t_n + \frac{1}{t_{n+1}}]^{-1}] = [t_1, [t_2, ..., t_{n+1}]^{-1}]$.

Proof of the theorem: for $n = 1$ we have $[a_1 + Tx] = \frac{1}{[\frac{1}{x}]+(\frac{1}{x}-[\frac{1}{x}])} = x$. Suppose the result holds for n and any irrational number x. Note that Tx is also an irrational number. We have $[a_1, a_2, ..., a_{n+1} + T^{n+1}(x)] = [a_1, [a_2, ..., a_{n+1} + T^{n+1}(x)]^{-1}]$ (by the lemma)
$= [a_1, (Tx)^{-1}]$ (by the induction hypothesis with Tx in place of x)
$= \frac{1}{a_1+Tx} = x$.
Now we fix a positive integer N and define $p_0 = 0, p_1 = 1, q_0 = 1, q_1 = a_1, p_n = a_n p_{n-1} + p_{n-2}, q_n = a_n q_{n-1} + q_{n-2}$ for $2 \leq n < N$ and $p_N = (a_N + T^N x)p_{N-1} + p_{N-2}, q_N = (a_N + T^N x)q_{N-1} + q_{N-2}$.

Theorem
$[a_1, a_2, ..., a_n] = \frac{p_n}{q_n}$ for $1 \leq n < N$ and $[a_1, a_2, ..., a_N + T^N x](= x) = \frac{p_N}{q_N}$.

Proof: the identity in the first part can be stated without any reference to the irrational number x. We start with positive numbers $a_1, a_2, ..., a_N$, define $p_n, q_n (1 \le k \le N)$ as above and prove that $[a_1, a_2, ..., a_n] = \frac{p_n}{q_n}$ for $1 \le n < N$. We have $[a_1] = \frac{1}{a_1} = \frac{p_1}{q_1}$. Suppose the result holds for all choices of $a_k' s$ when $n \le N-2$. Then $[a_1, a_2, ..., a_{n+1}] = [a_1, a_2, a_3, ..., a_{n-1}, a_n + \frac{1}{a_{n+1}}] = (\frac{r_n}{s_n})$ where $r_n = (a_n + \frac{1}{a_{n+1}})p_{n-1} + p_{n-2}, s_n = (a_n + \frac{1}{a_{n+1}})q_{n-1} + q_{n-2}$. Thus $\frac{r_n}{s_n} = \frac{(a_n + \frac{1}{a_{n+1}})p_{n-1} + p_{n-2}}{(a_n + \frac{1}{a_{n+1}})q_{n-1} + q_{n-2}} = \frac{p_n + \frac{1}{a_{n+1}}p_{n-1}}{q_n + \frac{1}{a_{n+1}}q_{n-1}} = \frac{a_{n+1}p_n + p_{n-1}}{a_{n+!}q_n + q_{n-1}} = \frac{p_{n+1}}{q_{n+1}}$. It remains to show that $[a_1, a_2, ..., a_N + T^N x] = x$.

The left side is $\frac{(a_{N-1} + \frac{1}{a_N + T^N x})p_{N-2} + p_{N-3}}{(a_{N-1} + \frac{1}{a_N + T^N x})p_{N-2} + p_{N-3}} = \frac{p_{N-1} + \frac{p_{N-2}}{a_N + T^N x}}{q_{N-1} + \frac{p_q N-2}{a_N + T^N x}} = \frac{p_N}{q_N}$.

Theorem
$\left| x - \frac{p_{N-1}}{q_{N-1}} \right| \le \frac{1}{q_{N-1}^2} \le \frac{1}{(N-1)^2}$.

Proof: we begin by verifying that $p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1}, 1 \le n \le N$. For $n = 1$ we have $p_n q_{n-1} - q_n p_{n-1} = 1 = (-1)^{1-1}$. Suppose the identity holds for some $n < N-1$. Then $p_{n+1}q_n - q_{n+1}p_n = (a_{n+1}p_n + p_{n-1})q_n - (a_{n+1}q_n + q_{n-1})p_n = p_{n-1}q_n - q_{n-1}p_n = -(p_n q_{n-1} - q_n p_{n-1}) = (-1)^n$. It remains to verify that the identity holds of $n = N$. We have $p_N q_{N-1} - q_N p_{N-1} = \{(a_N + T^N x)p_{N-1} + p_{N-2}\}q_{N-1} - \{(a_N + T^N x)q_{N-1} + q_{N-2}\}p_{N-1} = p_{N-2}q_{N-1} - q_{N-2}p_{N-1} = (-1)^{N-1}$.

We now have $\frac{p_N}{q_N} - \frac{p_{N-1}}{q_{N-1}} = \frac{(-1)^{N-1}}{q_N q_{N-1}}$ and $\frac{p_N}{q_N} = x$ so $\left| x - \frac{p_{N-1}}{q_{N-1}} \right| = \frac{1}{q_N q_{N-1}}$. To show that $q_N \ge q_{N-1}$ just note that $q_N = (a_N + T^N x)q_{N-1} + q_{N-2} \ge a_N q_{N-1} \ge q_{N-1}$. Finally we show that $1 \le q_1 < q_2 < ... < q_{N-1} < q_N$ (so that $q_{N-1} \ge N-1$ and the theorem follows from this). We have $q_{n+1} = a_{n+1}q_n + q_{n-1} \ge q_n + q_{n-1} > q_n$.

Corollary
$\{c^n : n = 1, 2, ...\}$ is dense in S^1 if $c \in S^1$ is not a root of unity.

[We have already proved this; what follows is an alternative proof]

Proof: let $c = e^{2\pi i \alpha}, \alpha \in [0, 1)$. Then α is irrational. Define p_n and q_n as above with $x = \alpha$.

p_n, q_n are positive integers such that $\left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}$. Also $p_n, q_n \to \infty$ and $\{\frac{p_{2n}}{q_{2n}}\}$ is increasing. [$p_{2n}q_{2n-1} - q_{2n}p_{2n-1} = (-1)^{2n-1} = -1$ and $p_{2n-1}q_{2n-2} - q_{2n-1}p_{2n-2} = (-1)^{2n-2} = 1$ so $\frac{p_{2n}}{q_{2n}} - \frac{p_{2n-2}}{q_{2n-2}} = -\frac{1}{q_{2n-1}q_{2n}} + \frac{1}{q_{2n-1}q_{2n-2}}$. Hence $\frac{p_{2n}}{q_{2n}} - \frac{p_{2n-2}}{q_{2n-2}} = \frac{q_{2n} - q_{2n-2}}{q_{2n}q_{2n-1}q_{2n-2}} \ge 0$].

Let $(a, b) \subset (0, 1)$ and $N = [\frac{a}{q_{2n}\alpha - p_{2n}}] + 1$. Let $x = N(q_{2n}\alpha - p_{2n})$. Clearly $x > a$. We also have $x \le a + q_{2n}\alpha - p_{2n} < b$ for n sufficiently large because $q_{2n}\alpha - p_{2n} < \frac{1}{q_{2n}}$. Thus any open interval $(a, b) \subset (0, 1)$ intersects $\{k\alpha + j : k \ge$

$1, j \in \mathbb{Z}\}$. In other words $\{k\alpha + j : k \geq 1, j \in \mathbb{Z}\}$ is dense in $[0,1]$. Now if $e^{i2\pi t} \in S^1$ with $0 \leq t \leq 1$ then $\left|e^{i2\pi t} - c^k\right| = \left|e^{i2\pi t} - e^{2\pi ik\alpha}e^{2\pi in}\right| \leq 2\pi \left|t - (k\alpha + n)\right|$ for any integer n. We can choose $k \geq 1$ and n such that $\left|e^{i2\pi t} - c^k\right| < \epsilon$.

[The fact that $\{k\alpha + j : k \geq 1, j \in \mathbb{Z}\}$ is dense in $[0,1]$ is also proved in Measure Theory by Halmos].

We now prove ergodicity of T under Gauss transformation.

For given positive integers $a_1, a_2, ...$ let $\Delta_{a_1,a_2,...,a_N} = \{x : a_k(x) = a_k, 1 \leq k \leq N\}$ where $a_k(x) = [\frac{1}{T^{n-1}(x)}]$. Let $p_n, q_n, 1 \leq n < N$ be defined as above and let $p_N = a_N p_{N-1} + p_{N-2}, q_N = a_N q_{N-1} + q_{N-2}$. We claim that (modulo a set of measure 0), $\Delta_{a_1,a_2,...,a_N} = [\frac{p_N}{q_N}, \frac{p_N+p_{N-1}}{q_N+q_{N-1}})$ if N is even and $\Delta_{a_1,a_2,...,a_N} = [\frac{p_N+p_{N-1}}{q_N+q_{N-1}}, \frac{p_N}{q_N})$ if N is odd. To see this let $f_{a_1,a_2,...,a_N}(t) = [a_1, a_2, ..., a_N + t]$. This is a monotone function and its range is the interval with end points $\frac{p_N}{q_N}$ and $\frac{p_N+p_{N-1}}{q_N+q_{N-1}}$. Clearly $\frac{p_N}{q_N} < \frac{p_N+p_{N-1}}{q_N+q_{N-1}}$ if N is even and the reverse inequality holds if N is even. This proves our claim. For fixed N these intervals form a partition of $[0,1)$ and the lengths of the intervals in the $N-th$ partition tend to 0 as $N \to \infty$. [$\left|\frac{p_N+p_{N-1}}{q_N+q_{N-1}} - \frac{p_N}{q_N}\right| = \frac{1}{(q_N+q_{N-1})q_N} \leq \frac{1}{N(2N-1)}$]. It follows from this that the sets of the type $\Delta_{a_1,a_2,...,a_N}$ generate the Borel sigma field of $[0,1)$. [If (a,b) is an open interval and $\omega \in (a,b)$ then we can choose N so large that the interval around ω with length $\frac{2}{N(2N-1)}$ is contained in (a,b). For this N there exist $a_1, a_2, ..., a_N$ such that $\omega \in \Delta_{a_1,a_2,...,a_N}$ since these sets form a partition of Ω. Clearly $\Delta_{a_1,a_2,...,a_N} \subset (a,b)$. It follows that (a,b) is a union of sets of the type $\Delta_{a_1,a_2,...,a_N}$. Thus the sigma field generated by sets of the type $\Delta_{a_1,a_2,...,a_N}$ contains all open intervals].

Ergodicity of T: We write f for $f_{a_1,a_2,...,a_N}$ (for fixed $a_i's$) and Δ for $\Delta_{a_1,a_2,...,a_N}$. The length of Δ is $\pm(f(1) - f(0))$. The interval $\{x : \alpha \leq T^N x < \beta\} \cap \Delta$ has length $\pm(f(\beta) - f(\alpha))$ (plus sign if N is even and minus sign otherwise). This follows from the fact that $x = [a_1, a_2, ..., a_N + T^N(x)]$ and the continuity and strict monotonicity of f. Hence $m(T^{-N}[\alpha,\beta)|\Delta) = \frac{f(\beta)-f(\alpha)}{f(1)-f(0)} = (\beta - \alpha)\frac{q_N(q_N+q_{N-1})}{(q_N+\alpha q_{N-1})(q_N+\beta q_{N-1})}$ using the fact that $f(t) = \frac{p_N+tp_{N-1}}{q_N+tq_{N-1}}$. [Note that $[a_1, a_2, ..., a_N + t]$ is obtained by replacing a_N by $a_N + t$ in $[a_1, a_2, ..., a_N] = \frac{a_N p_{N-1}+qp_{N-2}}{a_N q_{N-1}+q_{N-2}}$. Hence $f(t) = \frac{a_N p_{N-1}+p_{N-2}+tp_{N-1}}{a_N q_{N-1}+q_{N-2}+tq_{N-1}} = \frac{p_N+tp_{N-1}}{q_N+tq_{N-1}}$].

Noting that $q_{N-1} \leq q_N$ and hence $\frac{1}{2} \leq \frac{q_N(q_N+q_{N-1})}{(q_N+\alpha q_{N-1})(q_N+\beta q_{N-1})} \leq 2$ we see that $\frac{1}{2}(\beta-\alpha) \leq m(T^{-N}[\alpha,\beta)|\Delta) \leq 2(\beta-\alpha)$. Hence $\frac{1}{2}m(A) \leq m(T^{-N}A|\Delta) \leq 2m(A)$ for any Borel set A. If P denotes the Gauss measure ($dP = \frac{1}{(\ln 2)(1+x)}$) we get $c_1 m(A) \leq P(T^{-N}A|\Delta) \leq c_2 m(A)$ for suitable $c_1, c_2 \in (0,\infty)$ (because the density of P is bounded above and below). If A is invariant this gives $c_1 m(A)P(\Delta) \leq P(A \cap \Delta) \leq c_2 m(A)P(\Delta)$. Since intervals of the type Δ generate the Borel sigma field we can replace Δ by any Borel set. [This requires the $\pi - \lambda$ theorem. (c.f. Theorem 3.2 of Probability and Measure By Billingsley)]. If $P(A) > 0$ then $m(A) > 0$ and we can take $\Delta = A^c$ to get $P(A^c) = 0$.

Corollary:

$(a_1(x)a_2(x)...(a_n(x))^{1/n} \to \prod_{k=1}^{\infty}(1+\frac{1}{k^2+2k})^{(\ln k)/(\ln 2)}$ as $n \to \infty$ for almost all $x \in (0,1)$.

Remark: the limit is called *Khinchine's constant.* Is it rational or irrational? Algebraic or transcendental? The answers are not known!

Proof: let $f(x) = \ln a_1(x)$. Then $f(T^n(x)) = \ln a_{n+1}(x)$ and the ergodic theorem gives $(a_1(x)a_2(x)...(a_n(x))^{1/n} = e^{\frac{1}{n}\sum_{j=1}^{n} \ln a_j(x)} \to e^{\int \ln a_1(x)dP(x)} = e^{\int \ln[\frac{1}{x}]dP(x)}$ (here $[t]$ is the greatest integer not exceeding t). Now $\int \ln[\frac{1}{x}]dP(x) = \sum_{m=1}^{\infty} \frac{\ln m}{\ln 2} \int_{1/(m+1)}^{1/m} \frac{1}{1+x}dx = \sum_{m=1}^{\infty} \frac{\ln m}{\ln 2}[\ln(1+\frac{1}{m}) - \ln(1+\frac{1}{m+1})] = \sum_{m=1}^{\infty} \frac{\ln m}{\ln 2} \ln(1+\frac{1}{m^2+m})$. This completes the proof.

Remark: as a further application of ergodic theory to number theory it can be shown that $\frac{1}{n} \ln \left| x - \frac{p_{N-1}}{q_{N-1}} \right| \to -\frac{\pi^2}{6\ln(2)}$ a.e..

Theorem

$m\{x \in [0,1) : T^n(x) \leq a\} \to P([0,a])$ for all a.

[Equivalently $\int f(T^n)dm \to \int fdP$ for every bounded continuous function f on $[0,1)$]

Proof: this requires several steps. We first prove that T is a mixing transformation in the sense $P(T^{-n}A \cap B) \to P(A)P(B)$ for any two sets $A, B \in \mathcal{F}$. [This is stronger than ergodicity; by taking $B = A$ where A is invariant we see that mixing implies ergodicity]. To prove that T is mixing we prove that the following $0-1$ law holds: let $\mathcal{G}_n = \sigma\{a_n, a_{n+1}, ...\}$ where $a_n(x)$ is the $n-th$ integer in the continued fraction expansion of x. Let $\mathcal{G}_\infty = \bigcap_{n=1}^{\infty} \mathcal{G}_n$. Then for every A set in $\mathcal{G}_\infty$ we have $P(A) = 0$ or1. For this note that $A \in \mathcal{G}_n$ for each n. Fix n. There exists a set B in the Borel sigma field of $\mathbb{R}^\infty$ such that $A = T^{-n}B$ (why?). Hence $P(A) = P(T^{-n}(B)) = P(B) \leq 2P(T^{-N}B|\Delta_N)$ (by the third last line in the proof of ergodicity of T). Thus $P(A) \leq 2P(A|\Delta_N)$. Suppose $P(A) > 0$. Then the inequality $P(A)P(\Delta_N) \leq 2P(A \cap \Delta_N)$ for all intervals of the type Δ_N shows that $P(A)P(B) \leq 2P(A \cap B)$ for all A and B. Taking $B = A^c$ we get $P(B^c) = 0$ or $P(A) = 1$. This finishes the proof of the fact that $\mathcal{G}_\infty$ is trivial with respect to P. Now let $Y_n = P(A|\mathcal{G}_n) - P(A)$ where $A \in \mathcal{F}$ is arbitrary. By Martingale Convergence Theorem it follows that $Y_n \to P(A|\mathcal{G}_\infty) - P(A) = 0$ almost surely and in L^1 since A is independent of $\mathcal{G}_\infty$. For any $B \in \mathcal{F}$ we have

$T^{-n}B \in \mathcal{G}_n$ and hence $\int\limits_{T^{-n}(B)} Y_n dP (= P(A \cap T^{-n}(B)) - P(A)P(T^{-n}(B))) \to 0$ which says $P(A \cap T^{-n}(B)) - P(A)P(B) \to 0$. Thus T is mixing. To complete the proof we note that $m\{x \in [0,1) : T^n(x) \leq a\} = \int\limits_{T^{-n}A} \ln(2)(1+x)dP(x)$ where $A = [0,a]$. It suffices to show that $\int\limits_{T^{-n}A} f(x)dP(x) \to P(A)\int f dP$ for any bounded measurable function f. [Because $\ln(2)(1+x)$ is bounded and measurable and $\int \ln(2)(1+x)dP = 1$]. By simple function approximation it suffices to prove this when f is the indicator function of a measurable set B. We have to show that $P(B \cap T^{-n}(A)) \to P(A)P(B)$ which is precisely the statement that T is mixing.

Remark
The following general result follows from above arguments:

Theorem
Let $(\Omega, P, \mathcal{F}, T)$ be a DS. Let $\mathcal{G}_n$ be the sigma field $\{T^{-n}(A) : A \in \mathcal{F}\}$. Then $\mathcal{G}_{n+1} \subset \mathcal{G}_n$ for all n. If the sigma field $\mathcal{G} = \bigcap\limits_{n=1}^{\infty} \mathcal{G}_n$ is trivial (in the sense every set in this sigma field has probability 0 or 1 then T is mixing.
Details of the proof are left to the reader.

Further properties of continued fraction expansions:
We have $\omega = \frac{p_{N+T^n\omega p_{N-1}}}{q_{N+T^n\omega p_{N-1}}}$ and $\left|\omega - \frac{p_N}{q_N}\right| = \frac{1}{q_{N\{\frac{1}{T^N\omega}q_N + q_{N-1})}}$. Using the fact that $a_{N+1}(\omega) = [\frac{1}{T^N\omega}]$ we get $a_{N+1} \leq \frac{1}{T^N\omega} < a_{N+1} + 1$ and hence $\frac{1}{q_{N\{\frac{1}{T^N\omega}q_N+q_{N-1})}} \leq \frac{1}{q_{N\{a_{N+1}q_N+q_{N-1})}}$ and $\frac{1}{q_{N\{\frac{1}{T^N\omega}q_N+q_{N-1})}} > \frac{1}{q_{N\{\{a_{N+1}+1)q_N+q_{N-1})}} = \frac{1}{q_N\{q_{N+1}+q_N)}$. Thus $\frac{1}{q_N\{q_{N+1}+q_N)} < \left|\omega - \frac{p_N}{q_N}\right| \leq \frac{1}{q_{N\{a_{N+1}q_N+q_{N-1})}}$. We now show that

$\left|\log \frac{\omega}{p_N(\omega)/q_N(\omega)}\right| \leq \frac{1}{2^{N-2}}$. For $N = 1$ $\left|\log \frac{\omega}{p_N(\omega)/q_N(\omega)}\right| = |\log\{\omega a_1(\omega)\}| = \left|\log\{\omega[\frac{1}{\omega}]\}\right| = (\log\{\omega[\frac{1}{\omega}]\})^{-1}$. Considering the cases $\omega > 1/2$ and $\omega \leq 1/2$ we can see easily that the inequality holds in this case. From the definition of the sequences $\{p_n\}, \{q_n\}$ we see that $p_n \geq 2^{(n-2)/2}, q_n \geq 2^{(n-2)/2} (n = 2, 3, ...)$. It follows, by induction, that $\left|\frac{\omega}{p_N(\omega)/q_N(\omega)} - 1\right| \leq \frac{\sqrt{2}}{2^{N-1}} (n = 2, 3, ...)$. [$\left|\frac{\omega}{p_N(\omega)/q_N(\omega)} - 1\right| = |\omega - p_N(\omega)/q_N(\omega)| \frac{1}{p_N(\omega)/q_N(\omega)} \leq \frac{1}{q_{N\{a_{N+1}q_N+q_{N-1})}} \frac{1}{p_N(\omega)/q_N(\omega)}$ $= \frac{1}{\{a_{N+1}q_N+q_{N-1})} \frac{1}{p_N(\omega)} \leq \frac{1}{\{2^{(N-2)/2}+2^{(N-3)/2}\}2^{(N-2)/2}} \leq \frac{\sqrt{2}}{2^{N-1}}$. This gives $\left|\log \frac{\omega}{p_N(\omega)/q_N(\omega)}\right| \leq \frac{\sqrt{2}}{2^{N-1}-\sqrt{2}}$. [Indeed it is easy to check that $|t-1| \leq \frac{\epsilon}{1+\epsilon} \Rightarrow$

$|\log t| \leq \epsilon$ (: consider the cases $t < 1$ and $t \geq 1$ separately). Here we take $\epsilon = \frac{\sqrt{2}}{2^{N-1}-\sqrt{2}}$. We now use the inequality $\left|\log \frac{\omega}{p_N(\omega)/q_N(\omega)}\right| \leq \frac{\sqrt{2}}{2^{N-1}-\sqrt{2}}$ to prove the following

Theorem
$\frac{1}{N} \log q_N \to \frac{\pi^2}{12 \log 2}$ a.s..

Proof: claim: $\frac{1}{q_N} = \prod_{k=1}^{N} [a_k, a_{k+1}, ..., a_N]$. For this we first prove that $p_{n+1}(\omega) = q_n(T\omega)$. If we denote $q_n(T\omega)$ by $r_n(\omega)$ we get $r_n(\omega) = a_n(T\omega)q_{n-1}(T\omega) + q_{n-2}(T\omega) = a_{n+1}(\omega)r_{n-1}(\omega) + r_{n-2}(\omega)$. We also have $p_{n+1}(\omega) = a_{n+1}(\omega)p_n(\omega) + p_{n-1}(\omega)$. An inspection of these equations makes it clear that the equation $p_{n+1}(\omega) = q_n(T\omega)$ holds for all n provided it holds for $n = 0$ and 1. Since $p_{n+1}(\omega) = q_n(T\omega) = 1$ for $n = 1$ and $p_{n+1}(\omega) = q_n(T\omega) = a_2(\omega)$ for $n = 1$ we have finished the proof of $p_{n+1}(\omega) = q_n(T\omega)$. It follows that $\prod_{k=1}^{N} \frac{p_{N+1-k}(T^{k-1}\omega)}{q_{N+1-k}(T^{k-1}\omega)}$ is a telescopic product and its value is $\frac{1}{q_N(\omega)}$. But the product here is nothing but $\prod_{k=1}^{N} [a_k, a_{k+1}, ..., a_N]$ and so we have proved the claim. Using the inequality $\left|\log \frac{\omega}{p_N(\omega)/q_N(\omega)}\right| \leq \frac{\sqrt{2}}{2^{N-1}-\sqrt{2}}$ with ω changed to $T^{k-1}(\omega)$ we get

$\left|\log(T^{k-1}\omega) - \log[a_k, a_{k+1}, ..., a_N]\right| \leq \frac{\sqrt{2}}{2^{N-k-1}-\sqrt{2}}$. Now $\frac{1}{N} \log q_N = -\frac{1}{N} \sum_{k=1}^{N} \log[a_k, a_{k+1}, ..., a_N] = -\frac{1}{N} \sum_{k=1}^{N} \log T^{k-1}(\omega) + \alpha_N$ where $\alpha_N = -\frac{1}{N} \sum_{k=1}^{N} \frac{\sqrt{2}}{2^{N-k-1}-\sqrt{2}}$. By the Ergodic Theorem the first term converges to $-\frac{1}{\log 2} \int_0^1 \log x \frac{1}{1+x} dx = \frac{1}{\log 2} \int_0^1 \log(1+x) \frac{1}{x} dx$ [Note that $|\log x| \log(1+x) \leq x\,|\log x| \to 0$ as $x \to 0$ and $\log x \log(1+x) \to 0$ as $x \to 1$]. Expanding $\log(1+x)$ as $x - x^2/2 + x^3/3 + ...$ we get $\frac{1}{\log 2} \int_0^1 \log x \frac{1}{1+x} dx = \frac{1}{\log 2}[1 - \frac{1}{2^2} + \frac{1}{3^2} + ...] = \frac{\pi^2}{12 \log 2}$. Since $\alpha_n \to 0$ we can conclude that $\frac{1}{N} \log q_N \to \frac{\pi^2}{12 \log 2}$.

Theorem
$\frac{1}{n} \log \left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| \to -\frac{\pi^2}{6 \log 2}$ as $n \to \infty$.

Proof: recall that $\frac{1}{q_N\{q_{N+1}+q_N\}} < \left|\omega - \frac{p_N}{q_N}\right| \leq \frac{1}{q_N q_{N+1}}$. Hence $\frac{1}{n} \log \left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| \leq \frac{1}{n} \log \frac{1}{q_N q_{N+!}}$

$= -\frac{1}{n}\log q_n - \frac{1}{n}\log q_{N+1} \to -\frac{\pi^2}{6\log 2}$ and $\frac{1}{n}\log\left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| \geq \frac{1}{n}\log\frac{1}{q_N\{q_{N+1}+q_N\}} \geq \frac{1}{n}\log\frac{1}{2q_Nq_{N+1}} \to -\frac{\pi^2}{6\log 2}$.

Remark: $\left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| < e^{n[\epsilon - \frac{\pi^2}{6\log 2}]}$ for n sufficiently large. Thus the rational numbers $\frac{p_n(\omega)}{q_n(\omega)}$ converge to ω exponentially fast.

[END OF APPENDIX]

Mixing transformations

Let $(\Omega, \mathcal{F}, P, T)$ be a dynamical system. We say T is *mixing* if $P(T^{-n}(A) \cap B) \to P(A)P(B)$ as $n \to \infty$ for any two measurable sets A and B. We say T is *weakly mixing* if $\frac{1}{n}\sum_{k=1}^{n}\left|P(T^{-k}(A)\cap B) - P(A)P(B)\right| \to 0$ as $n \to \infty$ for any two measurable sets A and B.

Remark: 'mixing' is also called 'strong-mixing'.

Theorem

a) T is ergodic iff $\frac{1}{n}\sum_{k=1}^{n} P(T^{-k}(A)\cap B) \to P(A)P(B)$ for any two measurable sets A and B.

b) Weak mixing implies ergodic

c) Mixing implies weak mixing

If any of the properties above hold for A and B in a generating field they hold for all A and B.

Proof: for the 'if' part of a) take A to be an invariant set and $B = A$. This gives $P(A) = P^2(A)$ and hence T is ergodic. Now suppose T is ergodic. By the Ergodic Theorem $\frac{1}{n}\sum_{k=1}^{n} I_A(T^k(x)) \to \int I_A dP = P(A)$ a.e. and in L^1. Hence $\int_B \frac{1}{n}\sum_{k=1}^{n} I_A(T^k(x))dP \to P(A)P(B)$ and the left side is $\frac{1}{n}\sum_{k=1}^{n} P(T^{-k}(A)\cap B)$. This proves a).

b) is obvious from a).

c) is elementary.

The last part of the theorem is proved by straightforward arguments using that fact that if $\mathcal{F}_0$ is a field that generates $\mathcal{F}$ then, given $A \in \mathcal{F}$ and $\epsilon > 0$

there exists $B \in \mathcal{F}_0$ such that $P(A\Delta B) < \epsilon$. The details are left to the reader.

Theorem

a) T is ergodic iff $\frac{1}{n}\sum_{k=1}^{n} < U^k f, g > \to < f, 1 >< 1, g >$ for all $f, g \in L^2$ where $Uf = f \circ T$ and $<,>$ is the inner product in L^2.

b) T is weak mixing iff $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \to 0$ for all $f, g \in L^2$

iff $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, f > - < f, 1 >< 1, f >\right| \to 0$ for all $f \in L^2$

iff $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, f > - < f, 1 >< 1, f >\right|^2 \to 0$ for all $f \in L^2$

c) T is mixing iff $< U^k f, g > \to < f, 1 >< 1, g >$ for all $f, g \in L^2$

iff $< U^k f, f > \to < f, 1 >< 1, f >$ for all $f \in L^2$

Proof: 'if' part of a) follows by taking f and g to be indicators. If T is ergodic then

$\frac{1}{n}\sum_{k=1}^{n} U^k f \to < f, 1 >$ in L^1. Multiply by g and integrate to get $\frac{1}{n}\sum_{k=1}^{n} < U^k f, g > \to < f, 1 >< 1, g >$.

Next we show that T is weak mixing iff $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \to 0$ for all $f, g \in L^2$.

If T is weak mixing then $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \to 0$ whenever f and g are indicators. It follows easily that the same is true when f and g are simple functions. The general case follow from the estimate $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \leq \frac{1}{n}\sum_{k=1}^{n} [\varepsilon^2 + \varepsilon^2] = 2\varepsilon^2$ if $\|f\|_2 < \varepsilon$ and $\|g\|_2 < \varepsilon$. The converse part is trivial.

Now suppose this property holds when $f = g$. To prove that it holds for all f and g let $M_f = \{g \in L^2 : \frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \to 0\}$. It is easy to see (using triangle inequality as in the proof of b)) that this is a closed subspace of L^2. Note that $1 \in M_f$ and $f \in M_f$ (by hypothesis). Also $g \in M_f \Rightarrow Ug \in M_f$. Let $h \in M_f^{\perp}$. But this implies $< U^k f, h > - < f, 1 >< 1, h >= 0$ for all k [because $U^k f \in M_f$ and $1 \in M_f$] and so $h \in M_f$ by definition of M_f. Thus $h \in M_f^{\perp} \cap M_f = \{0\}$. This proves that $M_f = L^2$ and so $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, g > - < f, 1 >< 1, g >\right| \to 0 \ \forall f, g \in L^2$.

If $\{a_n\}$ is a bounded sequence of real (or complex) numbers then $\frac{1}{n}\sum_{k=1}^{n}|a_k| \to 0$ iff $\frac{1}{n}\sum_{k=1}^{n}|a_k|^2 \to 0$. [See Appendix below]. This completes b) and c) is proved similarly.

Remark: a m.p.t. T may be ergodic without T^2 being so: let $\Omega = \{-1, 1\}$ with the power set as the sigma filed and the uniform measure P. Let $T(1) = -1$ and $T(-1) = 1$. T is clearly ergodic but $\{1\}$ is an invariant set for T^2 with measure $\frac{1}{2}$.

Definition: a m.p.t. is *totally ergodic* if T^n is ergodic for $n = 1, 2, ...$

Theorem

A weak mixing transformation is totally ergodic.

Proof: given: $\frac{1}{n}\sum_{k=0}^{n-1}\left|P(T^{-k}(A)\cap B) - P(A)P(B)\right| \to 0$ as $n \to \infty$. If $m \in \mathbb{N}$ and $T^{-m}(A) = A$ then we get $\sum_{k=0}^{m-1}\left|P(T^{-k}(A)\cap A) - P(A)P(A)\right| = 0$. [Split the sum over $0 \le k < n$ into $0 \le k < m, m \le k < 2m, ...$ taking n to be a multiple of m]. In particular $P(T^{-0}(A)\cap A) - P(A)P(A) = 0$ so $P(A) = P^2(A)$.

Example: let α be an irrational number in $(0, 1)$ and define T on $[0, 1)$ with the Lebesgue measure by $Tx = x + \alpha \,\mathrm{mod}(1)$. Then T is totally ergodic but not weak mixing.

We already know that $T^n x = x + n\alpha \,\mathrm{mod}(1)$ is ergodic. The fact that T is not weak mixing will be proved later. [See the remark after the first theorem that follows the appendix below].

APPENDIX
A THEOREM ON CESARO CONVERGENCE

Theorem

Let $\{a_n\} \subset \mathbb{R}$ be bounded and $1 < p < \infty$. Then the following are equivalent:

a) $\frac{1}{n}\sum_{k=1}^{n}|a_k| \to 0$

b) there exists $A \subset \mathbb{N}$ such that $\lim_{n\notin A, n\to\infty} a_n = 0$ and $\frac{\#\{A\cap\{1,2,...,n\})}{n} \to 0$ as $n \to \infty$

c) $\frac{1}{n}\sum_{k=1}^{n}|a_k|^p \to 0$

Proof: it suffices to show that a) and b) are equivalent. Suppose $\frac{1}{n}\sum_{k=1}^{n}|a_k| \to 0$. For $k = 1, 2, ...$ let $I_k = \{n \geq 1 : |a_n| \geq \frac{1}{k}\}$. Claim: $\frac{\#\{I_k \cap [1,n]\}}{n} \to 0$ as $n \to \infty$ for each k. Indeed, this follows from the inequality $\frac{1}{n}\sum_{j=1}^{n}|a_j| \geq \frac{\#\{I_k \cap [1,n]\}}{nk}$. There exist integers $n_0 < n_1 < ...$ such that $n \geq n_k$ implies $\frac{\#\{I_{k+1} \cap [1,n]\}}{n} < \frac{1}{k+1}$. Let $I = \bigcup_{k=0}^{\infty}\{I_{k+1} \cap [n_k, n_{k+1})\}$. Let $n_k \leq n < n_{k+1}$. Then $I \cap [1,n] \subset \{I_k \cap [1, n_k]\} \cup \{I_{k+1} \cap [1,n]\}$. [Let $m \in I \cap [1,n]$. Then $m \in I_{r+1} \cap [n_r, n_{r+1})$ for some r. Since $n_r \leq m \leq n < n_{k+1}$ we have $r \leq k$. If $r = k$ then $m \in I_{r+1} \cap [n_r, n_{r+1}) = I_{k+1} \cap [1,n]$. If $r < k$ then $r+1 \leq k$ and $m \in I_{r+1} \subseteq I_k$. Remains to show that $m \leq n_k$. But $m < n_{r+1} \leq n_k$so we are done]. Hence $\frac{\#\{I \cap [1,n]\}}{n} \leq \frac{\#\{I_k \cap [1,n_k]\}}{n} + \frac{\#\{I_{k+1} \cap [1,n]\}}{n} < \frac{1}{k} + \frac{1}{k+1} \leq \frac{1}{k} + \frac{1}{k+1}$ if $n \geq n_k$. We have proved that $\frac{\#\{I \cap [0,n)\}}{n} \to 0$ as $n \to \infty$. If $n > n_k$ and $n \notin I$ then $n \notin I_{k+1}$ (if $n > n_k$ and $n \in I_{k+1}$ then there exists $\rho \geq k$ such that $n_\rho \leq n < n_{\rho+1}$ and $n \in I_{k+1} \subset I_{\rho+1}$ so $n \in I_{\rho+1} \cap [n_\rho, n_{\rho+1)} \subset I$). Thus $|a_n| < \frac{1}{k+1}$ for $n > n_k$, $n \notin I$ completing the proof of a) implies b). For the converse part let $|a_n| \leq C$ and let $\epsilon > 0$. There exists n_ϵ such that $|a_n| < \epsilon$ if $n > n_\epsilon$ and $n \notin I$. Also there exists m_ϵ such that $\frac{\#\{I \cap [0,n)\}}{n} < \epsilon$ if $n > m_\epsilon$. For $n > \max\{n_\epsilon, m_\epsilon\}$ we have $\frac{1}{n}\sum_{k=0}^{n-1}|a_k| < \epsilon + \epsilon C$.

END OF APPENDIX

Theorem

Let $(\Omega, \mathcal{F}, P, T)$ be a DS. Then the following are equivalent:

1) T is weak mixing

2) $T \times T$ is ergodic

3)$T \times T$ is weak mixing

Proof: 1) implies 3): it suffices to show that there is a set $S \subset \mathbb{N}$ such that $\frac{\#(S \cap [1,n])}{n} \to 1$ and $\lim_{n \in S}(P \times P)((T \times T)^{-n}(A \times B) \cap (C \times D)) \to P(A)P(B)P(C)P(D)$ for all measurable sets A, B, C, D. By 1) there exist sets $S_1, S_2 \subset \mathbb{N}$ such that $\frac{\#(S_i \cap [1,n])}{n} \to 1 (i = 1, 2)$ and $\lim_{n \in S_1} P(T^{-n}A \times B) \to P(A)P(B)$ and $\lim_{n \in S_2} P(T^{-n}C \times D)) \to P(C)P(D)$. But $(P \times P)((T \times T)^{-n}(A \times B) \cap (C \times D)) = P(T^{-n}A \times B)P(T^{-n}C \times D))$. To complete the proof of 1) implies 3) we only have to take $S = S_1 \cap S_2$. [Note that $\frac{\#(S^c \cap [1,n])}{n} \leq \frac{\#(S_1^c \cap [1,n])}{n} + \frac{\#(S_2^c \cap [1,n])}{n} \to 0$].

3) certainly implies 2). We now prove that 2) implies 1). Consider $\frac{1}{n}\sum_{k=1}^{n}\{P(T^{-k}A \cap B) - P(A)P(B)\}^2$

$= \frac{1}{n}\sum_{k=1}^{n} P^2(T^{-k}A \cap B) + P^2(A)P^2(B) - \{\frac{2}{n}\sum_{k=1}^{n} P(T^{-k}A \cap B)\}P(A)P(B)\}.$

In the last term we write $\frac{1}{n}\sum_{k=1}^{n} P(T^{-k}A\cap B) = \frac{1}{n}\sum_{k=1}^{n}(P\times P)((T\times T)^{-k}(A\times\Omega)\cap (B\times\Omega) \to P(A)P(B)$. Also $\frac{1}{n}\sum_{k=1}^{n} P^2(T^{-k}A\cap B) = \frac{1}{n}\sum_{k=1}^{n}(P\times P)((T\times T)^{-k}(A\times A)\cap(B\times B)) \to P^2(A)P^2(B)$. Hence $\frac{1}{n}\sum_{k=1}^{n}\{P(T^{-k}A\cap B) - P(A)P(B)\}^2 \to P^2(A)P^2(B) + P^2(A)P^2(B) - 2P^2(A)P^2(B) = 0$. This completes the proof.

We now study properties of an ergodic m.p. transformation T related to the spectrum of the operator $U : L^2 \to L^2$ defined by $Uf(\omega) = f(T(\omega))$. **In this discussion we take the scalar field to be the field of complex numbers**.

An eigen function for T with eigen value λ is a function f in $L^2\backslash\{0\}$ such that $f \circ T = \lambda f$. In other words it is eigen function for the operator U.

Lemma
$|\lambda| = 1$ and $|f|$ is a constant (a.e.)

Proof: taking L^2 norms in $f \circ T = \lambda f$ we get $|\lambda| = 1$. Hence $|f| = |\lambda f| = |f \circ T| = |f| \circ T$. Thus $|f|$ is invariant, hence constant a.e..

Lemma
Eigen functions corresponding to different eigen values are orthogonal.

Proof: this is true for any linear isometry from H into itself: $Uf_1 = \lambda_1 f_1, Uf_2 = \lambda_2 f_2, f_1 \neq 0, f_2 \neq 0, \lambda_1 \neq \lambda_2 \Rightarrow < f_1, f_2 > = < Uf_1, Uf_2 > = \lambda_1\bar{\lambda}_2 < f_1, f_2 > \Rightarrow < f_1, f_2 > = 0$ because $\lambda_1\bar{\lambda}_2 = \frac{\lambda_1}{\lambda_2} \neq 0$.

Lemma
The eigen space corresponding to a given eigen value is one dimensional.

Proof: let $Uf = \lambda f, Ug = \lambda g, f \neq 0$ and $g \neq 0$. Then $\frac{f\circ T}{f\circ T} = \frac{\lambda f}{\lambda g}$ and hence $\frac{f}{g}$ is invariant. It follows that $f = cg$ for some constant c. [Recall that $|g|$ is a constant. Hence $\{\omega : g(\omega) = 0\}$ is a null set].

Lemma
Eigen values of T form a subgroup of the multiplicative group S^1

Proof: this is obvious. [Just remember that $|f|$ is a constant for any eigen function f. If f and g are two eigen functions then $fg \neq 0$].

Theorem
For a DS $(\Omega, \mathcal{F}, P, T)$ where T is i.m.p the following are equivalent:
a) T is weak mixing

b) 1 is the only eigen value of the operator U on L^2 defined by $Uf = f \circ T$ and the eigen space corresponding to this eigen value is one dimensional.

Remark: if $Tz = az$ on S^1 where a is not a root of unity then T is ergodic, but it is not weak mixing. Indeed $Uf = af$ where $f(z) = z$ for all z.

The proof requires the following theorem from Functional Analysis:

Theorem [Spectral Theorem for unitary operators]

If U is a unitary operator on a complex Hilbert space H. If $x \in H$ then there is a unique complex Borel measure μ_x on S^1 such that $< U^n x, x >= \int z^n d\mu_x$ $\forall n \in \mathbb{Z}$. Further μ_x is a continuous measure if U has no eigen values.

[For the second part we refer to Taylor's book on Functional Analysis; see Theorem 6.5-E, p. 353]

Assuming this theorem we prove our theorem on weak mixing transformations as follows:

Proof of a) implies b): let $f \circ T = \lambda f, f \neq 0$. We have $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, f > - < f, 1 >< 1, f >\right| \to 0$. We claim that $|\lambda| = 1$ and, if $\lambda \neq 1$ then $< f, 1 >= 0$. This follows immediately from $\|f \circ T\| = |\lambda| \, \|f\|$, $\int f \circ T = \lambda \int f$ and T is m.p.. Suppose $\lambda \neq 1$. Then $\left|< U^k f, f > - < f, 1 >< 1, f >\right| =< f, f >$ for all k so $< f, f >= 0$, a contradiction. If $\lambda = 1$ then f is a constant since f is an invariant function and T is ergodic.

b) implies a): we have to prove that $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k g, g > - < g, 1 >< 1, g >\right|^2 \to 0$ for all $g \in L^2$. Let $f = g - \int g$ so $< f, 1 >= 0$. In this case we show that $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, f >\right|^2 \to 0$. A simple calculation shows that $\frac{1}{n}\sum_{k=1}^{n} \left|< U^k f, f >\right|^2 = \frac{1}{n}\sum_{k=1}^{n} \left|< U^k g, g > - < g, 1 >< 1, g >\right|^2$ and the proof would be complete. By the Spectral Theorem this reduces to $\frac{1}{n}\sum_{k=1}^{n} \left|\int \lambda^k d\mu_f\right|^2 \to 0$. Since f is orthogonal to the eigen functions of U it follows that μ_f is a continuous measure. We now prove that $\frac{1}{n}\sum_{k=1}^{n} \left|\int \lambda^k d\mu\right|^2 \to 0$ for any continuous measure μ on S^1. For this we may suppose that μ is a continuous *positive* measure. By Fubini's Theorem $\frac{1}{n}\sum_{k=1}^{n} \left|\int \lambda^k d\mu\right|^2 = \frac{1}{n}\sum_{k=1}^{n} (\int \lambda^k d\mu)(\int \lambda^{-k} d\mu) = \int\int \frac{1}{n}\sum_{k=1}^{n} \lambda^k \tau^{-k} (d\mu \times d\mu)(\lambda, \tau)$. Now for $\lambda \neq \tau$ we have $\frac{1}{n}\sum_{k=1}^{n} \lambda^k \tau^{-j} = \frac{1}{n} \frac{(\frac{\lambda}{\tau})^{n+1} - \frac{\lambda}{\tau}}{\frac{\lambda}{\tau} - 1} \to 0$. Since

$(\mu \times \mu)\{(\lambda, \tau) \in S^1 \times S^1 : \lambda = \tau\} = 0$ we see that $\frac{1}{n}\sum_{k=1}^{n}\left|\int \lambda^k d\mu\right|^2 \to 0$. [Dominated Convergence Theorem can be applied because $\left|\frac{1}{n}\sum_{k=1}^{n}\lambda^k\tau^{-k}\right| \leq 1$].

Theorem
Conditions a) and b) of previous theorem are also equivalent to:
c) $T \times T$ is ergodic on $(\Omega, \mathcal{F}, P) \times (\Omega, \mathcal{F}, P)$

Proof: (it is trivial to check that $T \times T$ is m.p.). Let T be weak mixing. To prove c) it suffices to show that $(P \times P)((T \times T)^{-n}(A \times B) \cap (C \times D)) \xrightarrow{c} P(A)P(B)P(C)P(D)$ for all $A, B, C, D \in \mathcal{F}$. ($\xrightarrow{c}$ stands for Cesaro convergence). This is true by an immediate application of the theorem on Cesaro convergence in the Appendix above.

[If $A \subset \mathbb{N}$ such that $\lim_{n \notin A, n \to \infty} |a_n - a| = 0$ and $\lim_{n \notin A, n \to \infty} |b_n - b| = 0$ where $\frac{card\{A \cap \{1,2,\ldots,n\}\}}{n} \to 0$ as $n \to \infty$ then $\lim_{n \notin A, n \to \infty} |a_n b_n - ab| = 0$. Of course union of two sets with the property $\frac{card\{A \cap \{1,2,\ldots,n\}\}}{n} \to 0$ also has this property, so the same set A can be used for $\{a_n\}$ and $\{b_n\}$].

Now suppose c) holds. Let $Uf = \lambda f, f \neq 0$. Then $F(x,y) = f(x)\bar{f}(y)$ defines an eigen function of $U \times U$ with eigen value 1 (i.e. an invariant function). $[(U \times U)F(x,y) = F(Tx, Ty) = f(T(x)\bar{f}(Ty) = \lambda f(x)\bar{\lambda}\bar{f}(y) = F(x,y)$ because $|\lambda| = 1]$. Hence $f(x)\overline{f(y)}$ is a.e. constant, say c. Now Fubini's Theorem implies that for almost all y, hence for at least one y, $f(x)\overline{f(y)} = c$. Thus f is a constant and hence $\lambda = 1$. We have proved that b) holds. The proof is complete.

Example: The rotation $Tz = az$ (a not a root of unity) on S^1 is ergodic but not weak mixing. This is clear because a is an eigen value other than 1 (with the eigen function $f(z) = z$).

It also follows from the equivalence of a) and c) that $T \times T$ is not ergodic even though T is!

Theorem
Let $(\Omega, \mathcal{F}, P, T)$ be an dynamical system with T ergodic and $(\Omega', \mathcal{F}', P', T')$ be another dynamical system with T' also ergodic. Let $Uf = f \circ T$ on $L^2(P)$ and $Vg = g \circ T'$ on $L^2(P')$. If $T \times T'$ is ergodic then U and V have no common eigen values other than 1.

Proof: if $Uf = \lambda f$ and $Vg = \lambda g$ $(\lambda \neq 1, f \neq 0, g \neq 0)$ then $(U \times V)F(x,y) = F(x,y)$ where $F(x,y) = f(x)\overline{g(y)}$. If $T \times T'$ is ergodic it would follow that F is constant. But then f and g are constants by Fubini's Theorem and this contradicts the fact $\lambda \neq 1$. Thus $T \times T'$ not is ergodic.

Remark: the converse of this is also true. We omit the proof.

Application to Markov Chain Theory

Let $P = ((p_{ij}))$ be an $N \times N$ stochastic matrix. Let π be a probability vector with $\pi P = P$. Let $S = \{1, 2..., N\}$. There exists a probability measure Q on $\Omega = S \times S \times ...$ with the sigma field $\mathcal{F}$ generated by cylinder sets such that $Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_j = i_j\} = \pi_{i_1} p_{i_1 i_2}, ..., p_{i_{j-1} i_j}$. Let T be the shift map $(T\omega)_n = \omega_{n+1}$.

Let $q_{ij} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$. To see that this limit exists for all i and j we first apply Birkhoff's ergodic theorem with $f = I_B(B \subset \{1, 2, ..., N\})$ to see that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_{T^{-k}(B)}$ exists a.e..and in L^1. If $A \subset \{1, 2, ..., N\}$ we can integrate over A to see that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q(T^{-k}(B) \cap A)$ exists. Now take $A = \{\omega : \omega_0 = i\}, B = \{\omega : \omega_0 = j\}$ to conclude that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q\{\omega : w_0 = i, \omega_k = j\} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} \pi_i$ exists. We assume that $\pi_i > 0$ for each i to conclude that q_{ij} exists for all i and j. [Those states i with $\pi_i = 0$ are irrelevant to the Markov chain].

Theorem
T is m.p.. on $(\Omega, \mathcal{F}, Q)$. Also the following are equivalent:
a) T is ergodic
b) P is irreducible
c) q_{ij} is independent of i
d) $q_{ij} > 0$ for all i and j

Proof:
T is ergodic iff
$\frac{1}{n} \sum_{k=0}^{n-1} Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r, \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\} \to$
$Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r\} Q\{\omega : \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\}$ for all choices of $i's$ and $j's$. Note also that the limit of $\frac{1}{n} \sum_{k=0}^{n-1} Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r, \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\}$ as $n \to \infty$ always exists. [The case $r = m = 1$ was already discussed above and the limit in this case is $q_{i_1 j_1}$].

These facts follow easily from the fact that cylinder sets generate $\mathcal{F}$. If T is ergodic then $q_{ij} = \pi_j$.[Indeed the limit is $\int_A E(I_B|\mathcal{I})/\pi_i$ where $\mathcal{I}$ is the invariant sigma field (which is trivial, in view of ergodicity). This says that the limit is $\frac{Q(B)Q(A)}{\pi_i} = \frac{\pi_i \pi_j}{\pi_i} = \pi_j$. The matrix $M = ((q_{ij}))$ is a stochastic matrix such that $MP = PM = M$ and $M^2 = M$. [Recall that $M = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k$ by definition. To prove the last relation note that $q_{ij}^{(2)} = \sum_{l=0}^{N} q_{il} q_{lj} = \sum_{l=0}^{N} q_{il} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{lj}^{(k)}$

$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \sum_{l=0}^{N} m_{il} p_{lj}^{(k)} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} m_{ij} = m_{ij.}$ [We used the fact that $MP^k = M$ for each k].

We also have $\pi M = \pi \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \pi = \pi$ since $\pi P^k = \pi$ for each k. We now prove a) implies b), i.e. if P is not irreducible then T is not ergodic. Let $A = \{\omega : \omega_0 \in C\}$ where C is a proper closed set of states. Then $0 < Q(A) < 1$. Since C is closed $Q(A \backslash T^{-1}(A)) = 0$. Since T is m.p. this implies $Q(A \Delta T^{-1}(A)) = 0$. This implies that T is not ergodic. Next we prove b) implies d). Let P be irreducible. Let $C = \{j : q_{ij} > 0\}$ with i fixed. This is a closed set. Indeed if $q_{ij} > 0$ and $q_{ik} = 0$ then i cannot lead to $j : 0 = q_{ik} \geq q_{ij} p_{jk}$ (because $M = MP$) which implies $p_{jk} = 0$. By hypothesis $C = S$. This proves b) implies d). Let us prove that d) implies c). Consider the system of equations $\sum_j q_{ij} t_j = t_i, i \in S$. Let $m = max\{t_i : i \in S\}$. If $t_i < m$ then $t_k = \sum_j q_{kj} t_j < m$ for all k (since $t_j \leq m$ for all $j, t_i < m$ and $q_{ij} > 0$ for all i, j). This contradiction shows that $\{t_i\}$ is necessarily constant. We apply this fact to the columns of M. Any column $\{q_{1l}, q_{2l}, ..., q_{Nl}\}$ of M is a solution of above system of equations because $M^2 = M$. Hence each column of M is a constant. This means that q_{ij} is independent of i. Now suppose c) holds. To prove that T is ergodic we have to show that $\frac{1}{n} \sum_{k=0}^{n-1} Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r, \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\} \to$

$Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r\} Q\{\omega : \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\}$ for all choices of $i's$ and $j's$. We have $Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r, \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\} = \pi_{i_1} p_{i_1 i_2} p_{i_2 i_3} ... p_{i_{r-1} i_r} p_{i_r j_1}^{(k-r)} p_{j_1 j_2} p_{j_2 j_3} ... p_{j_{m-1} j_m}$ for $k > r$. Thus $\frac{1}{n} \sum_{k=0}^{n-1} Q\{\omega : \omega_1 = i_1, \omega_2 = i_2, ..., \omega_r = i_r, \omega_{k+1} = j_1, ..., \omega_{k+m} = j_m\} =$

$p_{i_1 i_2} p_{i_2 i_3} ... p_{i_{r-1} i_r} \frac{1}{n} \sum_{k=r+1}^{n-1} p_{i_r j_1}^{(k-r)} p_{j_1 j_2} ... p_{j_{m-1} j_m}$

$\to p_{i_1i_2}p_{i_2i_3}...p_{i_{r-1}i_r}q_{i_rj_1}p_{j_1}p_{j_2}...p_{j_{m-1}j_m}$. Since $q_{i_rj_1}$ does not depend on i_r the limit is $Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r\}Q\{\omega:\omega_{k+1}=j_1,...,\omega_{k+m}=j_m\}$ where we used the fact that $q_{ij}=Q\{\omega:\omega_{k+1}=j\}$. [Since q_{ij} does not depend on i it must be π_i and each projection $\omega\to\omega_k$ has distribution (π_i)]. We have proved that T is ergodic. Thus a) implies b) implies d) implies c) implies a).

Remark: the four conditions above are also equivalent to the following two conditions:
e) $Px=x$ has at most one solution upto a constant factor
f) $xP=x$ has at most one solution upto a constant factor
We leave the details to the reader.

Theorem
With above notations T is mixing iff P is irreducible and aperiodic.
Proof: note that T is mixing iff
$\lim\limits_{n\to\infty}Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r,\omega_{n+1}=j_1,...,\omega_{n+m}=j_m\}$
$=Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r\}Q\{\omega:\omega_{n+1}=j_1,...,\omega_{n+m}=j_m\}$ for all choices of $i's$ and $j's$. If P is irreducible and aperiodic then $\lim\limits_{n\to\infty}p_{ij}^{(n)}=\pi_j$ for all j. If P is irreducible and aperiodic then previous theorem shows that $q_{ij}>0$ for all i,j and that q_{ij} is independent of i.

To prove that T is mixing we have to show that $Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r,\omega_{k+1}=j_1,...,\omega_{k+m}=j_m\}\to$

$Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r\}Q\{\omega:\omega_{k+1}=j_1,...,\omega_{k+m}=j_m\}$ as $k\to\infty$ for all choices of $i's$ and $j's$. We have $Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_k,\omega_{r+1}=j_1,...,\omega_{k+m}=j_m\}=\pi_{i_1}p_{i_1i_2}p_{i_2i_3}...p_{i_{k-1}i_k}p_{i_rj_1}^{(k-r)}p_{j_1}p_{j_2}...p_{j_{m-1}j_m}$ for $k>r$. Thus $Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r,\omega_{k+1}=j_1,...,\omega_{k+m}=j_m\}\to\pi_{i_1}p_{i_1i_2}p_{i_2i_3}...p_{i_{r-1}i_r}\pi_{j_i}p_{j_1}p_{j_2}...p_{j_{m-1}j_m}=Q\{\omega:\omega_1=i_1,\omega_2=i_2,...,\omega_r=i_r\}Q\{\omega:\omega_{k+1}=j_1,...,\omega_{k+m}=j_m\}$. Thus T is mixing. Conversely suppose T is mixing. Then $Q\{\omega:\omega_0=i,\omega_n=j\}\to Q\{\omega:\omega_0=i\}Q\{\omega:\omega_0=j\}$. Thus $\pi_ip_{ij}^{(n)}\to\pi_i\pi_j$ and so $p_{ij}^{(n)}\to\pi_j$. This implies that P is irreducible and aperiodic: $p_{ii}^{(n)}>0$ for all n sufficiently large, so period of each state i is 1; irreducibility is obvious since $p_{ij}^{(n)}>0$ for n sufficiently large.

Skew products
Let $(\Omega,\mathcal{F},P,T)$ be a dynamical system. Suppose for each $\omega\in\Omega$ there is a m.p.t. S_ω on a probability space $(\Omega',\mathcal{F}',P')$ such that the map $(\omega,\omega')\to S_\omega(\omega')$ is measurable from $(\Omega\times\Omega',\mathcal{F}\times\mathcal{F}')$ into $(\Omega',\mathcal{F}')$. The map $\tau:\Omega\times\Omega'\to\Omega\times\Omega'$ defined by $\tau(\omega,\omega')=(T\omega,S_\omega(\omega'))$ is called the *skew product* of T and $\{S_\omega\}_{\omega\in\Omega}$. A special case: let Ω' be a compact topological group and $S_\omega(\omega')=f(\omega)\omega'$ where $f:\Omega\to\Omega'$ is a given measurable function. If E is a Borel set in Ω' then $A=\{(a,b)\in\Omega'\times\Omega':ab\in E\}$ is a Borel set and $\{(\omega,\omega'):S_\omega(\omega')\in E\}=\{(\omega,\omega'):(f(\omega),\omega')\in A\}\in\mathcal{F}\times\mathcal{F}'$ by measurability of f. Hence τ is a skew product if we take P' to be the Haar measure.

Theorem

τ is m.p. on $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', P \times P')$.

Proof: it suffices to show that $(P\times P')(\tau^{-1}(A\times B)) = (P\times P')(A\times B)$ if $A \in \mathcal{F}$ and $B \in \mathcal{F}'$. The left side is $\int I_{\tau^{-1}(A\times B)} d(P\times P') = \int I_{T^{-1}(A)} \int I_{\{\omega' : S_\omega(\omega')\in B\}} dP'(\omega') dP\omega) = \int I_{T^{-1}(A)} P'(B) dP(\omega) = P(T^{-1}(A))P'(B) = (P\times P')(A\times B)$.

Ergodic theorem for flows:

Theorem

Let $\{T_t\}_{t\geq 0}$ be a flow on $(\Omega, \mathcal{F}, P)$. Then $\lim\limits_{\Delta\to\infty} \frac{1}{\Delta}\int\limits_0^\Delta f(T_t(\omega))dt$ exists a.e. and in L^1 for every $f \in L^1$. Also the limit exists in L^p if $f \in L^p$ where $1 \leq p < \infty$.

Proof: we have $\int\limits_0^\Delta f(T_t\omega)dt + \int\limits_\Delta^{[\Delta]+1} f(T_t\omega)dt = \sum\limits_{k=0}^{[\Delta]} \int\limits_k^{(k+1)} f(T_t\omega)dt$. Also $\int\limits_k^{(k+1)} f(T_t\omega))dt = \int\limits_0^1 f(T_{t+k}\omega)dt$

$= \int\limits_0^1 f(T_tT_k\omega)dt = g(T_k\omega)$ where $g(\omega) = \int\limits_0^1 f(T_t\omega)dt$. Hence $\frac{1}{\Delta}\int\limits_0^\Delta f(T_t\omega))dt = \frac{[\Delta]}{\Delta}\frac{1}{[\Delta]}\sum\limits_{k=0}^{[\Delta]} g((T_1)^k\omega) - \frac{1}{\Delta}\int\limits_\Delta^{[\Delta]+1} f(T_t(\omega))dt$.

g is clearly measurable. If $f \in L^p$ then $g \in L^p$. [This follows by Minkowski's inequality in integral form and that fact that each T_t is m.p.]. Further $\frac{1}{\Delta}\int\limits_{[\Delta]}^\Delta f(T_t(\omega))dt \to 0$ uniformly as $\Delta \to \infty$ if f is bounded. Combined with L^p- Ergodic Theorem and Birkhoff ergodic theorem we can draw the following conclusions: if $f \in L^p \cap L^\infty$ then $\frac{1}{\Delta}\int\limits_0^\Delta f(T_t\omega))dt$ converges in $L^p, p = 1, 2$. Noting that

$\left\| \frac{1}{\Delta}\int\limits_0^\Delta f_1(T_t\omega))dt - \frac{1}{\Delta}\int\limits_0^\Delta f_2(T_t\omega))dt \right\|_p \leq \frac{1}{\Delta}\int\limits_0^\Delta \|f_1(T_t\omega)) - f_2(T_t\omega))\|_p dt$

$= \frac{1}{\Delta}\int\limits_0^\Delta \|f_1 - f_2\|_p dt < \epsilon$ whenever $\|f_1 - f_2\|_p < \epsilon$ we can conclude that $\frac{1}{\Delta}\int\limits_0^\Delta f(T_t\omega))dt$ converges in L^p if $f \in L^p$ and $1 \leq p < \infty$. Birkhoff's theorem implies that $\frac{1}{N}\int\limits_{N}^{N+1} f(T_t(\omega))dt \to 0$ a.e. as $N \to \infty$ if $f \in L^1$. [$\frac{a_1+a_2+\ldots+a_n}{n} \to c$ implies $|a_n| < |c| + (n+n-1)\varepsilon$ for n sufficiently large, so $\frac{a_n}{n} \to 0$]. Suppose f is non-negative and integrable. Then $\frac{1}{\Delta}\int\limits_{\Delta}^{[\Delta]+1} f(T_t(\omega))dt \leq \frac{1}{\Delta}\int\limits_{[\Delta]}^{[\Delta]+1} f(T_t(\omega))dt \to 0$ as $\Delta \to \infty$. It follows that $\lim\limits_{\Delta\to\infty} \frac{1}{\Delta}\int\limits_0^{\Delta} f(T_t\omega))dt$ exists a.e. for any non-negative f in L^1. Of course, the same must be true for all f in L^1 since the limit is finite a.e.., In fact the limit is integrable by Fatou's Lemma.

Theorem [*Local ergodic theorem* for flows]

Let $\{T_t\}$ be a flow on $(\Omega, \mathcal{F}, P)$. Then $\frac{1}{2\varepsilon}\int\limits_{-\varepsilon}^{\varepsilon} f(T_s x)ds \to f(x)$ as $\varepsilon \to 0$ almost everywhere for any $f \in L^1(P)$.

Proof: $\int\int\limits_{-1}^{1} |f(T_s x)|\, ds dP(x) = \int\limits_{-1}^{1}\int |f(T_s x)|\, dP(x)ds = \int\limits_{-1}^{1} \|f\|_1\, ds = 2\,\|f\|_1 < \infty$. Hence there is a null set E in $(\Omega, \mathcal{F}, P)$ such that $x \notin E$ implies $\int\limits_{-1}^{1} |f(T_s x)|\, ds < \infty$. For any such x Lebesgue's Theorem implies that $\frac{1}{2\varepsilon}\int\limits_{-\varepsilon}^{\varepsilon} f(T_{s+t}x)ds \to f(T_t x)$ for almost all $t \in (-\frac{1}{2}, \frac{1}{2})$. Fix t such that $\frac{1}{2\varepsilon}\int\limits_{-\varepsilon}^{\varepsilon} f(T_{s+t}x)ds \to f(T_t x)$ for almost all x. [This is possible by Fubini's Theorem]. Now $P\{x : \frac{1}{2\varepsilon}\int\limits_{-\varepsilon}^{\varepsilon} f(T_s x)ds \to f(x)$ as $\varepsilon \to 0\} = P\{x : \frac{1}{2\varepsilon}\int\limits_{-\varepsilon}^{\varepsilon} f(T_{s+t}x)ds \to f(T_t x)$ as $\varepsilon \to 0\}$ because T_t is measure preserving. This completes the proof.

Flow under a function:

Let $(\Omega, \mathcal{F}, P)$ be a probability space and T be an i.m.p.t. on it. Let $f : \Omega \to (0, \infty)$ be measurable with $\int f dP = 1$. Assume that $\sum\limits_{n=0}^{\infty} f(T^n(\omega)) = \infty = \sum\limits_{n=1}^{\infty} f(T^{-n}(\omega))$ for every ω. Let $\Omega' = \Omega \times \mathbb{R}, \bar{\Omega} = \{(\omega, x) : 0 \leq x < f(\omega)\}$

with the sigma field obtained by restricting the product of $\mathcal{F}$ with the Borel sigma field of $\mathbb{R}$ to $\bar{\Omega}$. Note that $(P\times m)(\bar{\Omega}) = \int f dP = 1$. Define $T_t(\omega,x) =$

$$\begin{cases} (\omega, x+t) \text{ if } -x \le t < -x+f(\omega) \\ (T^n\omega, x+t-f(\omega)-...-f(T^{n-1}(\omega)) \text{ if } -x+\sum_{k=0}^{n-1} f(T^k\omega) \le t < -x+\sum_{k=0}^{n} f(T^k\omega) \\ (T^{-n}\omega, x+t+f(T^{-1}\omega)+...+f(T^{-n}(\omega)) \text{ if } -x-\sum_{k=0}^{n} f(T^{-k}\omega) \le t < -x-\sum_{k=0}^{n-1} f(T^{-k}\omega) \end{cases}$$

We call $\{T_t\}$ *the flow built under f on T.*

Theorem [Ambrose, Kakutani]

Any proper measurable flow is isomorphic to a flow built under a function.

For definition of a proper flow and a proof of the theorem we refer the reder to the following article:

"Structure and Continuity of Measurable Flows" by Warrwn Ambrose and Shizuo Kakutani, Duke Math. Jour., Vol. 9, No. 1, 1942. This article has also been reprinted in Selected Papers of Shizuo Kakutani, Volume 2,Birkhauser, 1986.

Unique ergodicity:

Consider a DS $(X,\mathcal{F},P,T)$ where X is a compact Hausdorff space and Then $\mathcal{F}$ is the Borel sigma field.

Let $M = \{\mu \in C(X)^* : \mu$ is a probability measure and $\mu\circ T^{-1} = \mu\}$. This is a compact convex set in $C(X)^*$ with the weak* topology. [Compactness follows from the fact that M is a closed subset of the closed unit ball which is compact by Banach-Alaoglu Theorem]. By Krein-Milman Theorem M is the closed convex hull of its extreme points.

Lemma: $\mu \in M$ makes T ergodic iff μ is an extreme point of M.

Proof of the lemma. suppose $\mu \in M$ makes T ergodic. If possible let $\mu = t\mu_1 + (1-t)\mu_2$ with $0 < t < 1, \mu_i \in M(i=1,2)$ and $\mu_1 \ne \mu_2$. If $T^{-1}(A) = A$ then $\mu(A) = 0$ or $\mu(A^c) = 0$. In the first case $\mu_1(A) = \mu_2(A) = 0$ and in the second case $\mu_1(A^c) = \mu_2(A^c) = 0$ so $\mu_1(A) = \mu_2(A) = 1$. Hence T is ergodic w.r.t. μ_1 and μ_2 also. By Birkhoff's Theorem we see that if $E \in \mathcal{F}$ then$\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^kx) \to \mu(E)$ a.e. $[\mu]$ and in $L^1(\mu)$, $\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^kx) \to \mu_1(E)$ a.e. $[\mu_1]$ and in $L^1(\mu)$, and $\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^kx) \to \mu_2(E)$ a.e. $[\mu_2]$ and in $L^1(\mu)$. Thus $\int\left|\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^kx) \to \mu(E)\right| d\mu(x) \to 0$ and this implies $\int\left|\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^kx) \to \mu(E)\right| d\mu_i(x) \to$

0 for $i = 1, 2$. But then there is a subsequence of $\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^k x)$ converging to $\mu(E)$ $a.e.[\mu_i]$ and so $\mu(E) = \mu_1(E) = \mu_2(E)$. This is true for any measurable set E. This contradiction shows that μ is an extreme point of M. Conversely suppose T is invariant but not ergodic w.r.t. μ. There is an invariant set A such that $0 < \mu(A) < 1$. Let $\mu_1(E) = \frac{\mu(E\cap A)}{\mu(A)}$ and $\mu_2(E) = \frac{\mu(E\cap A^c)}{\mu(A^c)}$. Then μ_1 and μ_2 are both probability measures. $\mu_1(T^{-1}(E)) = \frac{\mu(T^{-1}(E)\cap A)}{\mu(A)} = \frac{\mu(T^{-1}(E\cap A))}{\mu(A)} = \frac{\mu((E\cap A))}{\mu(A)} = \mu_1(E)$ so $\mu_1 \in M$. Similarly $\mu_2 \in M$. We have $\mu(E) = \mu(E\cap A) + \mu(E \cap A^c) = \mu(A)\mu_1(E) + \mu(A^c)\mu_2(E)$ for all E showing that μ is not an extreme point of M. [$\mu_1 = \mu_2$ would imply $\mu(A \cap A) = \mu(A)\mu_1(A) = \mu(A)\mu_2(A) = 0$, a contradiction].

We now prove uniform convergence in Birkhoff's theorem under certain conditions.

Definition: T is *uniquely ergodic* if there is a unique probability measure under which it is invariant.

Justification of this comes from the fact that when M is a singleton, $\{\mu\}$, then μ is necessarily an extreme point and so T is ergodic w.r.t. μ.

Theorem [Weyl]

Suppose the DS $(\Omega, \mathcal{F}, P, T)$ is uniquely ergodic where Ω is a compact metric space and $\mathcal{F}$ is the Borel sigma field. If T and f are continuous then $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges uniformly. Also $\frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k x_l} \to P$ in weak* topology.

Proof: we know that $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \int f dP$ a.e.. Suppose the convergence is not uniform. Then we can find $\delta > 0, n_1 < n_2 < ...$ and $\{x_l\} \subset \Omega$ such that $\left|\frac{1}{n_l}\sum_{k=0}^{n_l-1} f(T^k x_l) - \int f dP\right| \geq \delta\ \forall l$. Let $\mu_l = \frac{1}{n_l}\sum_{k=0}^{n_l-1} \delta_{T^k x_l}$. This sequence of probability measures has a subsequence converging in weak* topology to a probability measure. [Recall that $C(X)$ is separable and hence the unit ball of its dual is weak* metrizable]. This limiting measure belongs to M (by a direct verification which is left to the reader) and the hypothesis implies that it must be P. This clearly contradicts the fact that $\left|\frac{1}{n_l}\sum_{k=0}^{n_l-1} f(T^k x_l) - \int f dP\right| \geq \delta\ \forall l$. The second part follows easily from this proof.

Theorem [A converse of Weyl's Theorem].

Suppose $(\Omega, \mathcal{F}, P, T)$ is a DS where Ω is a compact metric space and $\mathcal{F}$ is the Borel sigma field. If T is continuous and ergodic and $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges uniformly for every $f \in C(\Omega)$ then T is uniquely ergodic. The same conclusion holds if we only know that some subsequence of $\{\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\}$ converges uniformly for each $f \in C(\Omega)$.

Proof: suppose Q is a probability measure on $(\Omega, \mathcal{F})$ such that $Q \circ T^{-1} = T$. Suppose $\frac{1}{n_j}\sum_{k=0}^{n_j-1} f(T^k x) \to g$ uniformly for some $\{n_j\} \uparrow \infty$. Then$\int f dQ = \int \frac{1}{n_j}\sum_{k=0}^{n_j-1} f(T^k x) dQ$ (because $Q \circ T^{-1} = T$). Letting $j \to \infty$ we get $\int f dQ = \int g dQ$. However ergodicity of T w.r.t. P shows that $g = \int f dP$ (a.e., hence evreywhere, be continuity). It follows that $\int f dQ = \int g dQ = \int\int f dP dQ = \int f dP$. Since this holds for all $f \in C(\Omega)$ we get $P = Q$, as stated.

Remark: suppose $(\Omega, \mathcal{F}, P, T)$ is a DS where Ω is a compact metric space and $\mathcal{F}$ is the Borel sigma field. If T is continuous and $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges uniformly to a constant for every $f \in C(\Omega)$ then T is uniquely ergodic. The same conclusion holds if we only know that some subsequence of $\{\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\}$ converges uniformly to a constant for each $f \in C(\Omega)$. The proof is same as the one above.

Example

Let α be an irrational number, $\Omega = [0,1], \mathcal{F} =$ borel sigma field, $P =$ Lebesgue measure and $T(x) = x + \alpha(\text{mod}(1))$. Suppose μ is a probability measure on Ω which makes T invariant. Then $\int e^{i2\pi nx} d\mu(x) = \int e^{i2\pi nx} d(\mu \circ T^{-1})(x) = \int e^{i2\pi n(y+\alpha)} d\mu(y) = e^{i2\pi n\alpha} \int e^{i2\pi ny} d\mu(y)$. Since $e^{i2\pi n\alpha} = 1$ iff $n = 0$ it follows that $\int e^{i2\pi nx} d\mu(x) = 0$ for all $n \neq 0$. [For $n = 0$ the integral is 1]. It follows that $\mu = P$. Thus T is uniquely ergodic. Hence $\frac{1}{n}\sum_{k=0}^{n-1} f(x + k\alpha)$ converges uniformly to $\int_0^1 f(x)dx$ for any continuous periodic function f with period 1.

Exercise

Show that above fact remains true if f is periodic and Riemann integrable (not necessarily continuous).

The fact that this is true for all intervals contained in $[0,1)$ is called Weyl's Equi-distribution Theorem.

Hints: we can find continuous functions g and h such that $g \leq f \leq h$ and $\int\limits_0^1 (h-g)$ is as small as necessary. (To see this consider step functions approximating f and modify them suitably on small intervals to get continuous functions).

Existence of invariant measures

Theorem

Suppose $T : \Omega \to \Omega$ be continuous where Ω is a compact metric space. Let $\mathcal{F}$ be the Borel sigma field. Then there exists a probability measure Q on $(\Omega, \mathcal{F})$ such that T is m.p. and ergodic w.r.t. Q.

Proof: let P be any probability measure on $(\Omega, \mathcal{F})$ and $\mu_n = \frac{1}{n} \sum\limits_{k=0}^{n-1} P \circ T^{-k}$. This sequence of probability measures has a subsequence, say $\{\mu_{n_j}\}$ converging in weak* topology of $C(X)^*$ to a probability measure μ. We claim that $\mu \circ T^{-1} = \mu$. It suffices to show that $\int f(T(x)) d\mu(x) = \int f d\mu$ for every continuous function f. Now $\int f(T(x)) d\mu(x) = \lim \frac{1}{n_j} \sum\limits_{k=0}^{n_j-1} \int f(T^{k+1}(x)) dP(x)$

$= \lim \frac{1}{n_j} \sum\limits_{k=1}^{n_j} \int f(T^k(x)) dP(x) = \lim \frac{1}{n_j} \{ \sum\limits_{k=0}^{n_j-1} \int f(T^k(x)) dP(x) - \int f dP +$
$\int f(T^{n_j}(x)) dP(x) \} = \lim \frac{1}{n_j} \{ \sum\limits_{k=0}^{n_j-1} \int f(T^k(x)) dP(x)$

$= \int f d\mu$. Thus μ makes T m.p.. Now the collection of all invariant probability measures is a non-empty compact convex set. By Krein - Milman Theorem this set has an extreme point Q. As seen earlier this Q makes T ergodic.

Remark: in above theorem can we find Q such that $Q << P$ where P is a given probability measure?. The answer is no! Let $\Omega = S^1, m$ denote normalized arc length measure and let P be any probability measure such that $P << m$ but m is not absolutely continuous w.r.t. P. [For instance P could be m restricted to the upper half of the circle divided by 2]. Let $Tz = az$ where $a \in S^1$ is not a root of unity. Suppose there is a probability measure Q such that $Q << P$ and $Q \circ T^{-1} = Q$ where $Tz = az$ and a is not a root of unity.. Then there exists f such that $Q(E) = \int\limits_E f dm$. Now $\int\limits_E f \circ T dm = \int\limits_{T(E)} f dm$ (because T is i.m.p. w.r.t. m)

$= Q(T(E)) = Q(E) = \int_E f dm$ for all Borel sets E so f is an invariant function on (S^1, m). Since T is ergodic we get $f = c\ [m]$ for some constant c. But then $Q(E) = \int_E f dm = cm(E)$ which implies that $m << Q << P$ contradicting the fact that m is not absolutely continuous w.r.t. P.

Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be bijective and bi-measurable. Suppose $P \circ T << P$ and $P \circ T^{-1} << P$. Then the following are equivalent:

a) there exists an invariant probability measure for T which is equivalent to Q

b) $\{\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(\omega))\}$ converges a.e. $[P]$ for every $f \in L^1$.

Proof: a) implies b) is obvious from Birkhoff's Ergodic Theorem. For b) implies a) let $Q(E) = \lim \frac{1}{n}\sum_{k=0}^{n-1} P(T^{-k}(E))$. The limit exists since $\frac{1}{n}\sum_{k=0}^{n-1} I_E(T^k(\omega))$ exists a.e. $[P]$, hence also in $L^1(P)$ be DCT. [That Q is indeed a probability measure follows by Vitali-Hahn-Saks-Theorem]. By hypothesis $Q << P$. Also $Q \circ T^{-1} = Q$. Now suppose $P(A) > 0$. Then $P\{\bigcup_n T^{-n}A\} \geq P(A) > 0$. Since $\bigcup_n T^{-n}A$ is an invariant set the definition of Q shows $Q(\bigcup_n T^{-n}A) = P(\bigcup_n T^{-n}A) > 0$. If $Q(A) = 0$ then $Q(T^{-n}(A)) = 0$ for all n implying that $Q(\bigcup_n T^{-n}A) = 0$ a contradiction. We have proved that $P(A) > 0$ implies $Q(A) > 0$ and the proof is complete.

Theorem [DOWKER]

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be a bijective map such that T and T^{-1} are measurable. Suppose $\liminf P(T^{-n}(A)) > 0$ whenever $P(A) > 0$ and $P \circ T^{-1} << P$. Then there exists a probability measure μ on $(\Omega, \mathcal{F})$ such that $\mu \circ T^{-1} = \mu$ and $\mu\tilde{}P$ (in the sense $\mu << P << \mu$). Conversely if there exists a probability measure μ on $(\Omega, \mathcal{F})$ such that $\mu \circ T^{-1} = \mu$ and $\mu\tilde{}P$ then $\liminf P(T^{-n}(A)) > 0$ whenever $P(A) > 0$.

We use Banach limits in the proof. [Let M be the set of all bounded sequences $\{a_n\}$ of real numbers such that $\lim \frac{a_1+a_2+\ldots+a_n}{n}$ exists considered as a subspace of the Banach space l^∞ of all bounded sequences of real numbers with the supremum norm. By Hahn Banach Theorem there exists a continuous linear map $L : l^\infty \to \mathbb{R}$ such that $\|L\| = 1$ and $L(\{a_n\}) = \lim \frac{a_1+a_2+...+a_n}{n}$ whenever $\{a_n\} \in M$. We claim that L has the following properties: $L(\{c\}) = c$ for any constant sequence $\{c\}, L(\{a_n\}) = c$ whenever $a_n \to c$, $\liminf a_n \leq L(\{a_n\}) \leq \limsup a_n$ for all $\{a_n\} \in l^\infty$, $L(\{a_{n+1} - a_n\}) = 0$ for all $\{a_n\} \in l^\infty$ and $L(\{a_n\}) \geq 0$ if $a_n \geq 0$ for all n. Proofs of these facts are easy: $\{a_{n+1} - a_n\} \in M$ and the cesaro limit of $\{a_{n+1} - a_n\}$ is 0. This gives $L(\{a_{n+1} - a_n\}) = 0$. If $a_n \geq 0$ for all n and $C = \sup\{a_n : n \in \mathbb{N}\}$ then $L(\{C - a_n\}) \leq \|L\| C = C$ so $C - L(\{a_n\}) \leq C$ and $L(\{a_n\}) \geq 0$. If $\epsilon > 0$ then there exists n_0 such that $a_n \geq \liminf a_n - \epsilon$ for all $n \geq n_0$ and so $L(\{a_{n_0+n} - \liminf a_n + \epsilon\}) \geq 0$. This implies $L(\{a_{n_0+n}\}) \geq \liminf a_n - \epsilon$. However $L(\{a_{n_0+n}\}) = L(\{a_{n_0+n-1}\}) = ... = L(\{a_n\})$ and ϵ is arbitrary so $L(\{a_n\}) \geq \liminf a_n$. Replacing $\{a_n\}$ by $\{-a_n\}$ we get $L(\{a_n\}) \leq \limsup a_n$].

Now let $Q(A) = L(\{P(T^{-n}(A))\}$. Q is a finitely additive non-negative set function and $Q(\Omega) = 1$. Also $P(A) = 0 \Rightarrow P(T^{-1}(A)) = 0 \Rightarrow P(T^{-2}(A)) = 0 \Rightarrow ...$ so $Q(A) = 0$. Note that $Q(T^{-1}(A)) = L(\{P(T^{-n}(T^{-1}A))\} = L(\{P(T^{-(n+1)}(A))\} = L(\{P(T^{-n}(A))\} = Q(A)$. We claim that $P(A) = 0$ iff $Q(A) = 0$. By hypothesis $\liminf P(T^{-n}(A)) > 0$ whenever $P(A) > 0$. Thus $P(A) > 0$ implies $Q(A) = L(\{P(T^{-n}(A))\} \geq \liminf P(T^{-n}(A)) > 0$. The claim follows.

We now construct a countably additive measure μ with the same properties.

Let $\mu(A) = \inf\{\sum Q(A_n) : A_n \in \mathcal{F}\ \forall n, A \subset \bigcup\limits_n A_n\}$. We first observe the following: if $\mu(A) = 0$ adn $\varepsilon > 0$ we can choose *disjoint* $A_1, A_2, ...$sets such that $A \subset \bigcup\limits_n A_n$ and $\sum Q(A_n) < \varepsilon$. [Replace $\{A_n\}$ by $\{A_1, A_2 \backslash A_1, A_3 \backslash A_1 \cup A_2, ...\}$]. Suppose E is the disjoint union of the sequence $\{E_n\}$ in $\mathcal{F}$. Given $\epsilon > 0$ and $j \geq 1$ there exists a sequene $\{A_{,j,n}\} \subset \mathcal{F}$ such that $\mu(E_j) + \frac{\epsilon}{2^j} > \sum\limits_n Q(A_{j,n})$ and $E_j \subset \bigcup\limits_n A_{j,n}$. Since $E \subset \bigcup\limits_{j,n} A_{j,n}$ we have $\mu(E) \leq \sum\limits_{j,n} Q(A_{j,n}) \geq \sum\limits_j \{\mu(E_j) + \frac{\epsilon}{2^j}\} = \sum\limits_j \mu(E_j) + \epsilon$. Thus $\mu(E) \leq \sum\limits_j \mu(E_j)$. Now $\mu(E) + \epsilon > \sum Q(A_n)$ for some sets $A_n \in \mathcal{F}$ with $E \subset \bigcup\limits_n A_n$. Since $E_j \subset \bigcup\limits_n (A_n \cap E_j)$ we have $\sum\limits_{j=1}^N \mu(E_j) \leq \sum\limits_{j=1}^N \sum\limits_n Q(A_n \cap E_j) = \sum\limits_n \sum\limits_{j=1}^N Q(A_n \cap E_j) = \sum\limits_n Q(A_n \cap (\bigcup\limits_{j=1}^N E_j)) \leq \sum\limits_n Q(A_n) < \mu(E) + \epsilon$. Since ϵ and N are arbitrary we get $\sum\limits_{j=1}^\infty \mu(E_j) \leq \mu(E)$. Hence μ is a measure. Note that $\mu(E) \leq Q(E) + Q(\emptyset) + Q(\emptyset) + ... = Q(E)$ for all E. Finally we show that $\mu \circ T^{-1} = \mu$ and $\mu\tilde{\ }P$. We have $\mu(T^{-1}(A)) = \inf\{\sum Q(A_n) : A_n \in \mathcal{F}\ \forall n, T^{-1}(A) \subset \bigcup\limits_n A_n\} = \inf\{\sum Q(T^{-1}A_n) : A_n \in \mathcal{F}$ $\forall n, T^{-1}(A) \subset \bigcup\limits_n T^{-1}(A_n)\}$

$= \inf\{\sum Q(A_n) : A_n \in \mathcal{F}\ \forall n, A \subset \bigcup_n A_n\} = \mu(A)$. Clearly $\mu(A) \leq Q(A) = 0$ whenever $P(A) = 0$. Now suppose $P(A) > 0$. If possible let $\mu(A) = 0$. There exist disjoint measurable sets $A_{j,n}, (j, n \geq 1)$ such that $A \subset \bigcup_j A_{j,n}$ and $\sum_{j=1}^{\infty} Q(A_{j,n}) < \frac{\epsilon}{2^n}$ for all n. Without loss of generalty we may assume that $A_{j,n}, j = 1, 2, ...$ are disjoint. Choose k_n such that $\sum_{j=k_n}^{\infty} P(A_{j,n}) < \frac{P(A)}{3^n}$. [Possible because $\sum_{j=1}^{\infty} P(A_{j,n}) \leq 1$ and $P(A) > 0$]. Let $C_n = \bigcup_{j=1}^{k_n - 1} A_{j,n}$ and $D_n = \bigcup_{j=k_n}^{\infty} A_{j,n}$. Then $A \backslash \bigcup_n D_n = \bigcap_n (A \backslash D_n) \subset \bigcap_n C_n = C$ (say) so $A \subset \bigcup_n D_n \cup C$ and $P(A) \leq \sum_n P(D_n) + P(C)$ However since $\sum_{j=k_n}^{\infty} P(A_{j,n}) < \frac{P(A)}{3^n}$ the definition of D_n gives $P(D_n) < \frac{P(A)}{3^n}$ so $P(A) \leq P(A) \sum_n \frac{1}{3^n} + P(C) = \frac{1}{2} P(A) + P(C)$ and $P(C) \geq \frac{1}{2} P(A) > 0$. This implies $Q(C) > 0$. However $Q(C) \leq \sum_{j=1}^{k_n - 1} Q(A_{j,n}) \leq \sum_{j=1}^{\infty} Q(A_{j,n}) < \frac{\epsilon}{2^n}$ and n is arbitrary so $Q(C) = 0$. This contradiction completes the proof.

The converse part is easy: since $Q(E) \to 0$ as $P(E) \to 0$ we see that $\liminf P(T^{-n}(A)) = 0$ implies $\liminf Q(T^{-n}(A)) = 0$ which gives $Q(A) = 0$, hence $P(A) = 0$.

Remark: the condition $P \circ T^{-1} << P$ can be replaced by the weaker condition that $\lim P(T^{-n}(A))$ exists and equals 0 whenever $P(A) = 0$. Indeed $Q(A) = L(\{P(T^{-n}(A))\} = 0$ in this case and the proof above works.

Converse part is easy: $P(A) = 0 \Rightarrow \mu(A) = 0 \Rightarrow \mu(T^{-1}A) = 0 \Rightarrow P(T^{-1}A) = 0$ so $P \circ T^{-1} << P$. If $P(A) > 0$ and $\liminf P(T^{-n}(A)) = 0$ then there exists $n_j \uparrow \infty$ such that $P(T^{-n_j}(A)) \to 0$ which implies $\mu(T^{-n_j}(A)) \to 0$. This is a contradiction because $\mu(T^{-n_j}(A)) = \mu(A) > 0$ for all j.

Remark: Dowker has proved that the condition $\liminf P(T^{-n}(A)) > 0$ whenever $P(A) > 0$ can be replaced by the condition $\limsup P(T^{-n}(A)) > 0$ whenever $P(A) > 0$.

Theorem [HAJIAN, KAKUTANI]

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be bijective and bimeasurable. Then T has an invariant measure equivalent to P iff the measures $P \circ T^{-n}, n = 1, 2, ...$ are uniformly absolutely continuous w.r.t. P

We do not prove this theorem here. For a proof see Theorem 3.18 of Introduction to Ergodic Theory by N. A. Friedman. For more information on this topic see: "The Problem of Finite Invariant Measures" by Daniel Glasscock.

Theorem [*Ergodic Decomposition*]

Let Ω be a compact metric space, $T : \Omega \to \Omega$ continuous and P a Borel probability measure such that $P \circ T^{-1} = P$. Then there exist Borel probability measures $P_\omega, \omega \in \Omega$ such that

1) $\int f dP = \int (\int f dP_\omega) dP(\omega)$ for all $f \in L^1(P)$

2) $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) \to \int f dP_\omega$ a.e. $[P]$ for all $f \in L^1(P)$

3) for each $\omega \in \Omega$ the map T is ergodic w.r.t. P_ω for each $\omega \in \Omega$.

Thus any continuous m.p. transformation is a 'mixture' of ergodic transformations.

We shall not prove this theorem here.

ENTROPY

Let $(\Omega, \mathcal{F}, P, T)$ be a dynamical system. If $\mathcal{A}$ is a finite sub-sigma field of $\mathcal{F}$ then there is a finite partition of Ω by sets in $\mathcal{F}$ such that the sets in the partition generate $\mathcal{A}$. In view of this we use symbols like $\mathcal{A}, \mathcal{B}$ etc for finite sigma fields as well as finite partitions.

Definition: $h(\mathcal{A}) = -\sum_{i=1}^{k} P(A_i) \log P(A_i)$ where $\mathcal{A}$ is the partition $\{A_1, A_2, ..., A_k\}$. [We use the convention $0 \log 0 = 0$]. This is called the *entropy* of $\mathcal{A}$. It is measure of the information contained in the partition.

If two partitions generate the same finite sigma field then they differ only be a permutation. Hence entropy of a partition depends only on the field generated by it.

We can also call $-\sum_{i=1}^{k} p_i \log p_i$ the entropy of the probability vector $\{p_1, p_2, ..., p_k\}$.

Theorem

1) $h(\mathcal{A}) \geq 0$

2) $h(\mathcal{A}) = 0$ iff $\mathcal{A}$ is trivial in the sense every set in it has probability 0 or 1.

3) $h(\mathcal{A}) \leq -\sum_{i=1}^{k} \frac{1}{k} \log \frac{1}{k} = \log k$.

Proof: 1) and 2) are obvious. For 3) we use Jensen's inequality. Since logarithm is concave we have $-\sum_{i=1}^{k} P(A_i) \log P(A_i) = \sum_{i=1}^{k} P(A_i) \log \frac{1}{P(A_i)} \leq \log \sum_{i=1}^{k} P(A_i) \frac{1}{P(A_i)} = \log k$.

Exercise

Trying to define a continuous version of entropy define $h(f) = -\displaystyle\int f \log f$ for a density function f. Show that $h(f)$ may be 0 or ∞! If g is another density function show that $-\displaystyle\int f \log \frac{g}{f} \geq 0$.

[$I_{(0,1)}$ is one counterexample. $\displaystyle\sum_{j=1}^{\infty} a_j I_{(j,.j+1)}$ where $a_j = \frac{c}{n[\log n]^2}$ gives the other counterexample. A suitable application of Jensen's inequality gives the last part].

Let $h(\mathcal{A}\ |\mathcal{B}) = -\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m} P(A_i \cap B_j) \log P(A_i|B_j)$ if $\mathcal{A} = \{A_1, A_2, ..., A_k\}$ and $\mathcal{B} = \{B_1, B_2, ..., B_{mk}\}$. In this definition we ignore the terms with $P(B_j) = 0$. Equivalently we define $P(A_i \cap B_j) \log P(A_i|B_j) = 0$ when $P(B_j) = 0$.

Let us first observe that $h(\mathcal{A}\ |\mathcal{B}) \geq 0$ since $P(A_i|B_j) \leq 1$.

Theorem

1) $h(\mathcal{A} \bigvee \mathcal{B}|\mathcal{C}) = h(\mathcal{A}|\mathcal{C}) + h(\mathcal{B}|\mathcal{A} \bigvee \mathcal{C})$
2) $h(\mathcal{A} \bigvee \mathcal{B}) = h(\mathcal{A}) + h(\mathcal{B}|\mathcal{A})$
3) $h(\mathcal{A}|\mathcal{C}) \leq h(\mathcal{B}|\mathcal{C})$ if $\mathcal{A} \subset \mathcal{B}$
4) $h(\mathcal{A}) \leq h(\mathcal{B})$ if $\mathcal{A} \subset \mathcal{B}$
5) $h(\mathcal{A}|\mathcal{B}) \leq h(\mathcal{A}|\mathcal{C})$ if $\mathcal{C} \subset \mathcal{B}$
6) $h(\mathcal{A}|\mathcal{B}) \leq h(\mathcal{A})$
7) $h(\mathcal{A} \bigvee \mathcal{B}|\mathcal{C}) \leq h(\mathcal{A}|\mathcal{C}) + h(\mathcal{B}|\mathcal{C})$
8) $h(\mathcal{A} \bigvee \mathcal{B}) \leq h(\mathcal{A}) + h(\mathcal{B})$

9) $h(T^{-1}\mathcal{A}|T^{-1}\mathcal{B}) = h(\mathcal{A}|\mathcal{B})$
10) $h(T^{-1}(\mathcal{A})) = h(\mathcal{A})$

Proof: note that $h(\mathcal{A}\ |\mathcal{B}) = -\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m} P(A_i \cap B_j) \log P(A_i|B_j) = h(\mathcal{A}\ )$ when $\mathcal{B}$ is trivial. Thus all the even numbered properties follow from the previous ones.

Proof of 1): we have $h(\mathcal{A} \bigvee \mathcal{B}|\mathcal{C}) = -\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{r=1}^{p} P((A_i \cap B_r \cap C_j) \log P(A_i \cap B_r|C_j) = -\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{r=1}^{p} P((A_i \cap B_r \cap C_j) \log[P(A_i|B_r \cap C_j)P(B_r \cap C_j|C_j)]$

$$= -\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{r=1}^{p} P((A_i \cap B_r \cap C_j) \log P(A_i|B_r \cap C_j)$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{r=1}^{p} P((A_i \cap B_r \cap C_j) \log P(B_r \cap C_j|C_j)$$

$= h(\mathcal{A}|\mathcal{B}\bigvee\mathcal{C}) + h(\mathcal{B}|\mathcal{C})$. We get 1) by switching $\mathcal{A}$ and $\mathcal{B}$

If $\mathcal{A}\subset\mathcal{B}$ then $\mathcal{A}\bigvee\mathcal{B}=\mathcal{B}$ and 1) gives $h(\mathcal{B}|\mathcal{C}) = h(\mathcal{A}|\mathcal{C}) + h(\mathcal{B}|\mathcal{A}\bigvee\mathcal{C}) \geq h(\mathcal{A}|\mathcal{C})$. We have proved 3).

To prove 5) we use the fact that the function $\phi(t) = -t\log t$ is concave. Hence $\phi(\sum_{j=1}^{m} P(A_i|C_j)P(C_j|B_r)) \geq \sum_{j=1}^{m}\phi(P(A_i|C_j))P(C_j|B_r)$ for all i and r. If $\mathcal{B}\subset\mathcal{C}$ then B_r is a union of $C_j's$' Hence $\sum_{j=1}^{m} P(A_i|C_j)P(C_j|B_r) = \sum P(A_i\cap C_j)/P(B_r)$ where the sum is over those j for which $C_j\subset B_r$. Thus $\sum_{j=1}^{m} P(A_i|C_j)P(C_j|B_r) = P(A_i|B_r)$ and we get $\phi(P(A_i|B_r)) \geq \sum_{j=1}^{m}\phi(P(A_i|C_j))P(C_j|B_r)$. Multiplying by $P(B_r)$ and summing over r we get $-\sum_{r=1}^{p} P(A_i\cap B_r)\log P(A_i|B_r) \geq$
$-\sum_{r=1}^{p}\sum_{j=1}^{m}\{P(A_i|C_j)\log P(A_i|C_j)\}P(C_j\cap B_r)\}$

$= -\sum_{j=1}^{m}\{P(A_i|C_j)\log P(A_i|C_j)\}P(C_j) = -\sum_{j=1}^{m} P(A_i\cap C_j)\log P(A_i|C_j)$. This says $h(\mathcal{A}\ |\mathcal{B}) \geq h(\mathcal{A}\ |\mathcal{C})$ and the proof of 5) is complete. To prove 7) note that $h(\mathcal{A}\bigvee\mathcal{B}|\mathcal{C}) = h(\mathcal{A}|\mathcal{C}) + h(\mathcal{B}|\mathcal{A}\bigvee\mathcal{C})$ (by 1)) and the second term on the right does not exceed $h(\mathcal{B}|\mathcal{C})$ by what we just proved, so $h(\mathcal{A}\bigvee\mathcal{B}|\mathcal{C}) \leq h(\mathcal{A}|\mathcal{C}) + h(\mathcal{B}|\mathcal{C})$. 9) is trivial.

Notation: $\bigvee_{i=0}^{n}\mathcal{A}_i$ denotes the sigma field generated by $\bigcup_{i=0}^{n}\mathcal{A}_i$.

Now $h(\mathcal{A}|\ \bigvee_{i=1}^{n}T^{-i}\mathcal{A}) = h(\bigvee_{i=0}^{n}T^{-i}\mathcal{A}) - h(\bigvee_{i=1}^{n}T^{-i}\mathcal{A})$ by 2) with $\mathcal{A}$ replaced by $\bigvee_{i=1}^{n}T^{-i}\mathcal{A}$ and $\mathcal{B}$ replaced by $\mathcal{A}$. Hence $h(\mathcal{A}|\ \bigvee_{i=1}^{n}T^{-i}\mathcal{A}) = h(\bigvee_{i=0}^{n}T^{-i}\mathcal{A}) - h(\bigvee_{i=0}^{n-1}T^{-i}\mathcal{A})$ (by 10)). Summing over n we get $\sum_{n=1}^{N} h(\mathcal{A}|\ \bigvee_{i=1}^{n}T^{-i}\mathcal{A}) = h(\bigvee_{i=0}^{N}T^{-i}\mathcal{A}) - h(\mathcal{A})$. This gives $\frac{1}{N}h(\bigvee_{i=0}^{N}T^{-i}\mathcal{A}) = \frac{1}{N}h(\mathcal{A}) + \frac{1}{N}\sum_{n=1}^{N} h(\mathcal{A}|\ \bigvee_{i=1}^{n}T^{-i}\mathcal{A})$. The sequence $\{h(\mathcal{A}|\bigvee_{i=1}^{n}T^{-i}\mathcal{A})\}$ is decreasing (by 5)) and non-negative, hence convergent. Therefore $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} h(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A})$ exists (and is finite).This proves that $\lim_{N\to\infty} \frac{1}{N} h(\bigvee_{i=0}^{N} T^{-i}\mathcal{A})$ exists and equals $\lim_{n\to\infty} h(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A})$.

Definition
$h(\mathcal{A}\ ,T) = \lim_{n\to\infty} \frac{1}{n} h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A}) = \lim_{n\to\infty} h(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A})$.

This is called the *entropy of* $\mathcal{A}$ *relative to* T.

Caution: it is not true that $h(\mathcal{A}\ ,T) = \lim_{n\to\infty} h(\mathcal{A}|\bigvee_{i=0}^{n} T^{-i}\mathcal{A})$. In fact the right side is 0 because $h(\mathcal{A}\ |\mathcal{B}) = 0$ if $\mathcal{A} \subset \mathcal{B}$. However in the equation $h(\mathcal{A}\ ,T) = \lim_{n\to\infty} \frac{1}{n} h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A})$ we can replace $\bigvee_{i=0}^{n} T^{-i}\mathcal{A}$ by $\bigvee_{i=1}^{n} T^{-i}\mathcal{A}$. This follows from the fact that $h(\bigvee_{i=1}^{n} T^{-i}\mathcal{A}) \leq h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A}) \leq h(\mathcal{A})+h(\bigvee_{i=1}^{n} T^{-i}\mathcal{A})$.

Theorem.

1) The sequences $\{\frac{1}{n} h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A})\}$ and $\{h(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A})\}$ (whose limits appear in above definition) are both decreasing.

2) $h(\mathcal{A}\ ,T) = \lim_{n\to\infty} h(T^{-n}\mathcal{A}|\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})$

3) It T is i.m.p. then $h(\mathcal{A}\ ,T) = \lim_{n\to\infty} h(\mathcal{A}|\bigvee_{i=1}^{n} T^{i}\mathcal{A})$.

4) $h(\mathcal{A}\ ,T) \leq h(\mathcal{B},T)$ if $\mathcal{A} \subset \mathcal{B}$

5) $h(\bigvee_{i=n}^{m} T^{-i}\mathcal{A},T) = h(\mathcal{A}\ ,T)$ if $m \geq n \geq 0$ and the same equation holds without the condition $n \geq 0$ when T is i.m.p.

6) $h(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A},T^{k}) = kh(\mathcal{A}\ ,T)$ for $k = 1, 2, ...$

7) $h(\mathcal{A}\ ,T) \leq h(\mathcal{B},T) + h(\mathcal{A}|\mathcal{B})$

Proof: we have already seen that the second sequence is decreasing. The first sequence is decreasing because $\frac{1}{N} h(\bigvee_{i=0}^{N} T^{-i}\mathcal{A}) = \frac{1}{N} h(\mathcal{A}) + \frac{1}{N}\sum_{n=1}^{N} h(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A})$. To prove that $h(\mathcal{A}\ ,T) = \lim_{n\to\infty} h(T^{-n}\mathcal{A}|\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})$ we note that

$h(T^{-k}\mathcal{A}|\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) = h(\bigvee_{i=0}^{k} T^{-i}\mathcal{A}) - h(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})$ (by 2) of previous theorem with $\mathcal{A}$ replaced by $\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}$ and $\mathcal{B}$ replaced by $T^{-k}\mathcal{A}$). Hence $\frac{1}{N}\sum_{k=1}^{N} h(T^{-k}\mathcal{A}|$ $\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) = \frac{1}{N}h(\bigvee_{i=0}^{N} T^{-i}\mathcal{A}) - \frac{1}{N}h(\mathcal{A})$. Also $h(T^{-k}\mathcal{A}|\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})$

$= h(T^{-(k+1)}\mathcal{A}|T^{-1}(\bigvee_{i=1}^{k} T^{-i}\mathcal{A})) = h(T^{-(k+1)}\mathcal{A}|\bigvee_{i=1}^{k} T^{-i}\mathcal{A}) \geq h(T^{-(k+1)}\mathcal{A}|$ $\bigvee_{i=0}^{k} T^{-i}\mathcal{A})$ so the sequence $\{h(T^{-k}\mathcal{A}|\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})\}$ is decreasing. It follows that the limit of this sequence is also the limit of $\frac{1}{N}\sum_{k=1}^{N} h(T^{-k}\mathcal{A}|\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) =$ $\frac{1}{N}h(\bigvee_{i=0}^{N} T^{-i}\mathcal{A}) - \frac{1}{N}h(\mathcal{A})$ which is $h(\mathcal{A}\ ,T)$.

We have proved 2) and 3) follows immediately. 4) is easy: $\mathcal{A} \subset \mathcal{B}$ implies $\bigvee_{i=0}^{n} T^{-i}\mathcal{A} \subset \bigvee_{i=0}^{n} T^{-i}\mathcal{B}$ and hence $h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A}) \subset h(\bigvee_{i=0}^{n} T^{-i}\mathcal{B})$. Dividing by $n+1$ and letting $n \to \infty$ we get $h(\mathcal{A}\ ,T) \leq h(\mathcal{B},T)$.

We have $\bigvee_{i=0}^{N-1} T^{-i}\bigvee_{j=n}^{m} T^{-j}\mathcal{A} = \mathcal{T}^{-n}\bigvee_{i=0}^{N+m-1-n} T^{-i}\mathcal{A}$. Therefore $\frac{1}{N}h(\bigvee_{i=0}^{N-1} T^{-i}\bigvee_{j=n}^{m} T^{-j}\mathcal{A}) = \frac{1}{N}h(\mathcal{T}^{-n}\bigvee_{i=0}^{N+m-1-n}$

$= \frac{N+m-n-1}{N}\frac{1}{N+m-n-1}h(\bigvee_{i=0}^{N+m-1-n} T^{-i}\mathcal{A})$. Letting $N \to \infty$ we get 5).

We now prove 6): $h(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}, T^k) = kh(\mathcal{A}\ ,T)$ for $k = 1, 2, ...$

We have $\frac{1}{n}h(\bigvee_{i=0}^{n-1} T^{-ik}(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A})) = k\frac{1}{nk}h(\bigvee_{j=0}^{nk-1} T^{-j}\mathcal{A})$ since $\{ik+j : 0 \leq i \leq n-1, 0 \leq j \leq k-1\} = \{0, 1, ..., nk-1\}$. Letting $n \to \infty$ we get $h(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A}, T) = kh(\mathcal{A}, T)$.

Before proving 7) let us observe the following:

Corollary

If T is rotation on S^1 by a root of unity then $h(\mathcal{A}, T) = 0$ for any $\mathcal{A}$.

Indeed $T^{-i} = I$ for some i and so $h(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A}, T) : k = 1, 2, ...\}$ is a finite set! It follows that $\{kh(\mathcal{A}, T)\}$ is bounded and hence $h(\mathcal{A}, T) = 0$.

We shall see later that the conclusion holds for all rotations of S^1.

Proof of property 7) of the theorem:
$h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}) \leq h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A} \bigvee \bigvee_{i=0}^{n-1} T^{-i}\mathcal{B}) = h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}) + h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}| \bigvee_{i=0}^{n-1} T^{-i}\mathcal{B})$.
By 8), 5) and 9) of previous theorem we have $h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}| \bigvee_{i=0}^{n-1} T^{-i}\mathcal{B}) \leq \sum_{j=0}^{n-1} h(T^{-j}\mathcal{A}| \bigvee_{i=0}^{n-1} T^{-i}\mathcal{B}) \leq \sum_{j=0}^{n-1} h(T^{-j}\mathcal{A}|T^{-j}\mathcal{B}) = \sum_{j=0}^{n-1} h(\mathcal{A}|\mathcal{B}) = nh(\mathcal{A}|\mathcal{B})$. This gives $\frac{1}{n}h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}) \leq \frac{1}{n}h(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}) + h(\mathcal{A}|\mathcal{B})$. The proof is completed by letting $n \to \infty$.

Definition: the entropy $h(T)$ of a m.p. transformation T is defined by $h(T) = \sup\{h(\mathcal{A}, T) : \mathcal{A}$ is a finite sigma field contained in $\mathcal{F}\}$.
Thus rotation by a root of unity has entropy 0.

Theorem [Kolmogorov, Sinai]

If T is i.m.p. and $\mathcal{A}$ is a finite sub sigma field such that $\bigvee_{n=-\infty}^{\infty} T^n(\mathcal{A}) = \mathcal{F}$ then $h(T) = h(\mathcal{A}, T)$.

Lemma
Let $\mathcal{A} \subset \sigma(\mathcal{A}_0)$ where $\mathcal{A}$ and $\mathcal{A}_0$ are fields contained in $\mathcal{F}$ and $\mathcal{A}$ is finite . Let $\epsilon > 0$. Then there is a finite field $\mathcal{B}$ contained in $\mathcal{A}_0$ such that $h(\mathcal{A}|\mathcal{B}) < \epsilon$. The hypothesis that $\mathcal{A} \subset \sigma(\mathcal{A}_0)$ can be weakened to the condition that each $A \in \mathcal{A}$ differs from a set E in $\sigma(\mathcal{A}_0)$ by a null set.

Let $\mathcal{A} = \{A_1, A_2, ..., A_k\}$ and assume (without loss of generality) that $P(A_i) > 0$ for each i. Let $\phi(t) = -t \log t, 0 < t \leq 1$ and $\phi(0) = 0$. We can find $\delta \in (0, 1)$ such that $\phi(t) < \epsilon/k$ for $0 \leq t \leq \delta$ as well as for $1 - \delta \leq t \leq 1$.

Claim: there is a finite field $\mathcal{B}$ contained in $\mathcal{A}_0$ generated by a partition ($\{B_1, B_2, ..., B_k\}$ with same number of sets as $\{A_1, A_2, ..., A_k\}$) such that $P(A_i|B_i) > 1 - \delta, i = 1, 2, ..., k$.

Let us first see how this claim proves the lemma. We have $P(A_j|B_i) < \delta$ for $j \neq i$ since $A_j \subset A_i^c$. Hence $h(\mathcal{A}|\mathcal{B}) = -\sum_{i,j=1}^{k} P(A_i \cap B_j) \log P(A_i|B_j) =$

$-\sum_{i \neq j} P(A_i \cap B_j) \log P(A_i|B_j) - \sum_{i=1}^{k} P(A_i \cap B_i) \log P(A_i|B_i)$

$= \sum_{i \neq j} P(B_j)\phi(P(A_i|B_j)) + \sum_{i=1}^{k} P(B_i)\phi(P(A_i|B_i)) < \frac{\epsilon}{k}\{\sum_{i \neq j} P(B_j) + \sum_{i=1}^{k} P(B_i)\} =$
ϵ. It remains to prove the claim. Let $\lambda > 0$ be so small that $(k-1)(\lambda + k(k-$

$1)\lambda) < \delta \min\limits_{1\leq i\leq k} \frac{P(A_i)}{2}$. Let $\rho > 0$ be such that $(k-1)[\rho + 2k(k-1)\rho] < \lambda$ For $1 \leq i \leq k$ choose $C_i \in \mathcal{A}_0$ such that $P(A_i \Delta C_i) < \rho$. Let $E = \bigcup\limits_{i\neq j}(C_i \cap C_j)$. Let $B_i = C_i \backslash E$ for $1 \leq i \leq k-1$ and $B_k = \Omega \backslash \bigcup\limits_{i=1}^{k-1} B_i$. We only have to verify that $P(A_i|B_i) > 1-\delta, i = 1, 2, ..., k$. For this it suffices to show that $P(A_i \Delta B_i) < \lambda$ for each i. For, we would then have $P(A_i) \leq P(B_i) + \lambda < P(B_i) + \frac{P(A_i)}{2}$ so $P(A_i) < 2P(B_i)$ which implies $P(B_i) - P(A_i \cap B_i) < \lambda < \delta P(B_i)$ and $P(A_i \cap B_i) > (1-\delta)P(B_i)$ or $P(A_i|B_i) > 1-\delta$ as required. If $i \neq j$ then $P(C_i \cap C_j) \leq P(C_i \Delta A_i) + P(A_j \Delta C_j) < 2\rho$ and $P(E) < 2k(k-1)\rho$. Thus, for $i < k$, $P(A_i \Delta B_i) < \rho + 2k(k-1)\rho < \lambda$. Finally $P(A_k \Delta B_k) = P(\bigcup\limits_{i=1}^{k-1} A_i \Delta \bigcup\limits_{i=1}^{k-1} B_i) \leq (k-1)[\rho + 2k(k-1)\rho] < \lambda$.

Proof of Kolmogorov-Sinai Theorem:

we have $\bigvee\limits_{n=-\infty}^{\infty} T^n(\mathcal{A}) = \mathcal{F}$. Let $\mathcal{B}$ be any finite subfield of $\mathcal{F}$. We have to show that $h(\mathcal{B}, T) \leq h(\mathcal{A}, T)$. Let $\mathcal{A}_N = \bigvee\limits_{n=-N}^{N} T^n(\mathcal{A})$. By 5) of previous theorem we have $h(\mathcal{A}_N, T) = h(\mathcal{A}, T)$. Hence $h(\mathcal{B}, T) \leq h(\mathcal{A}_N, T) + h(\mathcal{B}|\mathcal{A}_N)$
$= h(\mathcal{A}, T) + h(\mathcal{B}|\mathcal{A}_N)$. It suffices to show that $h(\mathcal{B}|\mathcal{A}_N) \to 0$ as $N \to \infty$. Let $\mathcal{A}_0 = \bigcup\limits_n \mathcal{A}_n$. $\mathcal{A}_0$ is a field which generates $\mathcal{F}$ which contains $\mathcal{B}$. Hence, if $\epsilon > 0$ is given we can find a partition $\mathcal{C} = \{C_1, C_2, ...C_m\}$ with each $C_i \in \mathcal{A}_0$ such that $h(\mathcal{B}|\mathcal{C}) < \epsilon$. There exists n_0 such that each $C_i \in \mathcal{A}_{n_0}$. For $N \geq n_0$ we have $h(\mathcal{B}|\mathcal{A}_N) \leq h(\mathcal{B}|\mathcal{A}_{n_0}) \leq h(\mathcal{B}|\mathcal{C}) < \epsilon$.

Remark

There is a version of this theorem when T is not invertible. If $\bigvee\limits_{n=0}^{\infty} T^{-n}(\mathcal{A}) = \mathcal{F}$ then $h(T) = h(\mathcal{A}, T)$. The proof is similar. [In place of $\bigvee\limits_{n=-N}^{N} T^n(\mathcal{A})$ in above proof we use $\bigvee\limits_{n=0}^{N} T^n(\mathcal{A})$. We omit the details].

Example: consider a stationary sequence $\{X_n\}_{-\infty}^{\infty}$ with state space $\{1, 2, ..., N\}$. The canonical version of this makes the projection maps $\{..., p_{-1}, p_0, p_1, p_2, ...\}$ a stationary sequence on $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$ for a suitable probability measure P. Let T be the shift transformation: $\{\omega_n\} \to \{\omega_{n+1}\}$. The sigma field $\mathcal{F} = \mathcal{B}^\infty$ is generated by $\bigcup\limits_{n=-\infty}^{\infty} T^{-n}\mathcal{A}$ where $\mathcal{A}$ is the field generated by the sets $\{\omega : \omega_0 =$

$i\}, 1 \le i \le N$. By Kolmogorov-Sinai Theorem we have $h(T) = h(\mathcal{A}, T)$. Hence $h(T) = \lim_{n\to\infty} h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A}))$. A partition for $\bigvee_{k=1}^{n} T^k(\mathcal{A})$ is the family of sets $\{\omega : \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$. Hence $h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A}))$

$= \sum_{i_{-1}, i_{-2}, ..., i_{-n}} P\{\{\omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \sum_{i_0} \phi(P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$. From this it is clear that $h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A})) \le \log N$. [$\sum_{i_0} \phi(P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$

$= -\sum_{i_0} P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \log P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$

$= \sum_{i_0} P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \log \frac{1}{P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}}$

$\le \log \sum_{i_0} P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \frac{1}{P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}} = \log N$].

Now we assume that T is a Bernoulli shift so that $p_n' s$ are i.i.d. with distribution $\{\alpha_1, \alpha_2, ..., \alpha_N\}$. We have

$h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A}))$

$= \sum_{i_{-1}, i_{-2}, ..., i_{-n}} P\{\{\omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \sum_{i_0} \phi(P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$

$= \sum_{i_{-1}, i_{-2}, ..., i_{-n}} \alpha_{i_{-1}} \alpha_{i_{-2}} ... \alpha_{i_{-n}} \sum_{i_0} \phi(\alpha_{i_0}) = -\sum_{i_0} \alpha_{i_0} \log \alpha_{i_0}$. Thus $h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A})) = -\sum_{i=1}^{N} \alpha_i \log \alpha_i$ for every n which implies that $h(T) = -\sum_{i=1}^{N} \alpha_i \log \alpha_i$.

In the case of a Markov shift with initial distribution $\{\alpha_1, \alpha_2, ..., \alpha_N\}$ and transition matrix $((q_{ij}))$ we get

$h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A}))$

$= \sum_{i_{-1}, i_{-2}, ..., i_{-n}} P\{\{\omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\} \sum_{i_0} \phi(P\{\omega_0 = i_0 | \omega_{-1} = i_{-1}, \omega_{-2} = i_{-2}, ..., \omega_{-n} = i_{-n}\}$

$= h(\mathcal{A}| \bigvee_{k=1}^{n} T^k(\mathcal{A}))$

$= \sum_{i_{-1}, i_{-2}, ..., i_{-n}} \alpha_{i_{-1}} q_{i_{-1} i_{-2}} ... q_{i_{-n} i_{-(n-1)}} \sum_{i_0} \phi(q_{i_{-1} i_0}) = \sum_{i_{-1}} \alpha_{i_{-1}} \sum_{i_0} \phi(q_{i_{-1} i_0}) =$

$\sum_{i,j} \alpha_i q_{ij} \log q_{ij}$ for each n and so $h(T) = \sum_{i,j} \alpha_i q_{ij} \log q_{ij}$.

Theorem

If T is i.m.p. and $\bigvee_{n=0}^{\infty} T^{-n}(\mathcal{A}) = \mathcal{F}$ then $h(T) = 0$.

Proof: $\mathcal{A} \subset \mathcal{F} \subset T^{-1}(\mathcal{F}) = \bigvee_{n=1}^{\infty} T^{-n}(\mathcal{A})$. Hence (by the lemma used in the proof of Kolomogorov-Sinai Theorem) we can find a finite field $\mathcal{B}$ contained in the field $\bigcup_m \bigvee_{n=1}^{m} T^{-n}(\mathcal{A})$ with $h(\mathcal{A}|\mathcal{B}) < \epsilon$. For large $n, \mathcal{B} \subset \bigvee_{n=1}^{m} T^{-n}(\mathcal{A})$ and $h(\mathcal{A}| \bigvee_{i=1}^{n} T^{-i}(\mathcal{A})) \leq h(\mathcal{A}|\mathcal{B}) < \epsilon$. Letting $n \to \infty$ we get $h(\mathcal{A}, T) \leq \epsilon$. Since ϵ is arbitrary we get $h(\mathcal{A}, T) = 0$. By Kolmogorov - Sinai Theorem we have $H(T) = h(\mathcal{A}, T) = 0$.

Corollary

All rotations on S^1 have entropy 0.

Proof: let $Tz = az$ and $\mathcal{A} = \{A, B\}$ where $A = \{e^{2\pi it} : 0 \leq t < \frac{1}{2}\}$ and $B = A^c$. We claim that $\bigvee_{n=0}^{\infty} T^n(\mathcal{A}) = \mathcal{F}$, the Borel sigma field of S^1. If a is not a root of unity then $\{a^n : n = 1, 2, ...\}$ is dense in S^1 and $\bigvee_{n=0}^{\infty} T^n(\mathcal{A})$ contains half circles starting at points of a dense set. Hence $\bigvee_{n=0}^{\infty} T^n(\mathcal{A}) = \mathcal{F}$ and T has entropy 0. If a is a root of unity we have already proved that the entropy is 0.

Theorem

$h(T^k) = kh(T)$ for any positive integer k. If T is invertible measure preserving then $h(T^{-1}) = h(T)$ and $h(T^k) = |k|\, h(T)$ for all integers k.

Proof: we claim that $h(\bigvee_{i=0}^{k-1} T^{-i}(\mathcal{A}), T^k) = kh(\mathcal{A}, T)$. Note that $\frac{1}{N} h(\bigvee_{i=0}^{N-1} T^{-ik}\{\bigvee_{j=0}^{k-1} T^{-j}(\mathcal{A})\} = k \frac{1}{Nk} h(\bigvee_{i=0}^{Nk-1} T^{-i}(\mathcal{A}))$. Letting $N \to \infty$ we get $h(\bigvee_{j=0}^{k-1} T^{-j}(\mathcal{A}), T^k) = kh(\mathcal{A}, T)$. On the one hand this gives $kh(\mathcal{A}, T) \leq h(T^k)$ for all $\mathcal{A}$ so $kh(T) \leq h(T^k)$ and on the other hand it gives $h(\mathcal{A}, T^k) \leq h(\bigvee_{j=0}^{k-1} T^{-j}(\mathcal{A}), T^k) = kh(T)$ so $h(T^k) \leq$

$kh(T)$. Thus $h(T^k) = kh(T)$. Now let T be i.m.p. Then $h(\bigvee_{j=0}^{k-1} T^j(\mathcal{A})) = h(\bigvee_{j=0}^{k-1} (T^{-1})^{-j}(\mathcal{A}))$. But $h(\bigvee_{j=0}^{k-1} T^j(\mathcal{A})) = h(T^{-(k-1)}(\bigvee_{j=0}^{k-1} T^j(\mathcal{A})))$ since T is m.p. It follows that $h(\bigvee_{j=0}^{k-1} (T^{-1})^{-j}(\mathcal{A})) = h(\bigvee_{j=0}^{k-1} T^j(\mathcal{A})) = h(\bigvee_{j=0}^{k-1} T^{-j}(\mathcal{A}))$. Divide by k and let $k \to \infty$ to get $h(T^{-1}) = h(T)$. If k is a negative integer then $h(T^k) = h((T^{-1})^k) = h(T^{-k}) = -kh(T)$. This completes the proof.

Remark

We state two results without proof:

1. If $\{\mathcal{A}_n\}$ is a sequence of finite fields whose union generates $\mathcal{F}$ then $h(T) = \lim h(\mathcal{A}_n, T)$

2. $h(T \times S) = h(T) + h(S)$

These facts are not used in these notes.

Suppose $(\Omega, \mathcal{F}, P, T)$ and $(\Omega', \mathcal{F}', P', T')$ are DS's. If there exists a bijection $\tau : \Omega \to \Omega$ such that τ and its inverse are measurable, $P'(\tau(A)) = P(A)$ and $\tau(T(\omega)) = T'(\tau(\omega))\ \forall \omega \in \Omega$ we say that the m.p. transformations T and T' are *isomorphic*. Actually, we modify this definition by allowing τ to be a bijective map between sets of full measure in the two spaces where the sets of full measure are invariant under the respective transformations. An important question in Ergodic Theory is: when are two m.p. transformations isomorphic. Entropy is an isomorphism invariant: if T and S are isomorphic then $h(T) = h(S)$. Is the converse true? The answer is no: if T is rotation on S^1 by a root of unity and T is rotation by a non-root of unity then $h(T) = h(S) = 0$. Since T is not ergodic and S is ergodic they are not isomorphic. The celebrated Ornstein Isomorphism Theorem says that two Bernoulli shift with same entropy are isomorphic. The proof of this is not included in these notes.

Definition: let $\mathcal{A}$ be a finite field contained in $\mathcal{F}$ and $\mathcal{G}$ a sub-sigma field of $\mathcal{F}$. We define $h(\mathcal{A}|\mathcal{G})$ as $E\sum_{i=1}^{n} \phi(P(A_i|\mathcal{G}))$ where) $\{A_1, A_2, ..., A_n\}$ is a partition generating $\mathcal{A}$ and $\phi(x) = -x\log x$. If $\mathcal{G}$ is generated by a finite partition $\{B_1, B_2, ..., B_m\}$ then $P(A_i|\mathcal{G}) = \sum_{j=1}^{m} P(A_i|B_j)I_{B_j}$ so $h(\mathcal{A}|\mathcal{G}) = \sum_{j=1}^{m}\sum_{i=1}^{n} \phi(P(A_i|B_j))P(B_j) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i|B_j)P(B_j)\log P(A_i|B_j)$ so our new definition agrees with the old definition in this case. We have $h(\mathcal{A}|\mathcal{G}) = -E\sum_{i=1}^{n} P(A_i|\mathcal{G})\log P(A_i|\mathcal{G}) = E\{-\sum_{i=1}^{n} I_{A_i}\log P(A_i|\mathcal{G})\}$. Indeed the last expression can be evaluated by conditioning on $\mathcal{G}$ and this yields the last equality. For future reference we state this as:

Theorem
$h(\mathcal{A}|\mathcal{G}) = E\{-\sum_{i=1}^{n} I_{A_i} \log P(A_i|\mathcal{G})\}.$

Theorem
1) $h(\mathcal{A}\bigvee \mathcal{B}|\mathcal{G}) = h(\mathcal{A}|\mathcal{G}) + h(\mathcal{B}|\mathcal{A}\bigvee \mathcal{G})$
2) $h(\mathcal{A}|\mathcal{G}) \leq h(\mathcal{B}|\mathcal{G})$ if $\mathcal{A} \subset \mathcal{B}$
3) $h(\mathcal{A}|\mathcal{G}_1) \leq h(\mathcal{A}|\mathcal{G}_2)$ if $\mathcal{G}_2 \subset \mathcal{G}_1$
4) $h(\mathcal{A}\bigvee \mathcal{B}|\mathcal{G}) \leq h(\mathcal{A}|\mathcal{G}) + h(\mathcal{B}|\mathcal{G})$
5) $h(T^{-1}\mathcal{A}|T^{-1}\mathcal{G}) = h(\mathcal{A}|\mathcal{G})$

Proof: we have $P(B|\mathcal{A}\bigvee \mathcal{G}) = \sum_i I_{A_i} \frac{P(B\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})}$ since $E\sum_i I_{A_i} \frac{P(B\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} I_C I_{A_j} =$ $EE(I_{A_j} \frac{P(B\cap A_j|\mathcal{G})}{P(A_j|\mathcal{G})} I_C|\mathcal{G})$
$= EP(A_j|\mathcal{G}) \frac{P(B\cap A_j|\mathcal{G})}{P(A_j|\mathcal{G})} I_C = E\sum_i P(B \cap A_j|\mathcal{G}) I_C = P(B \cap A_j \cap C)$ for all $C \in \mathcal{G}$ and for all j, proving that $E\sum_i I_{A_i} \frac{P(B\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} I_C I_A = P(B \cap A \cap C)$ for all $C \in \mathcal{G}$ and for all $A \in \mathcal{A}$ (because every set in $\mathcal{A}$ is a disjoint union of $A_j's$). Since $\mathcal{A}\bigvee \mathcal{G}$ is generated by sets of the type $A \cap C$ with $A \in \mathcal{A}$ and $C \in \mathcal{G}$ we have proved that $P(B|\mathcal{A}\bigvee \mathcal{G}) = \sum_i I_{A_i} \frac{P(B\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})}$.[The $\pi - \lambda$ Theorem may be used here].

Now $h(\mathcal{B}|\mathcal{A}\bigvee \mathcal{G}) = -E\sum_j P(B_j|\mathcal{A}\bigvee \mathcal{G}) \log P(B_j|\mathcal{A}\bigvee \mathcal{G})$
$= -E\sum_j \sum_i I_{A_i} \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} \log \sum_i I_{A_i} \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})}$
$= -E\sum_{i,j} I_{A_i} \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} \log \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})}$
$= -E\sum_{i,j} I_{A_i} \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} \log P(B_j\cap A_i|\mathcal{G}) + E\sum_{i,j} EI_{A_i} \frac{P(B_j\cap A_i|\mathcal{G})}{P(A_i|\mathcal{G})} \log P(A_i|\mathcal{G})$
$= -E\sum_{i,j} P(B_j \cap A_i|\mathcal{G}) \log P(B_j \cap A_i|\mathcal{G}) + E\sum_{i,j} P(B_j \cap A_i|\mathcal{G}) \log P(A_i|\mathcal{G})$
(where we used conditioning on $\mathcal{G}$)
$= h(\mathcal{A}\bigvee \mathcal{B}|\mathcal{G}) + E\sum_i P(A_i|\mathcal{G}) \log P(A_i|\mathcal{G}) = h(\mathcal{A}\bigvee \mathcal{B}|\mathcal{G}) - h(\mathcal{A}|\mathcal{G})$ proving that $h(\mathcal{A}\bigvee \mathcal{B}|\mathcal{G}) = h(\mathcal{A}|\mathcal{G}) + h(\mathcal{B}|\mathcal{A}\bigvee \mathcal{G})$. We have proved 1). 2) follows trivially from 1). It remains only to prove 3) since 4) follows from 3) and 1) (and 5) is trivial). To show $h(\mathcal{A}|\mathcal{G}_1) \leq h(\mathcal{A}|\mathcal{G}_2)$ if $\mathcal{G}_2 \subset \mathcal{G}_1$. Let $X = P(A|\mathcal{G}_1)$. We have $E(\phi(X)|\mathcal{G}_2) \leq \phi(E(X|\mathcal{G}_2) = \phi(P(A|\mathcal{G}_2))$. Hence

$$\sum_i E(\phi(P(A_i|\mathcal{G}_1))|\mathcal{G}_2) \leq E\sum_i \phi(P(A_i|\mathcal{G}_2)).$$ The right side is $h(\mathcal{A}|\mathcal{G}_2)$. Note that $E(\phi(P(A_i|\mathcal{G}_1))|\mathcal{G}_2) = E(\phi(P(A_i|\mathcal{G}_1))$ so the left side is $\sum_i E(\phi(P(A_i|\mathcal{G}_1)) = h(\mathcal{A}|\mathcal{G}_1)$. This finishes the proof.

Theorem
If $\mathcal{G}_n \uparrow \mathcal{G}$ then $h(\mathcal{A}|\mathcal{G}_n) \to h(\mathcal{A}|\mathcal{G})$ a.e.

We have $h(\mathcal{A}|\mathcal{G}_n) = E\sum_{i=1}^{k} \phi(P(A_i|\mathcal{G}_n)) \to E\sum_{i=1}^{k} \phi(P(A_i|\mathcal{G})) = h(\mathcal{A}|\mathcal{G})$ by Bounded Convergence Theorem.

Remark: we could have used this in the proof of Kolmogorv-Sinai Theorem: $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ and $\mathcal{A} \subset \bigvee_{k=-\infty}^{\infty} \mathcal{A}_n$ implies $\lim h(\mathcal{A}|\mathcal{A}_n) = h(\mathcal{A}|\bigvee_{k=-\infty}^{\infty} \mathcal{A}_n) \leq h(\mathcal{A}|\mathcal{A}) = 0$. In the proof of Kolmogorv-Sinai Theorem we had to prove that $h(\mathcal{B}|\bigvee_{k=-n}^{n} T^k\mathcal{A}) \to 0$. We know that $h(\mathcal{B}|\bigvee_{k=-n}^{n} T^k\mathcal{A}) \to h(\mathcal{B}|\bigvee_{k=-\infty}^{\infty} T^k\mathcal{A}) = h(\mathcal{B}|\mathcal{F}) = 0$.

SHANNON-McMILLAN-BREIMAN THEOREM

Let T be the shift associated with a stationary sequence.

Let $\Omega = \prod_{-\infty}^{\infty}\{1, 2..., N\}, \mathcal{F} =$ cylinder sigma field and P a stationary measure (i.e. a probability measure which makes the shift transformation $\{\omega_n\} \to \{\omega_{n+1}\}$ m.p.). Let $\mathcal{A}_n$ be the field generated by the partition $\{\omega : \omega_0 = i_0, \omega_1 = i_1, ..., \omega_n = i_n\}$ ($i'_j s \in \{1, 2, ..., N\}$). By Kolmogorov-Sinai Theorem $h(T) =$ where $h_{0,n-1} = h(\mathcal{A}_n, T)$. [$\bigvee_{n=-\infty}^{\infty} T^n(\mathcal{A}_0) = \mathcal{F}$ then $h(T) = h(\mathcal{A}_0, T) =. \lim_{n\to\infty} \frac{1}{n} h(\bigvee_{i=0}^{n} T^{-i}\mathcal{A}) = h(\mathcal{A}_n, T)$].

Let $p(i_0.i_1, ..., i_n) = P\{\omega : \omega_0 = i_0, \omega_1 = i_1, ..., \omega_n = i_n\}$. We have $h_{0,n-1} = E - \log p(Y_1, Y_2, ..., Y_n)$ where we have denoted the projection maps on Ω by $Y_n, n \in \mathbb{Z}$.

Hence $h(T) = -\lim_{n\to\infty} \frac{1}{n} E \log p(Y_1, Y_2, ..., Y_n)$.

Theorem [Shannon-McMIllan-Brieman]
If T is ergodic then $-\lim_{n\to\infty} \frac{1}{n} \log p(Y_1, Y_2, ..., Y_n) = h(T)$ a.e..

we first remark that if T is a Bernoulli shift then $p(Y_1, Y_2, ..., Y_n) = p_{Y_1}p_{Y_2}....p_{Y_n}$ so $\frac{1}{n}\log p(Y_1, Y_2, ..., Y_n) = \frac{1}{n}\sum_{k=1}^{n}\log p_{Y_k} \to E\log p_{y_1}$ (by SLLN or the ergodic theorem)$= \sum_{k=1}^{N} p_j \log p_j = -h(T)$.

Now let T be an ergodic Markov shift with stationary distribution $\{\pi_i\}$ and transition matrix $((p_{ij}))$. We have $-\lim_{n\to\infty}\frac{1}{n}\log p(Y_1, Y_2, ..., Y_n) = -\lim_{n\to\infty}\frac{1}{n}\log\{\pi_{Y_1}p_{Y_1Y_2}...p_{Y_{n-1}Y_n}\}$

$= -\lim_{n\to\infty}\frac{1}{n}\log\pi_{Y_1} - \sum_{k=1}^{n-1}\lim_{n\to\infty}\frac{1}{n}\log p_{Y_kY_{k+1}} = -E\log p_{Y_1Y_2}$ by the ergodic theorem since $Y_k = T^{-1}(Y_1)\ \forall k$. But $-E\log p_{Y_1Y_2} = -\sum_{i,j=1}^{N}\pi_i p_{ij}\log p_{ij} = h(T)$. Thus we have proved the theorem in these two cases. Now consider the general stationary shift. Let $g_k(\omega) = -\log\frac{p(Y_{-k}(\omega),...,Y_{-1}(\omega),Y_0(\omega))}{p(Y_{-k}(\omega),...,Y_{-1}(\omega))}(k \geq 1)$ with $g_0(\omega) = -\log p_{Y_0}(\omega)$. Let

$g_k^{(i)}(\omega) = -\log\frac{p(Y_{-k}(\omega),...,Y_{-1}(\omega),i)}{p(Y_{-k}(\omega),...,Y_{-1}(\omega))}(k \geq 1)$. Note that these numbers are all non-negative. Now $-\lim_{n\to\infty}\frac{1}{n}\log p(Y_0(\omega), Y_2(\omega), ..., Y_n(\omega)) = \frac{1}{n}\sum_{k=1}^{n-1} g_k(T^k(\omega))$.

This is because $\sum_{k=0}^{n-1} g_k(T^k(\omega)) = -\log p_{Y_0}(\omega) - \sum_{k=1}^{n-1}\log\frac{p(Y_0(\omega),...,Y_{k-1}(\omega),Y_k(\omega))}{p(Y_0(\omega),...,Y_{k-1}(\omega))}$

$= -\log\prod_{k=1}^{n-1}\frac{p(Y_0(\omega),...,Y_{k-1}(\omega),Y_k(\omega))}{p(Y_0(\omega),...,Y_{k-1}(\omega))} = -\log p_{Y_0}(\omega) - \log\frac{p(Y_0(\omega),...,Y_{n-2}(\omega),Y_{n-1}(\omega))}{p(Y_0(\omega))} =$
$-\log p(Y_0(\omega), ..., Y_{n-2}(\omega), Y_{n-1}(\omega))$ (*)

. Now $P\{Y_0 = i|Y_{-k}, Y_{-k-1}, ..., Y_{-1}\} \to P\{Y_0 = i|Y_{-1}, Y_{-2}, ...\}$ and hence $g_k(\omega) = -\log P\{Y_0 = i|Y_{-k}, Y_{-k-1}, ..., Y_{-1}\} \to -\log P\{Y_0 = i|Y_{-1}, Y_{-2}, ...\}$ on $\{Y_0 = i\}$. In other words $g_k(\omega) \to g(\omega)$

where $g(\omega) = -\log p\{Y_0|Y_{-1}, Y_{-2}, ...\}$. We now show that $E\sup\{g_k : k \geq 0\} < \infty$. For this let $\lambda > 0$ and $E_k = \{\omega : \max_{1\leq j<k} g_j \leq \lambda < g_k(\omega)\}$. Then $P(E_k) = \sum_i P\{\{Y_0 = i\} \cap E_k\}$. Let $F_{k,i} = \{\omega : \max_{1\leq j<k} g_j^{(i)}(\omega) \leq \lambda < g_k^{(i)}(\omega)\}$. Then $g_k = g_k^{(i)}$ on $\{Y_0 = i\}$ so $P(E_k) = \sum_i P\{\{Y_0 = i\} \cap F_{k,i}\}$. Noting that $F_{k,i} \in \sigma\{Y_{-k}, Y_{-k+1}, ..., Y_{-1}\}$ we get

$P\{Y_0 = i\} \cap F_{k,i}\} = \int_{F_{k,i}} e^{-g_k^{(i)}}dP$ [since $g_k^{(i)}(\omega) = -\log\frac{p(Y_{-k}(\omega),...,Y_{-1}(\omega),i)}{p(Y_{-k}(\omega),...,Y_{-1}(\omega))}$ so $e^{-g_k^{(i)}} = P\{Y_0 = i|Y_{-k}, ..., Y_{-1}\}$].

Hence $P\{Y_0 = i\} \cap F_{k,i}\} \leq e^{-\lambda}P\{F_{k,i}\}$. Thus $P\{\sup\{g_k : k \geq 0\} > \lambda\} = \sum_k P\{E_k\} = \sum_k\sum_i P\{\{Y_0 = i\} \cap F_{k,i}\} \leq e^{-\lambda}\sum_k\sum_i P\{F_{k,i}\} \leq e^{-\lambda}N$

since $\sum_k P\{F_{k,i}\} \leq P\{\Omega\} = 1$ for each i. It follows from this that $\sup\{g_k : k \geq 0\}$ has finite expectation. By dominated convergence theorem we get $Eg = \lim_{k\to\infty} Eg_k$. Since T is m.p. we have $E\frac{1}{n}\sum_{k=1}^{n} g_k(T^k(\omega)) = Eg_k$. Thus $Eg = \lim_{k\to\infty} E\frac{1}{n}\sum_{k=1}^{n} g_k(T^k(\omega)) = -\frac{1}{n}\lim_{k\to\infty} E\log p(Y_0, Y_1, ..., Y_k)$ (by (*))$= h(T)$ [by the statement immediately preceding the statement of this theorem]. We now write $\frac{1}{n}\sum_{k=1}^{n} g_k(T^k(\omega))$ as $\frac{1}{n}\sum_{k=1}^{n} g(T^k(\omega)) + \frac{1}{n}\sum_{k=1}^{n}(g_k - g)(T^k(\omega))$ and observe that the first term tends a.s. to Eg (by ergodicity).If we show that the second term tends to 0 a.s. we can conclude that $\frac{1}{n}\sum_{k=1}^{n} g_k(T^k(\omega))\{\equiv -\frac{1}{n}\log p(Y_0(\omega), ..., Y_{n-2}(\omega), Y_{n-1}(\omega))\}$ converges a.s., as required. Let $G_N = \sup_{n\geq N} |g_k - g|$. Then $\limsup_n \left|\frac{1}{n}\sum_{k=1}^{n}\{g_k(T^k(\omega)) - g(T^k(\omega))\}\right| \leq$

$\limsup_n \frac{1}{n}\sum_{k=1}^{n} \left|g_k(T^k(\omega)) - g(T^k(\omega))\right|$

$$\leq \limsup_n \frac{1}{n}\sum_{k=N}^{n} \left|g_k(T^k(\omega)) - g(T^k(\omega))\right| \leq \limsup_n \frac{1}{n}\sum_{k=1}^{n} G_N(T^k(\omega)) = EG_N$$

by the ergodic theorem. This is true for each N and $G_N \to 0$ a.s. as $N \to \infty$. This completes the proof.

TOPOLOGICAL DYNAMICS

Let X be a compact metric space and $T : X \to X$ a homeomorphism. We say T is *minimal* if the orbit $O_T(x) = \{T^n x : n \in \mathbb{Z}\}$ of x is dense in X for each $x \in X$.

Remark: we can study continous maps instead of homeomorphisms. We have assumed that T is a homeomorphism for simplicity.

Theorem

T is minimal iff C closed and $TC = C$ imply $C = \emptyset$ or $C = X$.

Proof: Suppose T is minimal, C is closed and $TC = C$. If $x \in C$ then $O_T(x) \subset C$. Since C is closed and $O_T(x)$ is dense we must have $C = X$. Conversely suppose C closed and $TC = C$ imply $C = \emptyset$ or $C = X$. Let $x \in X$ and C be the closure of $O_T(x)$. Then $TC = C$ but C is neither empty nor equal to X.

Remark: minimality is the topological analog of ergodicity.

Def. A *minimal set* is a non-empty closed set C such that $TC = C$ and the restriction of T to C is minimal.

Theorem
Minimal sets exist.

Proof: let $\mathcal{E}$ be the collection of all non-empty closed sets C such that $TC = C$. Clearly this collections is not empty. Order this by reverse inclusion. If $\{C_i\}$ is a totally ordered family of sets in $\mathcal{E}$ then $\bigcap_i C_i$ is nonempty (by compactness) and belongs to $\mathcal{E}$. Hence there is a minimal element C in $\mathcal{E}$. For any $x \in C$ the closure of the orbit of x is a non-empty closed invariant set. By minimality of C it must coincide with C. Hence C is minimal.

Def. T is called *semi-simple* if there exists a partition of X into closed sets $\{C_i\}$ such that $TC_i = C_i$ for all i and the restriction of T to C_i is minimal for each i.

It will be shown later that ergodic automorphisms of compact groups are not semi-simple.

Def. x is a periodic point of T if $T^n x = x$ for some positive integer n. The least integer with this property is called the period of x.

Theorem
Let T be a minimal homeomorphism. Then
1) $f \circ T = f, f$ continuous implies f is a constant
2) T has no periodic points unless X is a finite set

Proof: let $f \circ T = f, f$ continuous. For any real number a let $C = \{x : f(x) \leq a\}$. C is a closed invariant set. By minimality of T this closed must be empty or X. If $c = \sup\{a \in \mathbb{R} : C = \emptyset\}$ then $f(x) = c$ for all x.

Now suppose x is a periodic point. Then the orbit of x is a (finite, hence closed) set which is also dense, so X is finite.

Example
Converse of 1) is false. An ergodic continuous homomorphism of a compact metric group X is not minimal because the orbit of the identity is not dense; however any invariant function is almost everywhere constant (by ergodicity) and hence a constant if it is continuous. An explicit example is given after the next theorem.

Theorem
Let T be a continuous homomorphism of a compact metric group G such that $\gamma \circ T^n = \gamma$ for some positive integer n and some character γ of G implies $\gamma = 1$. Then T is ergodic. Converse also holds.

Proof: we prove the converse part first. Let T be ergodic and suppose $\gamma \circ T^n = \gamma$ but $\gamma \neq 1$. Let k be the least positive integer with $\gamma \circ T^k = \gamma$. Let $f = \gamma + \gamma \circ T + ... + \gamma \circ T^{k-1}$. Then f is invariant and hence constant a.e.. (hence a constant times the character 1). But the terms of this sum are all distinct characters, hence orthogonal to each other and this cannot happen. Now consider the direct part. Let f be invariant. Let $f = \sum c_n \gamma_n$ be the Fourier series of f. Then $\sum c_n \gamma_n(Tx) = \sum c_n \gamma_n(x)$ and $\gamma_n \circ T^j$ is also a character for each n and each j. If $\gamma_n, \gamma_n \circ T, \gamma_n \circ T^2, ...$ are all distinct characters then they are orthogonal and $< f, \gamma_n \circ T^j >=< f, \gamma_n >$ by invariance of f so the coefficients $< f, \gamma_n \circ T^j >$ in the Fourier expansion are all equal. This implies that they are all 0 since the coefficient sequence belongs to l^2. Thus $< f, \gamma_n >= 0$ in this case. If $f \neq 0$ then there exists n such that $< f, \gamma_n >\neq 0$. For this n it follows that $\gamma_n, \gamma_n \circ T, \gamma_n \circ T^2, ...$ are not all distinct, i.e. there exists $j < k$ such that $\gamma_n \circ T^j = \gamma_n \circ T^k$. Thus $\gamma_n \circ T^p = \gamma_n$ where $p = k - j \in \mathbb{N}$. By hypothesis we must have $\gamma_n = 1$. This conclusion holds for any n such that $< f, \gamma_n >\neq 0$. Thus there is only one non-zero term in the Fourier expansion of f and f is a constant.

Remark: in particular the map $T : S^1 \to S^1$ defined by $Tz = z^n$ (where n is a positive integer) is ergodic iff $z^{nm} = z^m, n \geq 1 \Rightarrow m = 0$. Thus T is ergodic iff $n \neq 1$. Also $z \to \frac{1}{z}$ is not ergodic.

Remark: consider now the map $T(a, b) = (a^3, b^2)$ on the torus. Any character γ is of the type $(a, b) \to a^r b^s$ where r and s are integers. If $\gamma \circ T^n = \gamma$ and $n \in \mathbb{N}$ then $a^{3^n r} b^{2^n s} = a^r b^s$ for all $a, b \in S^1$ which implies $r = s = 0$ and $\gamma = 1$. Thus T is ergodic. Since $T(1, 1) = (1, 1)$ this map is not minimal.

Examples

1. Let T be a rotation on a compact metric group $G : Tg = ag$ where $a \in G$. The T is minimal iff $\{a^n : n \in \mathbb{Z}\}$ is dense. In fact $O_T(e) = \{a^n : n \in \mathbb{Z}\}$ and if this dense then so is $O_T(g) = \{a^n g : n \in \mathbb{Z}\}$ for any g.

2. A continuous automorphism of a compact metric group G is minimal iff $G = \{e\}$. This is trivial since $Te = e$.

3. Consider the shift T on $\Omega = \prod_{-\infty}^{\infty} \{1, 2, ..., N\}$. Under addition modulo N on $\{1, 2, ..., N\}$ and pointwise addition on the product Ω becomes a compact metric group and T is a continuous automorphism. Hence T is not minimal.

Of course the orbit of $\{\omega_n\}$ where $\omega_n = 1$ for all n is a singleton set which is not dense.

Def. A homeomorphism T of a compact metric space X is called *topologically transitive* if $O_T(x)$ is dense for *some* $x \in X$.

Of course, if T is minimal then it is transitive

Theorem

FAE for a homeomorphism T of a compact metric space X
1) T is topologically transitive
2) C closed, $TC = C \Rightarrow C = X$ or C has no interior
3) U open, $TU = U \Rightarrow U = \emptyset$ or U is dense
4) U, V open and non-empty $\Rightarrow T^n(U) \cap V \neq \emptyset$ for some integer n
5) the set of points whose orbits are not dense is of first category

Proof:
1) implies 2): let $O_T(x_0)$ be dense, C closed and $TC = C$. If $x_0 \in C$ then $O_T(x_0) \subset C$ and so $C = X$. Otherwise, $O_T(x_0) \in C^c$ (because $TC = C$). Hence C^c is dense and C has no interior.

Equivalence of 2) and 3) is trivial.

3) implies 4): let U and V be non-empty open sets and suppose, if possible, $T^n(U) \cap V = \emptyset \ \forall n$. Then $\bigcup\limits_n T^n(U)$ is a non-empty open invariant set and hence it is dense by 3). But this set if disjoint from V which is a contradiction.

4) implies 5): let $\{U_n\}$ be a countable base for the topology of X. Then $O_T(x)$ is not dense $\Leftrightarrow O_T(x) \cap U_n = \emptyset$ for some $n \Leftrightarrow T^k x \in U_n^c \ \forall k$ for some $n \Leftrightarrow x \in \bigcup\limits_n \bigcap\limits_k T^{-k}(U_n^c)$. We have to show that $\bigcup\limits_n \bigcap\limits_k T^{-k}(U_n^c)$ is of first category. We prove that $\bigcap\limits_k T^{-k}(U_n^c)$ is nowhere dense. This set is closed so we have to show that it has no interior. If a nonempty open set V is contained in $\bigcap\limits_k T^{-k}(U_n^c)$ then $T^k(V) \cap U_n = \emptyset$ for all k which contradicts 4).

5) implies 1) follows from Baire Category Theorem.

Theorem

Consider a continuous map T of a compact metric space X. Let P be a Borel probability measure whose support is X. Suppose T is m.p. and ergodic w.r.t. P. The the set of points whose orbits are not dense is of measure 0.

Proof: let $\{U_n\}$ be a base for the topology of X. $\{T^n x : n = 0, 1, ...\}$ is dense in X if and only if $x \in \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=0}^{\infty} T^{-k} U_n$. Since $T^{-1}(\bigcup\limits_{k=0}^{\infty} T^{-k} U_n) \subseteq (\bigcup\limits_{k=0}^{\infty} T^{-k} U_n)$ and T is ergodic we must have $P(\bigcup\limits_{k=0}^{\infty} T^{-k} U_n) = 0$ or 1. Since $\bigcup\limits_{k=0}^{\infty} T^{-k} U_n$ contains U_n and $P(U_n) > 0$ we must have $P(\bigcup\limits_{k=0}^{\infty} T^{-k} U_n) = 1$. This is true for each n and hence $P(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=0}^{\infty} T^{-k} U_n) = 1$. Thus the *one-sided* orbit of almost all points are dense.

Corollary

Let T be an affine map on a compact connected metric abelian group G. Then T is ergodic (w.r.t. Haar measure) iff it is topologically transitive.

Proof: let $Tg = g_0Sg$ where $g_0 \in G$ and S is a continuous automorphism. Previous theorem shows that ergodic transformations are transitive. Let T be transitive. Claim: if γ is a character and $\gamma \circ S^k = \gamma$ with $k \in \mathbb{N}$ then $\gamma \circ S = \gamma$. For this let $\gamma_0(g) = \gamma(g^{-1}Sg)) = \frac{\gamma(Sg)}{\gamma(g)}$. Then $\gamma_0(T^k(g))$

$= \gamma_0(g_0Sg_0...S^{k-1}g_0S^kg) = \gamma_0(g_0Sg_0...S^{k-1}g_0)\gamma_0(S^kg) = \gamma(g_0^{-1}(S^kg_0))\gamma_0(S^kg) = \gamma_0(g)$

because $\gamma(S^k(g_0)) = \gamma(g_0)$ and $\gamma_0(S^kg) = \gamma_0(g)$. $[\gamma_0(S^kg) = \frac{\gamma(S^{k+1}g)}{\gamma(S^kg)} = \frac{\gamma(Sg)}{\gamma(g)} = \gamma_0(g)]$. Hence $\gamma_0 \circ T^k = \gamma_0$. Since T is transitive there exists g_1 such that $O_T(g_1)$ is dense. It follows that γ_0 takes only finite number of values on this dense set. Since G is connected γ_0 must be a constant. Thus $\frac{\gamma(Sg)}{\gamma(g)} = 1$ for all g and the claim is proved. Our next claim is that the smallest closed sub-group containing g_0 and the range of the map $g \to g^{-1}Sg$ is the whole of G. Otherwise there is a character γ such that $\gamma(g_0) = 1, \gamma(g^{-1}Sg) = 1$ for all g but $\gamma \neq 1$. [We take this result from Representation Theory for granted]. Note that $\gamma(Tg) = \gamma(g_0Sg) = \gamma(g_0)\gamma(Sg) = \gamma(Sg) = \gamma(g)$ and, by iteration $\gamma(T^ng) = \gamma(g)$ for all n. On the dense set $O_T(g_1)$ the character γ takes only the value $\gamma(g_1)$ and hence $\gamma = 1$, a contradiction. We now prove that the two claims above imply ergodicity of T. Let f be a T invariant function in L^2. Let $f = \sum_n c_n\gamma_n$ be the Fourier series of f. Then $\sum_n c_n\gamma_n(g_oSg) = \sum_n c_n\gamma_n(g)$ or $\sum_n c_n\gamma_n(g_0)\gamma_n(Sg) = \sum_n c_n\gamma_n$. by iteration $\sum_n c_n\gamma_n(g_0)\gamma_n \circ S^k = \sum_n c_n\gamma_n$ for each positive integer k. Fix i. If $\gamma_i, \gamma_i \circ S, \gamma_i \circ S^2, ...$ are distinct characters then, since $|\gamma_i(g_0)| = 1$, infinitely many coefficients in the series on the left will have the same modulus (viz. $|c_i|$) forcing this coefficient to be 0 [becuase the coefficient sequence is square integrable]. Thus $< f, \gamma_i > \neq 0$ implies $\gamma_i \circ S^n = \gamma_i \circ S^m$ with $n \neq m$. But then $\gamma_i \circ S^{|n-m|} = \gamma_i$. By the first claim above we conclude that $\gamma_i \circ S = \gamma_i$. But then $\gamma_i = 1$ on $\{g^{-1}Sg : g \in G\}$. From the equation $\sum_n c_n\gamma_n(g_0)\gamma_n(Sg) = \sum_n c_n\gamma_n$ we conclude that $\gamma_i(g_0) = 1$. By Claim 2 the subgroup generated by $\{g^{-1}Sg : g \in G\}$ and g_0 is dense. It follows that $\gamma_i = 1$ whenever $< f, \gamma_i > \neq 0$ implying that f is a constant. This finishes the proof.

Theorem

If T is topologically transitive and f is a continuous invariant function then f is a constant.

Proof: this is obvious.

Some examples:

we first give an example to show that the converse of above theorem is false. Let X be the disjoint union of two copies of the torus joined at the identity, i.e. $X = (\{S^1 \times S^1 \times \{0\}) \cup (\{S^1 \times S^1 \times \{1\})/\tilde{}$ where $\tilde{}$ identifies $(1,1,0)$ and $(1,1,1)$. Let S be an ergodic automorphism of the torus and $T(a,b,0) = (S(a,b),0), T(a,b,1) = (S(a,b),1)$. Then T maps each of the two tori into themselves so no orbit can be dense. If f is a continuous invariant function then it is constant on each of the two tori and the constants must be the same since there is a common point $(1,1,0) = (1,1,1)$.

Our next example shows that $O_T(x)$ can be dense for some x and a finite set for a set of points that is dense! Let T be a continuous ergodic automorphism of the torus $S^1 \times S^1$ and $E = \bigcup_{n=0}^{\infty} \{(a,b) \in S^1 \times S^1 : a^n = 1 = b^n\}$. First note that any point in S^1 can be approximated by a root of unity. [This just the statement that the set of rationals in dense in $\mathbb{R}$]. Given two points a and b in S^1 we can approximate them by roots of unity c and d and there exists n such that c^n and d^n are both 1. Hence E is dense in the torus. If $a^n = 1 = b^n$ then $T^n(a,b) = T\{(a,b)^n\} = \{T(a^n,b^n)\} = T(1,1) = (1,1)$. Hence the orbit of each point of E is a finite set. However, by an earlier theorem the set of points whose orbits are dense has full measure since T is ergodic.

Theorem

Let X be a compact metric space and $T : X \to X$ a topologically transitive homeomorphism. If there is an equivalent metric which makes T an isometry then T is minimal.

Proof: let d be a metric for X which makes T an isometry. Let x_0 be a point whose orbit is dense. Let $x \in X$. We have to show that the orbit of x is dense. Let $y \in X$ and $\epsilon > 0$. We can find integers n, m such that $d(x, T^n(x_0)) < \epsilon$ and $d(y, T^m(x_0)) < \epsilon$. Now $d(y, T^{m-n}(x)) \leq d(y, T^m(x_0)) + d(T^m x_0, T^{m-n}(x))$
$= d(y, T^m(x_0)) + d(T^n(x_0), x) < 2\epsilon$. This finishes the proof.

Remark: let $Tg = ag$ on a compact connected abelian metric group G. Then TFE

1) T is ergodic
2) T is minimal
3) T is topologically transitive
4) $\{a^n : n \in \mathbb{Z}\}$ is dense.

[we have proved 1) and 3) are equivalent. The equivalence of 2), 3) and 4) is trivial].

Conjugacy and spectrum:

let T and S be homeomorphisms of compact Hausdorff spaces X and Y respectively. We say that T and S are (topologically) conjugate if there is a homeomorphism $\phi : X \to Y$ such that $S \circ \phi = \phi \circ T$.

Conjugacy is an equivalence relation and minimality as well as transitivity are conjugacy invariants. [This means that if T and S are conjugate and one of them has one of these properties so does the other]

Let T be a homeomorphism of compact metric space X. If $f \circ T = \lambda f$ where $f \in C(X)\backslash\{0\}$ and $\lambda \in \mathbb{C}$ then λ is called an eigen function of T corresponding to eigen value λ.

In the next four theorems T is a transitive homeomorphism of a compact metric space.

Theorem
If and $f \circ T = \lambda f$ where $f \in C(X)\backslash\{0\}$ and $\lambda \in \mathbb{C}$ then $|\lambda| = 1$ and $|f|$ is a constant.

Proof: $\sup\{|f(T(x))| : x \in X\} = \sup\{|f(x)| : x \in X\} \neq 0$ since T is bijective. However the left side also equals $\sup\{|\lambda|\,|f(x)| : x \in X\}$. It follows that $|\lambda| = 1$. It follows that $|f|\,(Tx) = |\lambda|\,|f(x)| = |f(x)|$ so $|f|$ is constant on the orbit of any point. Since there is a dense orbit,$|f|$ must be a constant.

Theorem
The eigen space corresponding to a given eigen value is one dimensional.

Proof: let $f \circ T = \lambda f, g \circ T = \lambda g, f \neq 0, g \neq 0$. By previous theorem f and g never vanish. It follows that $\frac{f}{g}$ is invariant. It is constant on a dense set hence on X.

Theorem
Eigen functions corresponding to distinct eigen values are linearly independent.

Suppose $f_i \circ T = \lambda_i f_i, 1 \leq i \leq N$, with $\lambda_i's$ distinct. Suppose $\sum_{i=1}^{N} a_i f_i = 0$. We get $\sum_{i=1}^{N} a_i \lambda_i^j f_i(x) = \sum_{i=1}^{N} a_i f_i(T^j(x)) = 0$ for $0 \leq j \leq N-1$. View this as a system of N linear equations in the N variables $a_i f_i(x), 1 \leq i \leq N$. The coefficient matrix of this system of linear equations is non-singular.[This is a vander Monde matrix and the determinant is the product of the numbers $(\lambda_k - \lambda_j)$ with $k < j$]. Hence $a_i f_i(x) = 0$ for all i and x. But $f_i \neq 0$ implies $f_i(x) \neq 0$ for all x and hence $a = 0$ for all i.

Theorem
Eigen values of T form a subgroup of S^1.

Proof: this is trivial.

Theorem

T has at most countably many eigen values.

Proof: let $f\circ T=\lambda f, f\neq 0, \lambda\neq 1$. We may suppose that $f(x_0)=1$ for some x_0. We claim that $\|f-1\|\geq 1/2$ where $\|\|$ is the supremum norm. Choose $n\in\mathbb{N}$ such that $|\lambda^n-1|\geq 1/2$. [If λ is not a root of unity then $\{\lambda^n : n\geq 1\}$ is dense. If it is an $N-th$ root of unity where $N\geq 2$ let $\lambda_1=\lambda e^{2\pi ik/N}$ where $k=N/2$ or $(N-1)/2$ according as N is even or odd. We claim that either $|\lambda-1|\geq 1/2$ or $|\lambda_1-1|\geq 1/2$. If these inequalities are both false then $|\lambda-\lambda_1|<1$ which implies $\left|1-e^{2\pi ik/N}\right|<1$. This gives $2-2\cos(2\pi k/N)<1$ or $\cos(2\pi k/N)>1/2$. When N is even this gives the contradiction $-1>1/2$ and when N is odd we get $\cos(\pi(1-\frac{1}{N}))>1/2$ which is again a contradiction since $\pi/2\leq\pi(1-\frac{1}{N})<\pi$. Now observe that λ_1 is also an $N-th$ root of unity and hence it is of the type λ^n for some positive integer n].

Then $\|f-1\|\geq|f(T^n x_0)-1|=|\lambda^n f(x_0)-1|\geq 1/2$. Now if there are uncountable many eigen values then there are uncountable many eigen functions $f_i(i\in I)$ associated with distinct eigen values such that $|f_i|\equiv 1\ \forall i$. Thus $f_i/f_{i'}$ is an eigen function associated with an eigen value $\neq 1$ so $\|f_i/f_{i'}-1\|\geq 1/2$ whenever $i\neq i'$. But then $\|f_i-f_{i'}\|\geq 1/2$ whenever $i\neq i'$. This contradicts the separability of $C(X)$.

Definiton: we say T has a *topological discrete spectrum* (tds) if the the closed subspace of $C(X)$ spanned by eigen functions is $C(X)$.

Note that if T has tds and T is also transitive then there exist continuous functions $\{f_n : n=1,2,...\}$ such that $f_n\circ T=\lambda_n f_n, f_n's$ are linearly independent and the closed subspace of $C(X)$ spanned by eigen functions is $C(X)$.

Definition: Let T and S be homeomorphisms of compact metric spaces X and Y respectively. We say T is *topologically conjugate* to S if there exists a homeomorphism $\phi : X\to Y$ such that $\phi\circ T=S\circ\phi$.

Theorem [Halmos and von Neumann]

Let T be a homeomorphism of compact metric space X. TFE

1) T is topologically transitive and it is an isometry for some equivalent metric

2) T is conjugate to a minimal rotation on a compact abelian metric group

3) T is minimal and has topological discrete spectrum

4) T is topologically transitive and has topological discrete spectrum

Proof: 1) implies 2): our aim is to make X itself a compact abelian group so that T becomes a minimal rotation on it and the identity map is the required conjugacy map. Let d be an equivalent metric that makes T an isometry. Let $O_T(x_0)$ be dense. Define $*$ on $O_T(x_0)$ by $T^n(x_0)*T^m(x_0)=T^{n+m}(x_0)$. Since $d(T^n(x_0)*T^m(x_0),T^p(x_0)*T^q(x_0))=d(T^{n+m}(x_0),T^{p+q}(x_0))$

$\leq d(T^{n+m}(x_0),T^{p+m}(x_0))+d(T^{p+m}(x_0),T^{p+q}(x_0))=d(T^n(x_0),T^p(x_0))+d(T^m(x_0),T^q(x_0))$

the map $*: O_T(x_0) \times O_T(x_0) \to O_T(x_0)$ is uniformly continuous, so it extends to a continuous map : $X \times X \to X$. Also $d(T^{-n}x_0, T^{-m}x_0) = d(T^n x_0, T^m x_0)$ so the map $T^n(x_0) \to T^{-n}(x_0)$ is uniformly continuous and extends to X. Thus X becomes an abelian topological group (in which x_0 is the identity and the inverse of $T^n x_0$ is $T^{-n}x_0$). Note that $T(T^n(x_0)) = T(x_0) * T^n(x_0)$ so T acts as rotation by $T(x_0)$ on X. By an earlier theorem this rotation is minimal. Clearly T is conjugate to this rotation (via the identity map).

2) implies 3):

Let S be a minimal rotation on a compact abelian metric group G. We have to show that S has discrete spectrum. Each character γ is an eigen function. Linear span of characters is an algebra which contains constants and separate points. Also the complex conjugate of a character is a character. By Stone-Weirstrass Theorem eigen functions span a sense subspace.

3) implies 4) is trivial.

4) implies 1): suppose $f_n \circ T = \lambda_n f_n, f_n \neq 0, |f_n| \equiv 1, f_n^\prime s$ linearly independent and span a dense subspace of $C(X)$. Let $D(x,y) = \sum_{n=1}^{\infty} \frac{|f_n(x) - f_n(y)|}{2^n}$. $D(Tx,Ty) = \sum_{n=1}^{\infty} \frac{|f_n(Tx) - f_n(Ty)|}{2^n} = \sum_{n=1}^{\infty} |\lambda_n| \frac{|f_n(x) - f_n(y)|}{2^n} = D(x,y)$. All that remains is to show that D induces the original topology of X. Since each f_n is continuous convergence in the original metric implies convergence in D. If $D(x_j, x) \to 0$ then $f_n(x_j) \to f_n(x)$ as $j \to \infty$ for each n. This implies $f(x_j) \to f(x)$ for any $f \in C(X)$ [because $f_n^\prime s$ span a dense subspace of $C(X)$]. This implies that $x_j \to x$ in the original metric. [Let $\epsilon > 0$. There is a continuous function $f: X \to [0,1]$ such that $f(x) = 0$ and $f(y) = 1$ if $y \in X \backslash B(x,\epsilon)$. Since $f(x_j) \neq 1$ for j sufficiently large we get $x_j \in B(x,\epsilon)$ for such j].

Theorem [Topological Discrete Spectrum Theorem]

Two minimal homeomorphisms of compact metric spaces are topologically conjugate iff they have the same eigen values.

We do not prove this theorem.

Theorem

Let T be uniquely ergodic and P be its unique invariant measure. Then T is minimal iff P has full support.

Proof: If T is minimal and $P(U) = 0$ for some non-empty open set U then $X = \bigcup_{n=-\infty}^{\infty} T^n(U)$ and this is a contradiction because $P(T^n(U)) = 0$ for all n.

Suppose P has full support. If T is not minimal then there is a proper closed set C such that $TC = C$. There is a probability measure Q on C which is invariant for the restriction of T to C. Let $Q_1(E) = Q(E \cap C)$ for all Borel sets E in X. Then $Q_1 \circ T^{-1} = Q_1$ and $Q_1 \neq P$ because $P(C^c) \neq 0 = Q_1(C^c)$. This contradicts the hypothesis.

Theorem

TFE for a homeomorphism T of a compact metric space X :

1) $\frac{1}{n}\sum_{k=0}^{n} f(T^k(x)) \to c$ uniformly for some constant c for each $f \in C(X)$

2) $\frac{1}{n}\sum_{k=0}^{n} f(T^k(x)) \to c$ pointwise for some constant c for each $f \in C(X)$

3) $\frac{1}{n}\sum_{k=0}^{n} f(T^k(x)) \to \int f dP$ for each x, for each $f \in C(X)$ for some invariant p.m. P

4) T is uniquely ergodic

Proof: 1) implies 2) is trivial.

2) implies 3): let $\Lambda(f) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} f(T^k(x))$. [The limit is independent of x by hypothesis]. Λ is linear and continuous on $C(X)$ and hence there is a complex measure P such that $\Lambda(f) = \int f dP$ for all $f \in C(X)$. Since Λ is positive, P is a positive measure and since $\Lambda(1) = 1$, P is a probability measure. Since $\Lambda(f \circ T) = \Lambda(f)$ for all $f \in C(X)$ we see that P is invariant. Since $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} f \circ T^k$ exists in L^1 [by Birkhoff's Ergodic Theorem] the constant c in 2) must be $\int f dP$.

3) implies 4): suppose Q is an invariant measure. By 3), $\frac{1}{n}\sum_{k=0}^{n} f(T^k(x)) \to \int f dP$ for each x and Dominated convergence theorem gives $\int f dQ = \int f dP$ for all f. Hence $Q = P$.

4) implies 1): this is Weyl's Theorem.

Theorem

A rotation $Tg = ag$ on a compact metric group G is uniquely ergodic iff it is minimal.

Proof: since Haar measure has full support the theorem previous to above theorem shows that unique ergodicity implies minimality. If T is minimal that $\{a^n : n \in \mathbb{N}\}$ is dense and, for any character $\gamma \neq 1$, we have $\gamma(a) \neq 1$ and $\frac{1}{n}\sum_{k=0}^{n-1} \gamma(T^k x) = \frac{1}{n}\sum_{k=0}^{n-1} \gamma^k(a)\gamma(x) = \frac{1}{n}\frac{\gamma^n(a)-1}{\gamma(a)-1}\gamma(x) \to 0$ for all x. Thus, if Q is an invariant measure then (since T is necessarily ergodic) $\int \gamma dQ = 0$ for all characters γ except 1. It follows that $\int \gamma dQ = \int \gamma dP$ (where P is the Haar measure). Note that this holds for $\gamma = 1$ also. Since characters span a dense subset of $C(X)$ we get $\int f dQ = \int f dP$ for all $f \in C(X)$ and hence $Q = P$. This completes the proof.

Remark: let G be a compact metric group and $T : G \to G$ be a continuous automorphism. Then the Haar measure P and the measure δ_1 are both invariant and they are unequal so T is not uniquely ergodic (unless $G = \{1\}$).

Theorem

An affine transformation on a compact metric group G is uniquely ergodic iff it is minimal.

Proof: we do not prove that 'if' part. This is available in "Minimal dynamical systems with quasi-discrete spectrum", by Hahn and W. Parry, Jour. Lond. Math Soc, Vol 40, pp 309-323, 1965

Now suppose $Tg = aAg$ is uniquely ergodic. Then the unique invariant measure is the Haar measure which has full support and hence T is minimal.

Remark: it has been shown that any i.m.p. ergodic transformation is isomorphic to a uniquely ergodic transformation. Thus, measure theoretically unique ergodicity is not a useful concept!

Theorem [Birkhoff Recurrence Theorem]

Let Ω be a compact metric space and T be a homeomorphism of Ω. Then there exists $\omega \in \Omega$ such that every neighbourhood of ω contains $T^n\omega$ for infinitely many n.

Proof: we know that there a minimal set , i.e. a non-empty closed set C such that $TC = C$ and $T : C \to C$ is minimal. Every neighbourhood of a point ω of C contains $T^n\omega$ for infinitely many n. [$\omega \in C = \{T^n\omega : n \in \mathbb{Z}\}^-$. If $\{T^n\omega : n \in \mathbb{Z}\}$ is a finite set then $T^k\omega = \omega$ for some positive integer k; in this case any neighbourhood of ω contains $T^{kj}\omega = \omega$ for all positive integers j. If $\{T^n\omega : n \in \mathbb{Z}\}$ is an infinite set it is clear that every neighbourhood of ω contains $T^n\omega$ for infinitely many n].

Isomorphisms and spectral invariants

Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be probability spaces and $T_i : \Omega_i \to \Omega_i$ be m.p. for $i = 1, 2$. We say T_1 is *isomorphic* to T_2 and write $T_1\tilde{\ }T_2$ if there exist sets $E_i \in \mathcal{F}_i, i = 1, 2$ such that $P_i(E_i) = 1$ $(i = 1, 2)$ and

1) $T_i(E_i) \subset E_i$ $(i = 1, 2)$

2) there exists a bijection $\phi : E_1 \to E_2$ such that ϕ and ϕ^{-1} are measurable and $\phi \circ T_1 = T_2 \circ \phi$.

Note that if we replace E_1 by $\bigcap_{n=-\infty}^{\infty} T_1^n E_1$ and E_2 by $\bigcap_{n=-\infty}^{\infty} T_2^n E_2$ then we have $T_i(E_i) = E_i$ $(i = 1, 2)$ instead of 1).

Note also that $T_1 \tilde{} T_2 \Rightarrow T_1^n \tilde{} T_2^n$ $\forall n$.

Measure algebras and *conjugacy*:

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let X be the set of all equivalence classes $[E]$ of sets E in $\mathcal{F}$ under the equivalence relation $E \tilde{} F$ if $P(E\Delta F) = 0$. X is a complete metric space under the metric $d([E], [F]) = P(E\Delta F)$. We can define set theoretic operations on X in the obvious way. For instance $[E] \cap [F] = [E \cap F]$. Countable unions, countable intersections and complements can be defined similarly. We can define $[P]([E]) = P(E)$. We call $(X, [P])$ a measure algebra. For each $[E] \in X$, I_E is a well defined element of L^p (for any p). If T is m.p. on $(\Omega, \mathcal{F}, P)$ then $T^{-1} : X \to X$ can defined by $T^{-1}([E]) = [T^{-1}(E)]$.

Definition: let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be measure spaces and X, Y be the corresponding metric spaces obtained by above construction. A map $\phi : Y \to X$ is an *isomorphism of measure algebras* if it is a bijection, preserves countable unions and complements and $[P_1]\phi([E]) = [P_2]([E])$ for all $E \in \mathcal{F}_2$.

Definition: T_1 and T_2 are *conjugate* if there is a measure algebra isomorphism $\phi : Y \to X$ such that $\phi \circ T_2^{-1} = T_1^{-1} \circ \phi$.

Note that isomorphism implies conjugacy. It can be shown that the converse is also true if the probability spaces involved are Lebesgue spaces, i.e. spaces isomorphic to $[0, 1]$ with Borel sigma field and Lebesgue measure together with a countable number of atoms.

Definition: With above notations T_1 and T_2 are *spectrally isomorphic* if there is an isometric isomorphism W of $L^2(P_2)$ onto $L^2(P_1)$ such that $W(f) \circ T_1 = W(f \circ T_2)$ for all $f \in L^2(P_2)$.

Theorem

If T_1 and T_2 are isomorphic then they are conjugate and if they are conjugate then they are spectrally isomorphic.

Proof: we only have to prove the second implication. Suppose there is a measure algebra isomorphism $\phi : Y \to X$ such that $\phi \circ T_2^{-1} = T_1^{-1} \circ \phi$. Define $W : L^2(P_2) \to L^2(P_1)$ by $WI_E = I_F$ where $\phi([E]) = [F]$. (Note that the indicators are not well defined pointwise but they are well defined as L^2 functions). W extends to an isometric isomorphism of $L^2(P_2)$ onto $L^2(P_1)$ and yields the desired spectral isomorphism.

Theorem

Let $W : L^2(P_2) \to L^2(P_1)$ be an isometric isomorphism. Suppose $W(L^\infty(P_2)) \subset L^\infty(P_1)$, $W^{-1}(L^\infty(P_1)) \subset L^\infty(P_2)$ and W is multiplicative on $L^\infty(P_2)$. Then there is an isomorphism ϕ of measure algebras with $WI_E = I_F$ where $\phi([E]) = [F]$.

Proof: we first observe that $(WI_E)^2 = WI_E^2 = WI_E$ so WI_E is an indicator. Hence we can define ϕ by $\phi([E]) = [F]$ where $WI_E = I_F$. It is easy to check that ϕ is a well defined bijection. Note that $[P_2]([E]) = P_2(E) = \int I_E \bar{I_E} dP_2 =< WI_E, WI_E >=< I_F, I_F >= P_1[F]$ or $[P_1]\phi([E]) = [P_2]([E])$. It remains to show that ϕ preserves complements and countable unions. The equation $I_E + I_{E^c} = 1$ gives $WI_E + WI_{E^c} = W1$. Since $Wf = W1Wf$ for all bounded measurable functions f we get $W1 = 1$. Thus $WI_E + WI_{E^c} = 1$ so ϕ preserves complements. [If $WI_E = I_F$ and $WI_{E^c} = I_H$ then $I_F + I_H = 1$ so $[H] = [F]^c$]. To prove that ϕ preserves finite unions we use the identity $I_{A\cup B} = I_A + I_B - I_A I_B$. Using the fact that $\bigcup_{k=1}^{n} A_k \uparrow \bigcup_{k=1}^{\infty} A_k$ we get $I_{\bigcup_{k=1}^{n} A_k} \to I_{\bigcup_{k=1}^{\infty} A_k}$ in L^2 and this shows that ϕ preserves countable unions. [Details are left to the reader].

Corollary

If W is an isometry such that $W(L^\infty(P_2)) \subset L^\infty(P_1)$, $W^{-1}(L^\infty(P_1)) \subset L^\infty(P_2)$ and W is multiplicative on $L^\infty(P_2)$ and if $U_{T_1} \circ W = W \circ U_{T_2}$ then T_1 and T_2 are conjugate.

A property of a m.p. transformation which is preserved by isomorphism/conjugacy/spectral isomorphism is called *isomorphism/conjugacy/spectral invariant*. Thus any spectral invariant is a conjugacy invariant and any conjugacy invariant is an isomorphism invariant.

Theorem

Ergodicity, weak and strong mixing are spectral invariants (hence also conjugacy and isomorphism invariants).

Proof: T is ergodic iff $\{f \in L^2 : f \circ T = f\}$ is one-dimensional. If $W(f) \circ T_1 = W(f \circ T_2)$ (as in the definition of spectral invariance) and T_1 is ergodic then $f \circ T_2 = f$ iff $W(f) \circ T_1 = W(f)$ iff $Wf \in \{g \in L^2 : g \circ T = g\}$ which shows that $\{f \in L^2 : f \circ T_2 = f\}$ is one dimensional.

T is weak mixing iff it is ergodic and 1 is the only eigen value of U_T. If $W(f) \circ T_1 = W(f \circ T_2)$ then λ is eigen value of T_2 and f is an eigen function corresponding to it iff λ is eigen value of T_2 and Wf is an eigen function corresponding to λ.

Suppose T_1 is strong mixing and $W(f) \circ T_1 = W(f \circ T_2)$ for some isometry W. We have to show that $< f \circ T_2^n, g > \to < f, 1 >< 1, g >$ for all f, g. Since this holds when f or g is a constant we may suppose $\int f = \int g = 0$. Since T_1 is ergodic and conjugate to T_2 it follows that T_2 is ergodic. W maps T_2 invariant functions to T_1 invariant functions, so it maps constants to constants and hence it maps the orthogonal complement of the space of constants to this space. Thus $\int Wf = 0 = \int Wg$. Now $< f \circ T_2^n, g > = < W(f \circ T_2^n), Wg > = < W(f) \circ T_1^n, Wg > \to 0$ since T_1 is strong mixing.

Examples

Rotation on S^1 by a root of unity and rotation by a non root of unity are not not spectrally isomorphic (hence not conjugate or isomorphic). This is because the second one is ergodic and the first one is not. Also the second map is not weak mixing. [$Tz = az$ has a an an eigen value and hence it does not have discrete spectrum]. Thus this map is not spectrally isomorphic to any weak mixing map.

Definition: an i.m.p. transformation $T : \Omega \to \Omega$ is said to have *countable Lebesgue spectrum* if $L^2(P)$ has an orthonormal basis of the type $\{1\} \cup \{U_T^n f_j : j = 1, 2, ..., n \in \mathbb{Z}\}$.

Example: let T be the $(1/2, 1/2)$ two sided shift on $\prod_{-\infty}^{\infty} \{-1, 1\}$. Let $g_0 = 1$ and $g_{n_1,n_2,...,n_k}(\{a_n\}) = a_{n_1} a_{n_2} ... a_{n_k}$ for $n_1 < n_2 < ... < n_k, k \geq 1$. We have $U_T g_{n_1,n_2,...,n_k}(\{a_n\}) = a_{n_1+1} a_{n_2+1} ... a_{n_k+1} = g_{n_1+1, n_2+1, ..., n_k+1}$. If we write the collection $\{g_{n_1,n_2,...,n_k}\}$ as a sequence $\{f_j\}$ the the functions $\{1\} \cup \{U_T^n f_j : j = 1, 2, ..., n \in \mathbb{Z}\}$ form an orthonormal basis for $L^2(P)$

Let T and S both have countable Lebesgue spectrum. Let $\{1\} \cup \{U_T^n f_j : j = 1, 2, ..., n \in \mathbb{Z}\}$ and $\{1\} \cup \{U_S^n g_j : j = 1, 2, ..., n \in \mathbb{Z}\}$ be the corresponding bases. There is an L^2 isometry W which maps $U_T^n f_j$ to $U_S^n g_j$ and such that $W(f) \circ T = W(f \circ S)$ and hence T and S are spectrally isomorphic.

Theorem

If T has countable Lebesgue spectrum then it is strong mixing.

Proof: let $\{1\} \cup \{U_T^n f_j : j = 1, 2, ..., n \in \mathbb{Z}\}$ be a basis. As $m \to \infty$, $< U_T^m U_T^n f_k, U_T^l f_j > \to < U_T^n f_k, 1 >< 1, U_T^l f_j > = 0$. Now $\{f : < U_T^m f, U_T^l f_j > \to < f, 1 >< 1, U_T^l f_j >$ as $m \to \infty\}$ is a closed subspace of L^2 which contains the basis $\{1\} \cup \{U_T^n f_j : j = 1, 2, ..., n \in \mathbb{Z}\}$. Hence $< U_T^m f, U_T^l f_j > \to < f, 1 >< 1, U_T^l f_j >$ as $p \to \infty$ for all f. Now $\{g : U_T^m f, g > \to < f, 1 >< 1, g >$ as $m \to \infty\}$ is a closed subspace which contains an orthonormal set so it is equal to L^2.

This proves the theorem.

We now consider m.p.t.'s with *pure point spectrum* (PPP) i.e. those for which there is an orthonormal basis of L^2 consisting of eigen functions of U_T.

The following elementary facts about eigen functions may appear to be repetitive but we shall go through them nonetheless. Let T be m.p. and ergodic. By eigen values/eigen functions of T we mean those of U_T in L^2. If λ is an eigen function with eigen value λ then $|\lambda| = 1$ and $|f|$ is a constant: the first property follows by taking L^2 norms on both sides of $f \circ T = \lambda f$ and the second property follows from the facts that $|f|$ is invariant and T is ergodic. Next we note that eigen functions corresponding to different eigen values are orthogonal: $f \circ T =$

$\lambda_1 f, g \circ T = \lambda_2 g, \lambda_1 \neq \lambda_2, f \neq 0, g \neq 0$ imply that $\int [f \circ T][g \circ T]^- =< f, g >$ (because T is m.p.) and $\int [f \circ T][g \circ T]^- = \lambda_1 [\lambda_2]^- < f, g >= \frac{\lambda_1}{\lambda_2} < f, g >$. Also eigen spaces are one dimensional: if $f \circ T = \lambda f, g \circ T = \lambda g, f \neq 0, g \neq 0$ then $\frac{f}{g}$ is invariant, hence constant. Finally we observe that eigen values form a subgroup of S^1.

Theorem [Discrete Spectrum Theorem due to Halmos and Von Neumann]

Let T_i be an ergodic m.p.t. on $(\Omega_i \mathcal{F}_i, P_i), i = 1, 2$ and assume that both these have PPP. Then the following are equivalent:

1) T_1 and T_2 are spectrally isomorphic

2) T_1 and T_2 have same eigen values

3) T_1 and T_2 are conjugate

Proof: 1) implies 2) is trivial. 3) implies 1) by definition. 2) implies 1) is straightforward: we get orthonormal bases index by the common eigen values and this gives an isometric isomorphism of L^2 which is a spectral isomorphism. 2) implies 3) requires the following algebraic result proved earlier:

Lemma

Let H be an abelian group and K a subgroup of H such that $k \in K, n \in \mathbb{N} \Rightarrow k = g^n$ for some $g \in K$. Then there exists a homomorphism $\phi : H \to K$ such that ϕ is the identity on K.

[This is purely algebraic. We are asserting that K is an algebraic retract of H].

Proof of the lemma: let $\mathcal{R} = \{(M, \phi) : K \leq M \leq H, \phi : M \to K$ is a homomorphism with $\phi =$ identity on $K\}$. ($G_1 \leq G_2$ means G_1 is a subgroup of G_2). This contains (M, ϕ) if $M = K$ and ϕ is the identity. Order this class by saying $(M_1, \phi_1) \leq (M_2, \phi_2)$ if $M_1 \leq M_2$ and $\phi_2 | M_1 = \phi_1$. It is clear that any totally ordered subfamily of $\mathcal{R}$ has an upper bound. By Zorn's Lemma there is a maximal element (M_0, ϕ_0). Claim: $M_0 = H$. Suppose $g \in H \backslash M_0$. Let M be the group generated by M_0 and g. We consider two cases:

Case 1: $g^n \notin M_0$ for any integer n. In this case $M = \{g^n h : n \in \mathbb{Z}, h \in M_0\}$ and the representation of elements of M in the form $g^n h$ is unique. Let $\psi(g^n a) = \phi_0(a)$. This gives an element (M, ψ) strictly larger than (M_0, ϕ_0) which is a contradiction.

Case 2: there is a least positive integer N such that $g^N \in M_0$. Each element of the group M is uniquely expressible as $g^n a$ where $a \in M_0$ and $0 \leq n < N$. There exists $h \in K$ such that $\phi_0(g^N) = h^N$. We define $\psi(g^n a) = h^n \phi_0(a)$. Once again this gives an element (M, ψ) strictly larger than (M_0, ϕ_0) which is a contradiction.

we now prove 2) implies 3). Let Φ be the group formed by the common eigen values of T_1 and T_2. Let $\{f_\lambda : \lambda \in \Phi\}$ and $\{g_\lambda : \lambda \in \Phi\}$ be orthonormal bases of eigen functions. We may and do assume that $|f_\lambda| = 1$ and $|g_\lambda| = 1$ everywhere. We have $U_{T_i}(f_\lambda f_\mu) = \lambda\mu f_\lambda f_\mu$ and $U_{T_i}(f_{\lambda\mu}) = \lambda\mu f_{\lambda\mu}$. Since eigen spaces are one dimensional we see that $f_{\lambda\mu} = c(\lambda,\mu) f_\lambda f_\mu$ for some $c(\lambda,\mu) \in S^1$. We use the lemma above to reduce the proof to the case when $c(\lambda,\mu) = 1$ for all λ and μ. Let H be the the product $(S^1)^{\Omega_1}$ (the collection of all functions from $\tau : \Omega_1 \to S^1$) under pointwise multiplication and G be the subgroup of constant functions. Denote by h_a the constant function $h_a(\omega) = a\ \forall \omega$. The hypothesis of above lemma is satisfied, so we get a homomorphism $\tau : H \to G$ such that τ is the identity on G. If $F_\lambda = [\tau(f_\lambda)]^- f_\lambda$ then one can easily check that $F_{\lambda\mu} = F_\lambda F_\mu$. Hence we assume hence forth that $c(\lambda,\mu) = 1$ for all λ and μ and $f_{\lambda\mu} = f_\lambda f_\mu$. Similarly we may suppose $g_{\lambda\mu} = g_\lambda g_\mu$. There is an isometric isomorphism W from $L^2(\Omega_2)$ onto $L^2(\Omega_1)$ such that $W g_\lambda = f_\lambda$ for all λ. Note that $W(gh) = (Wg)(Wh)$ for all $g, h \in N$, the vector space spanned by the functions $\{g_\lambda : \lambda \in \Lambda\}$ L^∞ If we prove that this equation holds for all L^∞ functions f and g we can obtain an isomorphism of measure algebras from W as described earlier and we can then conclude that T_1 and T_2 are conjugate. Let M the vector space spanned by the functions f_λ. If $g \in L^\infty$ then there is a sequence $\{g_n\} \subset N$ such that $g_n \to g$ in L^2. It follows that $g_n g_\lambda \to g g_\lambda$ in L^2 and since $W(g_n g_\lambda) = W(g_n) W(g_\lambda)$ for each n we get $W(g g_\lambda) = W(g) W(g_\lambda)$. [We used the fact that convergence in L^2 implies a.e. convergence for a subsequence]. In particular we have proved that $W(g)W(g_\lambda) (= W(g g_\lambda)) \in L^2$ for all $g \in L^\infty$ for all λ. Now let $g \in L^\infty$ and $h \in L^\infty$. There exists $\{h_n\} \subset N$ such that $h_n \to h$ in L^2. Now $\int |W(g)W(h)|^2 \leq \liminf \int |W(g)W(h_n)|^2 = \liminf \int |W(gh_n)|^2 = \int |W(gh)|^2 < \infty$. Thus $W(g)W(h) \in L^2$ for all $g, h \in L^\infty$. Next we take $\xi \in L^2$ and choose a sequence $\{g_n\} \subset N$ such that $g_n \to \xi$ in L^2. We get $\int |W(\xi)W(g)|^2 \leq \liminf \int |W(g_n)W(g)|^2 = \liminf \int |W(g_n g)|^2 = \int |W(\xi g)|^2 < \infty$. It follows that $W(\xi)W(g) \in L^2$ for all $\xi \in L^2$. This implies (by a standard argument using Uniform Boundedness Principle) that $W(g) \in L^\infty$. Thus W maps L^∞ into itself. If $g, h \in L^\infty$ choose $\{g_n\} \subset N$ and $\{h_n\} \subset N$ such that $g_n \to g$ and $h_n \to h$ in L^2. Since $W(g_n h_m) = W(g_n)W(h_m), g_n \to g$ in $L^2, W(g_n)W(h_m) \to W(g)W(h_m)$ in L^2 we get $W(g h_m) = W(g)W(h_m)$. Now $g h_m \to gh$ in L^2 as $m \to \infty$ and $W(g)W(h_m) \to W(g)W(h)$ in L^2 in view of the fact that $W(g) \in L^\infty$. Hence $W(gh) = W(g)W(h)$. This finishes the proof.

Theorem

Any ergodic rotation $T(g) = ag$ on a compact abelian metric group has pure point spectrum. Eigen functions of T are constant multiples of characters and the set of eigen values coincides with $\{\gamma(a) : \gamma \in \hat{G}\}$.

Proof: we have $\gamma(Tg) = \gamma(ag) = \gamma(a)\gamma(g)$ so $\gamma(a)$ is an eigen value with eigen function γ. Since characters span a sense subspace of L^2 it follows that

T has pure point spectrum. If $f \circ T = \lambda f, f \neq 0, \lambda \notin \{\gamma(a) : \gamma \in \hat{G}\}$ then f is orthogonal to each character, hence to L^2. If $\lambda = \gamma(a)$ for some γ then f is a multiple of γ because (by ergodicity) the eigen space corresponding to eigen value $\gamma(a)$ is one dimensional.

Theorem [Representation Theorem]

An ergodic m.p. transformation with pure point spectrum is conjugate to an ergodic rotation on a compact abeliian group.

Proof: let Λ be the group of eigen values of the given transformation. Give Λ the discrete topology. Let G be the dual group. Then G is a compact abelian group under pointwise convergence topology. [Think of G as a subset of $(S^1)^\Lambda$. Pointwise convergence topology is the product topology relativized to G. A straightforward argument show that G is closed in $(S^1)^\Lambda$. By Tychonoff's Theorem G is compact]. The inclusion map $s : \Lambda \to S^1$ is a character and hence it belongs to G. Define $V : G \to G$ by $V(g) = sg$. V is thus a rotation on G. Suppose f is an invariant function for V. Let $f = \sum < f, \gamma_j > \gamma_j$ be the Fourier series of f. Then $< f, \gamma_j >= [\gamma(s)]^- < f, \gamma_j >$ so $< f, \gamma_j >= 0$ unless $\gamma_j(s) = 1$. However $\gamma_j(s)$ is an eigen value and $< f, \gamma_j >= 0$ except when this eigen value is 1. This shows f is a constant and hence that V is ergodic. By previous theorem V has pure point spectrum. T and S have the same eigen values and hence they are conjugate by • Neumann Theorem.

Corollary

Any subgroup of S^1 is the group of eigen values of an ergodic rotation on a compact abelian group with pure point spectrum.

[Such a rotation was constructed in the proof above].

TOPOLOGICAL ENTROPY

We write log for logarithm to base 2. Let X be a compact Hausdorff space.

Definition: the entropy $h(\mathcal{U})$ of an open cover $\mathcal{U}$ of X is defined by $h(\mathcal{U}) = \log N(\mathcal{U})$ where $N(\mathcal{U})$ is the smallest number of sets from the cover $\mathcal{U}$ required to cover X.

In the following we write $\mathcal{U}_1 \sqcup \mathcal{U}_2$ for the open cover consisting of sets of the type $A \cap B$ with $A \in \mathcal{U}_1$, $B \in \mathcal{U}_2$. We say that $\mathcal{U}_2$ is a refinement of $\mathcal{U}_1$ if every set in $\mathcal{U}_2$ is a subset of some set in $\mathcal{U}_1$. We write $\mathcal{U}_1 \leq \mathcal{U}_2$ in this case.

Note that $h(\mathcal{U}) = 0$ iff $X \in \mathcal{U}$. Also $h(\mathcal{U}_1) \leq h(\mathcal{U}_2)$ if $\mathcal{U}_1 \leq \mathcal{U}_2$. We claim that $h(\mathcal{U}_1 \sqcup \mathcal{U}_2) \leq h(\mathcal{U}_1) + h(\mathcal{U}_2)$. This follows from the simple fact that if $\{A_i : 1 \leq i \leq n\}$ and $\{B_i : 1 \leq i \leq m\}$ both cover X then so does the collection $\{A_i \cap B_j : 1 \leq i \leq n, 1 \leq j \leq m\}$ which has nm elements.

If $T : X \to X$ is continuous then $h(T^{-1}\mathcal{U}) \leq h(\mathcal{U})$ where $T^{-1}\mathcal{U} = \{\mathcal{T}^{-1}A : A \in \mathcal{U}\}$. If T is continuous and its range is all of X then $h(T^{-1}\mathcal{U}) = h(\mathcal{U})$.

Theorem

If $\mathcal{U}$ is an open cover and $T : X \to X$ is continuous then $\lim_{n\to\infty} \frac{1}{n} h(\mathcal{U} \sqcup T^{-1}\mathcal{U} \sqcup ... \sqcup T^{-(n-1)}\mathcal{U})$ exists.

Proof: let $a_n = h(\mathcal{U} \sqcup T^{-1}\mathcal{U} \sqcup ... \sqcup T^{-(n-1)}\mathcal{U})$. Then $a_{n+m} \leq a_n + a_m$. This implies $\lim_{n\to\infty} \frac{1}{n} a_n$ exists.

Definition: $\lim_{n\to\infty} \frac{1}{n} h(\mathcal{U} \sqcup T^{-1}\mathcal{U} \sqcup ... \sqcup T^{-(n-1)}\mathcal{U})$ is called *the entropy of T relative to $\mathcal{U}$* and is denoted by $h(T,\mathcal{U})$.

Note that $h(T,\mathcal{U}_1) \leq h(T,\mathcal{U}_2)$ if $\mathcal{U}_1 \leq \mathcal{U}_2$. In particular this inequality holds if $\mathcal{U}_2$ is a subcover of $\mathcal{U}_1$.

Now we note that since $h(\mathcal{U} \sqcup T^{-1}\mathcal{U} \sqcup ... \sqcup T^{-(n-1)}\mathcal{U}) \leq nh(\mathcal{U})$ we have $h(T,\mathcal{U}) \leq h(\mathcal{U})$.

Definition: the topological entropy $h(T)$ of a continuous map $T : X \to X$ is defined by

$h(T) = \sup\{h(T,\mathcal{U}) : \mathcal{U}$ is an open cover of $X\}$.

Note that $h(T) = 0$ if T is the identity map of X.

Claim: $h(T) = \sup\{h(T,\mathcal{U}) : \mathcal{U}$ is a finite open cover of $X\}$.

This follows from the fact that every open cover $\mathcal{U}$ has a finite subcover $\mathcal{U}_1$ and $h(T,\mathcal{U}) \leq h(T,\mathcal{U}_1)$.

Theorem

Let $T_i : X_i \to X_i$ be continuous maps for $i = 1, 2$ (where X_1, X_2 are compact Hausdorff spaces). If T_1 and T_2 are topologically conjugate then they have the same entropy.

This is straightforward. We omit the details.

Another result proved by a routine argument is that the entropies of T and T^{-1} are the same.

We now state a few theorems on topological entropy and its relation to entropy w.r.t. measures. We do not prove these theorem here. References for the original articles containing the proofs of these results are available in Peter Walter's Ergodic Theory.

Theorem [L. W. Goodwyn]

The entropy of a homeomorphism w.r.t. a an invariant measure P does not exceed the topological entropy.

Theorem [T.N.T. Goodman]

The topological entropy of T is the supremum of entropies w.r.t. all invariant measures.

Theorem [K. Berg]

The topological entropy of an automorphism of a compact metric group coincides with the entropy w.r.t. Haar measure.

Theorem [Bowen]

If $\{T_t : t \in \mathbb{R}\}$ is a group of homeomorphisms of a compact metric space X then $h(T_t) = |t|\, h(T_1)$.

KAKUTANI TOWERS AND ROKHLIN's LEMMA

Let $(\Omega, \mathcal{F}, P, T)$ be a dynamical system. Assume that T is i.m.p.. If $\{B, T(B), ..., T^{N-1}(B)\}$ are disjoint we call this collection a column with base B and height N. We call $T^{N-1}(B)$ the roof of the column and $B \cup T(B) \cup ... \cup T^{N-1}(B)$ the carrier of the column.

A (Kakutani) tower is a countable collection of disjoint towers. Let $\mathcal{T} = \{ \mathcal{C}_1, \mathcal{C}_2, ...\}$ where $\mathcal{C}_k = \{B_k, T(B_k), ..., T^{N_k-1}(B_k)\}, k = 12, ...$are disjoint columns. The set $U_{\mathcal{T}} = \bigcup_k U_{\mathcal{C}_k}$ where $U_{\mathcal{C}_k} = B_k \cup T(B_k) \cup ... \cup T^{N_k-1}(B_k)$ is the carrier of the tower $\mathcal{T}$. Its base is the union of the bases of its columns,viz. $B = B_1 \cup B_2 \cup .....$ Its roof is $T^{N_1-1}(B_1) \cup T^{N_2-1}(B_2) \cup .....$ A set of the type $\{T^k(x) : 0 \leq k \leq N_k(x)-1\}$ where $N_k(x)$ is the height of the column containing x is called a fibre.

Theorem [Rokhlin's Lemma]

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be i.m.p. and ergodic. Assume that P is non-atomic. Given $N \in \mathbb{N}$ and $\epsilon > 0$ there exists a measurable set B such that $B, T(B), ..., T^{N-1}(B)$ are disjoint and carrier of this column has measure $> 1 - \epsilon$.

Proof: let C be any measurable set with $0 < P(C) < \frac{\epsilon}{N}$. [This is the only place where non-atomicity of P is used. We remark that this property may hold even in purely atomic spaces; for example $P\{n\} = \frac{1}{2^n}$ in $\mathbb{N}$]. Let $\tau_C(x) = \min\{n \geq 0 : T^n x \in C\}$. This is finite a.e. on C by Poincare's recurrence Theorem. Let $B_k = \{x \in C : \tau_C(x) = k\}, k = 1, 2, ...$ and let $\mathcal{C}_k = \{B_k, T(B_k), ..., T^{k-1}(B_k)\}$. Display these sets as a column of disjoint sets. [The sets in these columns are disjoint: if $x \in T^i(B_k) \cap T^j(B_k)$ with $0 \leq i < j < k$ then $T^{-i}x$ and $T^{-j}x$ both belong to B_k which implies $\tau_C(T^{-i}x) = \tau_C(T^{-j}x) = k$. This is a contradiction because $\tau_C(T^{-i}x) = i - j + \tau_C(T^{-j}x)$]. This proves that $\mathcal{C}_k$ is a column. Suppose $x \in T^i B_k \cap T^j(B_m)$ with $0 \leq i < k$ and $0 \leq j < m$. Suppose $i < j$. Let $u = T^{-i}x$ and $v = T^{-j}x$ so that $u \in B_k \subseteq C, v \in B_m$. Since $T^{j-i}v = u \in C$ and $v \in B_m$ we must have $m \leq j - i$. This is a contradiction because $j - i \leq j < m$. Similarly if $j < i$ and $z = T^{-j}x$ we can see that $z \in C$ and hence $T^{i-j}(T^{-i}x) \in C$ which implies $k \leq i - j \leq i < k$ a contradiction. Thus, we must have $i = j$. But then

$T^{-i}x \in B_k \cap B_m = \emptyset$ unless $k = m$. Thus we have constructed a Kakutani tower.

Now we divide $\mathcal{C}_k$ into blocks of size N starting from the base. [There will be some sets, not exceeding $(N-1)$) which are not included in these blocks. Pick the first set in each of these blocks and let B be the union of these sets. Clearly $B, TB, ..., T^{N-1}(B)$ are disjoint. Note that T maps the union Ω_0 of all the sets in the tower into itself and hence Ω_0 has measure is 1.[If $x \in T^i(B_k)$ with $i < k-1$ then $Tx \in T^{i+1}(B_k) \subset \Omega_0$ and if $x \in T^{k-1}(B_k)$ then $Tx \in B = \bigcup_k B_k \subset \Omega_0$. Thus $T\Omega_0 \subset \Omega_0$ and this implies $P(\Omega_0) = 1$]. Now the union of N $B, TB, ..., T^{N-1}(B)$ covers all of this set of measure 1 except for the sets are outside the blocks we have marked in each column. Since T is m.p. the sets in the columns have the same measure and so the the leftover part of each column has measure not exceeding $NP(B_k)$. Thus $P\{(B \cup TB \cup ... \cup T^{N-1}(B))^c\} < N \sum_k P(B_k) = NP(C) < \epsilon$. This completes the proof.

Ergodic theorem in a Banach space

Theorem

Let X be a Banach space and $T : X \to X$ be a bounded operator. Let $S_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k$. Assume that

1) $\sup_n \|S_n\| < \infty$

2) $\frac{1}{n} \|T^n\| \to 0$ as $n \to \infty$

Then $\{S_n x\}$ converges in the norm of X as $n \to \infty$ for all x such that $\{S_n x\}$ has a weakly convergent subsequence.

Proof: clearly, $(I-T)S_n = S_n(I-T) = \frac{1}{n}(I - T^n)$. By 2) $\|(I-T)S_n\| = \|S_n(I-T)\| \to 0$.

We have $I - S_n = \frac{1}{n} \sum_{k=0}^{n-1} (I - T^k) = (I-T)\frac{1}{n} \sum_{k=0}^{n-1} kS_k$.

Let $S_{n_j}x \to y$ weakly. We claim that $Ty = y$. We have $TS_{n_j}x \to Ty$ weakly and $(I-T)S_{n_j}x \to y - Ty$ weakly because T is weak-weak continuous. Thus $|x^*(y) - x^*(Ty)| = \lim \left|x^*(I-T)S_{n_j}x\right| \leq \limsup \|x^*\| \|x\| \left\|(I-T)S_{n_j}\right\| = 0$. Since x^* is arbitrary we get $y = Ty$. Let $z = x - y$. Then $S_n z = S_n x - S_n y = S_n x - y$ so $S_{n_j} z \to y - y = 0$ weakly. Now $z - S_{n_j} z = (I-T)\frac{1}{n_j} \sum_{k=0}^{n_j - 1} kS_k \in (I-T)(X)$. Thus z belongs to the weak closure of the range of $(I-T)$. Let $\epsilon > 0$. Since the weak closure coincides with the original closure we can find $u \in X$ such that $\|z - (I-T)u\| < \epsilon$. Now $\|S_n(I-T)u\| \to 0$ and $\|S_n z\| \leq \|S_n(z - (I-T)u)\| + \|S_n(I-T)u\| \leq C\epsilon + \|S_n(I-T)u\|$ where $C = \sup_n \|S_n\|$.

This proves that $\|S_n z\| \to 0$. Thus $\|S_n x - S_n y\| \to 0$ which means $\|S_n x - y\| \to 0$, i.e. $S_n x \to y$ in the norm.

Remarks: the condition that $\{S_n x\}$ has a weakly convergent subsequence is satisfied automatically if X is separable and reflexive (by Banach Alaoglu Theorem). In particular this holds if X is a separable Hilbert space.

Let $(\Omega, \mathcal{F}, P, T)$ be a DS with $\mathcal{F}$ countably generated. Then $L^2(P)$ is a separable Hilbert space. Define $U : L^2 \to L^2$ by $U(f) = f \circ T$. Then the hypothesis of above theorem are all satisfied and we can conclude that $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=0}^{n-1} f \circ T^k$ exists in the norm of L^2 for each $f \in L^2$. Hence we can view above theorem as a generalization of von Neumann's Ergodic Theorem to Banach spaces.

When $X = L^1(\Omega, \mathcal{F}, P)$ we have the following theorem:

Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X = L^1(\Omega, \mathcal{F}, P)$. Let $T : X \to X$ be a bounded operator such that the operators $S_n = \frac{1}{n} \sum\limits_{k=0}^{n-1} T^k, n = 1, 2, ...$ map L^∞ into itself. Suppose the norms of these operators are bounded both as operators on L^1 and as operators on L^∞. Then $\{S_n f\}$ converges in the norm of X for every $f \in X$ if and only if $\frac{\|T^n f\|_1}{n} \to 0$ as $n \to \infty$ for every $f \in X$.

We omit the lengthy proof of this theorem. [See "Linear Operators" by Dunford and Schwartz, Part I, p 662 (Corollary 5) for a proof].

Suppose T is a measurable map on $(\Omega, \mathcal{F}, P)$ but T is not measure preserving. When does the map $f \to f \circ T$ map L^1 into L^1 and when is this linear map bounded? We have the following:

Theorem

Let $1 \leq p < \infty$. Then $f \to f \circ T$ is a bounded operator on L^p if and only if there exists $C \in (0, \infty)$ such that $P(T^{-1}(E)) \leq CP(E)$ for all meaurable sets E and $f \circ T \in L^p$ for all $f \in L^p$.

Proof: suppose there exists $C \in (0, \infty)$ such that $P(T^{-1}(E)) \leq CP(E)$ for all E and $f \circ T \in L^p$ for all $f \in L^p$.. We first observe that $f = g$ a.e. implies $f \circ T = g \circ T$ a.e. [If $f(\omega) = g(\omega)$ for $\omega \in A$ where $P(A^c) = 0$ then $P(T^{-1}(A^c)) \leq CP(A^c) = 0$ and $f(T(\omega)) = g(T(\omega))$ if $\omega \in T^{-1}(A)$]. Next we show that $f \to f \circ T$ is bounded. For this we apply the Closed Graph Theorem. If $f_n \to f$ in L^p and $f_n \circ T \to g$ in L^p then we can find integers $n_k \uparrow \infty$ such that $f_{n_k} \to f$ a.e. and $f_{n_k} \circ T \to g$ a.e. If $f_{n_k}(\omega) \to f(\omega)$ for $\omega \in B$ with $P(B^c) = 0$ then $P(T^{-1}(B^c)) = 0$ and $f_{n_k}(T(\omega)) \to f(T(\omega))$ for $\omega \in T^{-1}(B)$. Thus $f_{n_k} \circ T \to f \circ T$ a.e. and $f_{n_k} \circ T \to g$ a.e. implying that $g = f \circ T$ a.e. This proves that $f \to f \circ T$ is a bounded operator on L^p. Conversely, suppose

$f \to f \circ T$ is a bounded operator on L^p. Then $\displaystyle\int (I_E \circ T)^p dP \leq C \int I_E dP$ where C is the norm of the operator $f \to f \circ T$. Thus $P(T^{-1}(E)) \leq CP(E)$.

Remark: if the map $f \to f \circ T$ is a bounded operator on L^p for some p then it is so on every L^q; it maps positive functions to positive functions and L^∞ functions to L^∞ functions. In "Ergodic Theory of Markov Processes" by Shaul Foguel the basic object of study is a bounded operator on L^1 which is positive and has norm ≤ 1. This book also contains theorems on existence of invariant measures proved using properties of such operators. A number of conditions equivalent to the existence of an 'invariant measure' for such an operator which is equivalent to P are given.

Theorem [Pointwise Ergodic Theorem in L^1]

Suppose U is an operator of norm at most 1 on $L^1(= L^1(\Omega, \mathcal{F}, P))$ which also acts as an operator of norm at most 1 on L^∞. Then $\lim \frac{1}{n} \displaystyle\sum_{k=0}^{n-1} U^k f$ exists a.e. for every $f \in L^1$.

Ref. Theorem 6, page 675 of Dunford and Schwartz, Part I.

APPENDIX

Existence and uniqueness of Haar measure

[Ref.: Measure Theory by Cohn]

Throughout G is a locally compact Hausdorff topological group. Our interest is mainly in compact metric groups, but we prove the existence theorem in the case of locally compact groups.

Theorem

If $f : G \to \mathbb{C}$ is continuous and has compact support then f is left and right uniformly continuous.

The proof is left as an exercise.

Theorem

Let μ be a regular Borel measure on G. If $f : G \to \mathbb{C}$ is continuous and has compact support then $x \to \displaystyle\int f(xy) d\mu(y)$ and $x \to \displaystyle\int f(yx) d\mu(y)$ are continuous.

Proof: exercise.

Theorem

Any open subgroup H of G is closed.

Proof: we have $G\backslash H = \bigcup_{x \notin H} (xH)$ which is open.

Theorem

There exists an open (hence closed) subgroup of G which is $\sigma-$ compact.

Proof: there exists U open such that $e \in U$ and $\bar{U}$ is compact. There exists a symmetric open set V such that $e \in V \subset \bar{V} \subset U$. Let $V_1 = V, V_2 = VV, ..., V_{n+1} = V^n V, ...$ Let $E = \bigcup_{n=1}^{\infty} V_n$. Clearly E is an open subgroup of G. It follows that E is a also closed. Note that $\bar{V_n}$ is compact and $\bar{V_n} \subset E$. Thus E is $\sigma-$ compact.

We use the following notations: ${}_x f(y) = f(x^{-1}y). f_x(y) = f(yx^{-1})$.

Theorem

Haar measure on G exists.

Proof: let $K \subset G$ be compact and $A \subset G$ have non-empty interior A^0. The $K \subset \bigcup_x xA^0$ and so there is a finite set $\{x_1, x_2, ..., x_n\}$ with $K \subset \bigcup_{i=1}^n x_i A^0$. Let $N(K, A)$ be the smallest integer n for which such a finite set exists. [Set $N(K, A) = 0$ if $K = \emptyset$]. Let K_0 be a fixed compact set with non-empty interior. For each open set U containing e and each compact set K define $\phi_U(K) = \frac{N(K,U)}{N(K_0,U)}$. We note that $0 \leq \phi_U(K) \leq N(K, K_0), \phi_U(K_0) = 1, \phi_U(K) \leq \phi_U(K')$ if $K \subset K', \phi_U(K \cup K') \leq \phi_U(K) + \phi_U(K')$ and equality holds in this last inequality if $KU^{-1} \cap K'U^{-1} = \emptyset$. [For the last part let $K \cup K' \subset \bigcup_{i=1}^n x_i U$ where $n = \phi_U(K \cup K')$. Assume that $x_i U$ intersects $K \cup K'$ for each i. Since , $KU^{-1} \cap K'U^{-1} = \emptyset$, for each i either $x_i U \cap K = \emptyset$ or $x_i U \cap K' = \emptyset$. Thus $\{x_1, x_2, .., x_n\} = \{x_i : x_i U \cap K = \emptyset\} \cup \{x_i : x_i U \cap K' = \emptyset\}$. Note that $K \subset \bigcup_J x_i U$ and $K' \subset \bigcup_I x_i U$ where $I = \{x_i : x_i U \cap K = \emptyset\}, J = \{x_i : x_i U \cap K' = \emptyset\}$. Since I and J are disjoint we get $N(K, U) + N(K', U) \leq \#(I) + \#(J) = N(K \cup K', U)$]. This proves that $\phi_U(K \cup K') = \phi_U(K) + \phi_U(K')$ if $KU^{-1} \cap K'U^{-1} = \emptyset$. Let $I_K = [0, N(K, K_0)]$ and $X = \prod I_K$ the product taken over all compact sets K. With the product topology X is compact. Note that $\phi_U \in X$ for every open set U containing e. If V is open and $e \in V$ let $S_V = \{\phi_U : U \subset V, U$ open, $e \in U\}^-$.We claim that $\bigcap S_V$, where the intersection is over all open sets V containing e, is non-empty. Since each S_V is non-empty and compact we only have to verify finite intersection property. If $V_1, V_2, .., V_m$ are open sets containing e then $\phi_{V_1 \cap V_2 \cap ... \cap V_m}$ belongs to each of the sets $S_{Vi}, 1 \leq i \leq m$. We have proved the claim. Let $\phi \in \bigcap S_V$. We claim that for any compact sets K, K' and any $x \in G$ we have: $\phi(K) \geq 0, \phi(\emptyset) = 0, \phi(K_0) = 1, \phi(xK) = \phi(K), \phi(K) \leq \phi(K')$ if $K \subset K', \phi(K \cup K') \leq \phi(K) + \phi(K')$ with equality if $K \cap K' = \emptyset$. After proving this claim we define $\mu^*(U) = \sup\{\phi(K) : K \subset U, K$ compact$\}$ for U open, $\mu^*(A) = \inf\{\mu^*(U) : A \subset U, U$ open$\}$; we show that μ^* is an outer measure on the power set of G, that each Borel set is μ^* measurable and that the restriction of μ^* to the Borel sigma field is a left Haar measure on G.

Proof of the claim: we only have to prove that $\phi(K_1 \cup K_2) \leq \phi(K_1) + \phi(K_2)$ with equality if $K_1 \cap K_2 = \emptyset$. The map $\theta(h) = h(K_1) + h(K_2) - h(K_1 \cup K_2)$ is continuous on X. Since $\phi_U(K_1) + \phi_U(K_2) - \phi_U(K_1 \cup K_2) \geq 0$ it follows that θ is non-negative at each point of S_V for any open set V containing e. Hence $\theta(\phi) \geq 0$ which proves the first part of the claim. Now let $K_1 \cap K_2 = \emptyset$. There exist disjoint open sets U_1, U_2 such that $K_1 \subset U_1$ and $K_2 \subset U_2$. [This separation result holds in any Hausdorff space!]. We can find open sets V_1, V_2 containing e such that $K_1 V_1 \subset U_1$ and $K_2 V_2 \subset U_2$. [If $x \in K_1$ there exists neighbourhoods S_x, T_x of e such that $xS_x \subset U_1$ and $T_x T_x \subset S_x$. By compactness $K_1 \subset x_1 T_{x_1} \cup x_2 T_{x_2} \cup ... \cup x_n T_{x_n}$ for some finite subset $\{x_1, x_2, ..., x_n\}$ of K_1. Let $V_1 = T_{x_1} \cap T_{x_2} \cap ... \cap T_{x_n}$. Any point of $K_1 V_1$ belongs to $x_i T_{x_i} T_{x_i}$ for some i and $x_i T_{x_i} T_{x_i} \subset x_i S_{x_i} \subset U_1$]. Let $V = V_1 \cap V_2$. Since $K_1 V \cap K_2 V \subset K_1 V_1 \cap K_2 V_2 \subset U_1 \cap U_2 = \emptyset$ we get $\phi_U(K_1 \cup K_2) = \phi_U(K_1) + \phi_U(K_2)$ whenever U is open, contains e and is contained in V^{-1}. [Because $K_1 U^{-1} \cap K_2 U^{-1} = \emptyset$]. The map θ above is 0 on points of the type $h = \phi_U$. Hence it is 0 on each S_V and so it is 0 at ϕ. This proves the claim.

Now we study μ^* defined by $\mu^*(U) = \sup\{\phi(K) : K \subset U, K$ compact$\}$ for U open, $\mu^*(A) = \inf\{\mu^*(U) : A \subset U, U$ open$\}$. If U_n is open for each n then $\mu^*(\bigcup_n U_n) \leq \sum_n \mu^*(U_n)$. To see this take any compact set $K \subset \bigcup_n U_n$. We have $K \subset \bigcup_{n=1}^N U_n$ for some N and there exist compact sets $K_1, K_2, ..., K_N$ with $K_i \subset U_i$ for each i and $K \subset \bigcup_{n=1}^N K_n$. [It is enough to prove this when $N = 2$ since the general case follows by induction. There exists an open set W such that $K \backslash U_2 \subset W \subset \bar{W} \subset U_1$. Take $K_1 = \bar{W}$ and $K_2 = K \backslash W$]. Hence $\phi(K) \leq \sum_{i=1}^N \phi(K_i) \leq \sum_{i=1}^N \mu^*(U_i) \leq \sum_n \mu^*(U_n)$. Taking supremum over K we get $\mu^*(\bigcup_n U_n) \leq \sum_n \mu^*(U_n)$. Now let $A_n, n = 1, 2...$ be arbitrary and $\epsilon > 0$. There exist open sets U_n with $A_n \subset U_n$ and $\mu^*(A_n) + \frac{\epsilon}{2^n} > \mu^*(U_n)$. Now $\mu^*(\bigcup_n A_n) \leq \mu^*(\bigcup_n U_n) \leq \sum_n \mu^*(U_n) < \epsilon + \sum_n \mu^*(A_n)$. It follows that μ^* is indeed an outer measure.

Our next claim is that $\mu^*(V) \geq \mu^*(V \cap U) + \mu^*(V \cap U^c)$ for any open sets U and V with $\mu^*(V) < \infty$.

Once this claim is established we can see easily that every open set if μ^* measurable: let $\mu^*(A) < \infty$ and $\epsilon > 0$. There is an open set V such that $\mu^*(A) + \epsilon > \mu^*(V)$ and $A \subset V$. Thus $\mu^*(A) + \epsilon > \mu^*(V \cap U) + \mu^*(V \cap U^c) \geq \mu^*(A \cap U) + \mu^*(A \cap U^c)$. Thus and sub-additivity show that $\mu^*(A) = \mu^*(A \cap U) + \mu^*(A \cap U^c)$. Now we prove the claim. There exists a compact set K such that $K \subset U \cap V$ and $\phi(K) > \mu^*(U \cap V) - \epsilon$. There is a compact set $H \subset V \backslash K$ such that $\phi(H) > \mu^*(V \backslash K) - \epsilon$. Note that $K \cap H = \emptyset$. Also $V \backslash U \subset V \backslash K$ so $\phi(H) > \mu^*(V \backslash K) - \epsilon \geq \mu^*(V \backslash U) - \epsilon$. Hence $\mu^*(V \cap U) + \mu^*(V \cap U^c) - 2\epsilon < \phi(K) + \phi(H) = \phi(H \cup K) \leq \mu^*(V)$. It follows that every open set, and hence every Borel set, is μ^* measurable. Let μ be the restriction of μ^* to the Borel sigma field.

We now prove that μ is left invariant: the fact that $\phi(xK) = \phi(K)$ shows that $\mu^*(xU) = \mu^*(U)$ and hence $\mu^*(xA) = \mu^*(A)$. [Here K, U, A are typical compact, open and arbitrary subsets].

This finishes the proof.

Remark: it can be shown that μ is regular. However this doesn't require a proof if the group G is a locally compact Polish space and this is the case we are interested in. (In fact we are interested only in compact metric groups).

Theorem

Left/right Haar measure on a locally compact Polish group G is unique up to a constant. If G is a compact metric group then there is a unique Haar measure which is also a probability measure.

Proof: let μ and ν be two left Haar measures. Let $g : G \to \mathbb{C}$ be a non-negative continuous function with compact support which is not identically 0. Let $f : G \to \mathbb{C}$ be a continuous function with compact support. We first show that $\mu(U) > 0$ for any non-empty open set U. Since μ is regular there is a compact set K with $\mu(K) > 0$. [By definition a Haar measure is not identically 0!]. Since $K \subset \bigcup\limits_{x \in G} xU$ there is a finite subcover. Suppose $K \subset \bigcup\limits_{i=1}^{n} x_i U$. Clearly, $\mu(x_i U) > 0$ for some i and hence $\mu(U) = \mu(x_i U) > 0$. We can now conclude that $\displaystyle\int g d\mu > 0$. [$g > 0$ on some non-empty open set]. For the first part of the theorem It is enough to show that $\frac{\int f d\mu}{\int g d\mu} = \frac{\int f d\nu}{\int g d\nu}$. If ξ is the continuous complex function with compact support on

$G \times G$ defined by $\xi(x,y) = \frac{f(x)g(yx)}{\int g(zx)d\nu(z)}$ it follows that $\int\int \xi(x,y)d\nu(y)d\mu(x) =$
$\int\int \xi(y^{-1}x,y)d\mu(x)d\nu(y) = \int\int \xi(y^{-1}x,y)d\nu(y)d\mu(x) = \int\int \xi(y^{-1},xy)d\nu(y)d\mu(x)$
(we have changed y to xy to get the last equality) and hence that $\int f d\mu =$
$\int g d\mu \int \frac{f(y^{-1})}{\int g(zy^{-1})d\nu(z)} d\nu(y)$. [The function ξ defined here is obviously continuos on $G \times G$; its support is contained in $C \times DC^{-1}$ where C and D are the supports of f and g respectively]. We have proved that $\frac{\int f d\mu}{\int g d\mu}$ does not depend on μ proving that first part of the theorem. If G is compact then $\mu(G)$ and $\nu(G)$ are finite and hence $\frac{\mu}{\mu(G)} = \frac{\nu}{\nu(G)}$ proving that there is a unique left invariant probability measure in this case.

If G is compact then the unique left invariant probability measure is also the unique right invariant probability measure. We shall not prove this here since our interest is mainly in abelian groups.

END OF APPENDIX

APPENDIX

ANALYTIC SETS AND ISOMORPHISM THEOREMS

Borel isomorphism and analytic sets [From Cohn's "Measure Theory"]

A *Polish* space is a topological space which can be metrized to become a complete separable metric space.

It is well known fact that a subspace of a Polish space is a Polish space iff it is a G_δ set.

A subset A of a Polish space X is called *analytic* if it is the continuous image of a Polish space.

Theorem

Open sets and closed sets in a Polish space are analytic.

Proof: open subsets and closed subsets are themselves Polish.

Theorem

Countable unions and countable intersections of analytic sets are analytic.

Proof: let $f_n : Z_n \to A_n$ be continuous and onto where Z_n is a Polish space. Let $Z = \bigcup(Z_n \times \{n\})$ and declare that a set $B \subset Z$ is open if, for each n,

$B \cap (Z_n \times \{n\}) = U_n \times \{n\}$ for some open set U_n in Z_n. This defines a topology on Z. Let d_n be a metric for the topology of Z_n which makes it complete and separable. Assume (w.l.o.g.) that $d_n \leq 1$. Define $d(z_1, z_2) = d_n(x, y)$ if $z_1 = (x, n)$ and $z_2 = (y, n)$ and define $d(z_1, z_2)$ to be 1 if z_1, z_2 are not of this type. Then (Z, d) is complete and separable and the map $\phi : Z \to \bigcup\limits_n A_n$ defined by $\phi((x, n)) = f_n(x)$ is a continuous surjective map. This proves that $\bigcup\limits_n A_n$ is analytic. Now let $H = \prod\limits_{n=1}^{\infty} Z_n$ and define $D(\{x_n\}, \{y_n\}) = \sum\limits_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1+d_n(x_n, y_n)}$. (H, D) is a complete separable metric space and $K = \{\{x_n\} \in H : f_n(x_n) = f_1(x_1)\}$ is a closed subset of H. Hence K is a Polish space and the range of the map $\{x_n\} \to f_1(x_1)$ is $\bigcap A_n$. Hence $\bigcap A_n$ is also analytic.

Theorem
Every Borel subset of a Polish space is analytic.

Lemma:
Let (X, τ) be any Hausdorff topological space. The smallest class of sets containing all open sets and all closed sets closed under countable intersections and countable disjoint unions coincides with the Borel sigma field.

Proof: let $\mathcal{B}$ be the class mentioned in the statement. Let $\mathcal{B}_0 = \{A \in \mathcal{B} : A^c \in \mathcal{B}\}$. If $\{A_n\} \subset \mathcal{B}_0$ then $\bigcup\limits_n A_n = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash \{A_1 \cup A_2\}) \cup ...$ is a disjoint union of sets in $\mathcal{B}$ and hence it belongs to $\mathcal{B}$. Since $\mathcal{B}$ is closed under complementation also it is a sigma field. Since it contains open sets it contains all Borel sets. The Borel sigma field is therefore contained in $\mathcal{B}$. Since the Borel sigma field also satisfies the properties satisfied by $\mathcal{B}$ it follows that $\mathcal{B} \subset$ Borel sigma field.

The theorem follows immediately: the class of analytic sets also satisfies above properties and hence it contains $\mathcal{B}$, which is the same as the Borel sigma field.

Theorem
Product of analytic sets is analytic.

Proof: let A_n be an analytic set in a Polish space Z_n for $n = 1, 2, ...$ We have to show that $\prod\limits_{n=1}^{\infty} A_n$ is an analytic set in the Polish space $\prod\limits_{n=1}^{\infty} Z_n$. Define $f : \prod\limits_{n=1}^{\infty} Z_n \to \prod\limits_{n=1}^{\infty} A_n$ by $f(\{z_n\}) = (f_n(z_n))$ where $f_n : Z_n \to A_n$ is continuous and onto.

This map is continuous and the image of $\prod\limits_{n=1}^{\infty} Z_n$ under it is $\prod\limits_{n=1}^{\infty} A_n$.

Theorem

Let X and Y be Polish and $A \subset X$ analytic. Let $f : A \to Y$ be Borel measurable, A_1 be analytic in X and A_2 be analytic in Y. Then $f(A \cap A_1)$ and $f^{-1}(A_2)$ are analytic.

Remark: taking $A = X$ we see that images and inverse images of analytic sets under Borel measurable maps are analytic.

Proof: let $\pi_2(x, y) = y \ \forall (x, y) \in X \times Y$. We claim that the graph $G_f = \{(a, f(a)) : a \in A\}$ of f is a Borel subset of $A \times Y$. Indeed $(a, y) \to ((f(a), y)$ is Borel measurable from $A \times Y \to Y \times Y$ and the graph of f is the inverse image under this map of $\Delta = \{(y_1, y_2) \in Y \times Y : y_1 = y_2\}$. Hence $G_f = (A \times Y) \cap B$ for some Borel set B in $X \times Y$. Now $G_f \cap (A_1 \times Y)$ is an analytic subset of $X \times Y$ because it is the intersection of two analytic sets. Let $\xi : Z \to G_f \cap (A_1 \times Y)$ be continuous and onto where Z is a Polish space. Now $(\pi_2 \circ \xi)(Z) = f(A \cap A_1)$ and hence this last set is analytic. Now let $\pi_1(x, y) = x$. The set $G_f \cap (X \times A_2)$ is an analytic set in $X \times Y$ and $f^{-1}(A_2) = \pi_1(G_f \cap (X \times A_2))$. Hence $f^{-1}(A_2)$ is analytic.

Theorem

Any Polish space is the continuous image of $\mathbb{N}^{\mathbb{N}}$ (with product topology).

Proof: recall that $\mathbb{N}^{\mathbb{N}}$ is complete and separable under the metric $d(\{n_k\}, \{m_k\}) = \sum\limits_{k=1}^{\infty} \frac{1}{2^k} \frac{|n_k - m_k|}{1+|n_k - m_k|}$ $d(\{n_k\}, \{m_k\}) < \frac{1}{2^{N+1}}$ implies that (the first N terms in the sum are necessarily 0) and $n_j = m_j$ for $1 \leq j \leq N$. Let d be a complete separable metric for a Polish space X. With each finite sequence $(n_1, n_2, ..., n_k)$ of positive integers we associate a set $C(n_1, n_2, ..., n_k)$ in such a way that each of these is a non-empty closed set, the diameter of $C(n_1, n_2, ..., n_k)$ does not exceed $1/k$, $C(n_1, n_2, ..., n_{k-1}) = \bigcup\limits_{j=1}^{\infty} C(n_1, n_2, ..., n_{k-1}, j)$ and $X = \bigcup\limits_{n_1=1}^{\infty} C(n_1)$. [To see that such sets exists we take a countable dense set $\{x_n\}$ and define $C(j)$ to be the closed ball with center x_j and radius $\frac{1}{2}$. Suppose we have constructed $C(n_1, n_2, ..., n_k)$. We can write this set as a union of a countable number of closed balls of radius $\frac{1}{2(k+1)}$ each and we denote these balls by $C(n_1, n_2, ..., n_k, j), j = 1, 2, ...$ This completes the construction]. Now let $\{n_1, n_2, ...\}$ be an element of $\mathbb{N}^{\mathbb{N}}$. Then the intersection of the sets $C(n_1, n_2, ..., n_k)$ over k is a singleton set $\{x\}$. we consider the map $\{n_1, n_2, ...\} \to x$ from $\mathbb{N}^{\mathbb{N}}$ into X. It is clear from the properties of the sets $C(n_1, n_2, ..., n_k)$ that this map is onto X.

Suppose x and y are the images of $\{n_k\}$ and $\{m_k\}$ and $d(\{n_k\}, \{m_k\}) < \frac{1}{2^{N+1}}$. Then $n_j = m_j$ for $1 \leq j \leq N$. In particular x and y both belong to $C(n_1, n_2, ..., n_N) = C(m_1, m_2, ..., m_N)$ which implies that $d(x, y) \leq \frac{1}{N}$. This proves continuity of the map $\{n_1, n_2, ...\} \to x$. This completes the proof.

Corollary

If A is a non-empty analytic set in a Polish space X then A is a continuous image of $\mathbb{N}^{\mathbb{N}}$.

Proof: this is obvious from the definition of analytic sets and above theorem.

Theorem

Let X be Polish and $A \subset X$. Then A is analytic if and only if there is a closed set C in $\mathbb{N}^{\mathbb{N}} \times X$ whose projection (on the second coordinate) is A.

Proof: 'if' part is obvious. Suppose A is analytic. By above corollary there is a continuous map f from $\mathbb{N}^{\mathbb{N}}$ onto A. The graph G_f of f is a closed subset of $\mathbb{N}^{\mathbb{N}} \times X$ and its projection on the second coordinate is A.

A topological space is *zero dimensional* if there is a basis consisting of clopen (i.e. closed and open) sets. Discrete spaces and subspaces/products/disjoint unions of zero dimensional spaces are zero dimensional. $\mathbb{N}^{\mathbb{N}}$ and $\{0,1\}^{\mathbb{N}}$ are zero dimensional.

Theorem

If X is Polish and A is a Borel set in X then there is a zero dimensional Polish space Z and a continuous one-to-one map $f : Z \to X$ with $f(Z) = A$.

The map $\{a_n\} \to \sum\limits_{n=1}^{\infty} \frac{a_n}{2^n}$ is a continuous map of the zero dimensional space $\{0,1\}^{\mathbb{N}}$ onto $[0,1]$. Restricting it to sequences with infinitely many $1's$ together with the point $\{0,0,...\}$ we get a continuous injective map of a zero dimensional space onto $[0,1]$. Also the domain of this restriction is a G_δ in $\{0,1\}^{\mathbb{N}}$ (since a countable set is an F_σ) and hence it is Polish. This proves that $[0,1]$ is a continuous injective image of a zero dimensional Polish space Z. It follows that $[0,1]^{\mathbb{N}}$ is a continuous injective image of the zero dimensional Polish space $Z^{\mathbb{N}}$. [Because a product of zero dimensional Polish spaces is a zero dimensional Polish space]. Let $\chi : Z^{\mathbb{N}} \to [0,1]^{\mathbb{N}}$ be a 1-1 continuous map with range $[0,1]^{\mathbb{N}}$. Now let X be a Polish space. X is homeomorphic to a subspace of $[0,1]^{\mathbb{N}}$. [If $\{x_n\}$ is dense in X and d is a complete metric with $d(x,y) \leq 1$ for all x, y then $x \to \{d(x, x_n)\}$ is the desired homeomorphism]. The range of this homeomorphism is Polish and hence it is a G_δ in $[0,1]^{\mathbb{N}}$. Let $\phi : X \to S \subset [0,1]^{\mathbb{N}}$ be a homeomorphism (onto S) where S is a G_δ in $[0,1]^{\mathbb{N}}$. Now $\chi^{-1}(S)$ is a G_δ in $Z^{\mathbb{N}}$ (because S is). Thus, $\chi^{-1}(S)$ is Polish and zero dimensional. Note that $\phi^{-1}(\chi(\chi^{-1}(S))) = X$ and $\phi^{-1} \circ \chi$ is continuous and one-to-one. This proves the theorem when $A = X$. Consider the class $\mathcal{G}$ of all Borel sunsets of X that are continuous one-to-one images of zero dimensional Polish spaces. All open and closed sets in X belong to this family. We claim that countable intersections and countable disjoint unions of sets in $\mathcal{G}$ belong to $\mathcal{G}$. Once this is proved we can conclude that $\mathcal{G}$ contains every Borel set and the proof of the theorem would be complete. Let $\{A_n\} \subset \mathcal{G}$. Let Z_n be a zero dimensional Polish space and $f_n : Z_n \to A_n$ be one-to-one, continuous and onto. Let $\Delta = \{\{z_n\} \in \prod Z_n : f_n(z_n) = f_1(z_1)\ \forall n\}$ and define $f : \Delta \to X$ by $f(\{z_n\}) = f_1(z_1)$. This gives a one-to-one continuous map of Δ onto $\bigcap_n A_n$ and hence $\bigcap_n A_n \in \mathcal{G}$. If $Z_0 = \bigcup_n (Z_n \times \{n\})$ is the 'disjoint union' of $Z_n's$ (see the proof of the fact that disjoint union of analytic sets are analytic for details) and if $f : Z_0 \to X$ is defined by $f(z_n \times \{n\}) = f_n(z_n)$ then we get a one-to-one continuous map of Z_0 onto $\bigcup_n A_n$ provided $A_n's$ are disjoint and hence $\bigcup_n A_n \in \mathcal{G}$ in this case.

Lemma

Let X be a zero dimensional separable metric space and U be an open subset which is not compact. Let $\epsilon > 0$. We can find an infinite sequence of disjoint clopen (non-empty) sets $A_1, A_2, ...$ each having diameter less than ϵ such that $U = \bigcup_n A_n$.

Proof: let $\{U_i : i \in I\}$ be an open cover of U with no finite subcover. Consider the collection $\mathcal{G}$ of all clopen sets which are contained in some U_i and which have diameter less than ϵ. Every point of U lies in an open ball of radius less than $\epsilon/2$ contained in $U \cap U_i$ and there is a clopen set containing the point and contained in this open ball. Hence U coincides with the union of all the members of $\mathcal{G}$. By separability we can write this union as a disjoint union, say $\bigcup_n V_n$. Writing this union as $V_1 \cup (V_2 \backslash V1) ... (V_n \backslash \{V_1 \cup V_2 \cup ... \cup V_{n-1}\}) \cup ...$ we see that U is the union of disjoint clopen (non-empty) sets each having diameter less than ϵ. Note that each of these clopen sets is a subset of some U_i. Since $\{U_i\}$ has no finite subcover our collection $\{A_n\}$ is necessarily infinite.

Lemma

Let C be the set of all condensation points (i.e. points such that every neighbourhood of them is uncountable) of a separable metric space X. Then C is closed and $X \backslash C$ is at most countable.

Proof: let $\{U_n\}$ be a countable basis for X. Note that $x \notin C$ iff $B(x, r)$ is countable for some r iff there exists U_n such that $x \in U_n$ and U_n is countable. In this case every point of U_n is also in $X \backslash C$ and hence $X \backslash C$ is open and C is closed. It is also clear that $X \backslash C$ is a (countable) union of countable sets U_n and hence it is countable.

Theorem

Let X be Polish and B be an uncountable Borel subset. There exists a continuous one-to-one map $f : \mathbb{N}^{\mathbb{N}} \to X$ whose range $f(\mathbb{N}^{\mathbb{N}})$ is contained in B such that $B \backslash f(\mathbb{N}^{\mathbb{N}})$ is at most countable.

Proof: there exists a zero dimensional Polish space Z and a continuous injective map $g : Z \to X$ whose range is B. We claim that there is a continuous injective map $\phi : \mathbb{N}^{\mathbb{N}} \to Z$ such that $Z \backslash \phi(\mathbb{N}^{\mathbb{N}})$ is countable. Once this is proved the function $f = g \circ \phi$ has the desired properties. The set C of all condensation points of Z is Polish and zero dimensional. Also $Z \backslash C$ is countable. Note that every point of C is a condensation point of C. [Because $Z \backslash C$ is countable]. Let d be a complete metric for C. We construct sets $A(n_1, n_2, ..., n_k)$ for $k \geq 1$ and $n_i's \in \mathbb{N}$ as follows: Let U be obtained from C by removing one point. Clearly U is open and non-compact. By one of the lemmas above U is the union of an infinite sequence of disjoint sets each of which is non-empty and clopen and has diameter less than 1. Call these sets $A(n_1), n_1 = 1, 2, ...$ Note that each point of $A(n_1)$ is a condensation point of it. [This is because these sets are open: if $x \in A(n_1)$ and V is an open set containing x then $A(n_1) \cap V$ is also a neighbourhood of x and so it contains uncountable many points of C]. We can repeat this construction by replacing C by any of the sets $A(n_1)$. By an induction argument we can construct sets $A(n_1, n_2, ..., n_k)$ with the following properties: these sets are clopen nonempty, diameter of $A(n_1, n_2, ..., n_k)$ is less than $\frac{1}{k}$ and $A(n_1, n_2, ..., n_k)$ is the union of $A(n_1, n_2, ..., n_k, n_{k+1})$ together with a singleton. Define $h : \mathbb{N}^{\mathbb{N}} \to Z$ by $h(\{n_k\}) = z$ where $\{z\} = \bigcap_k A(n_1, n_2, ..., n_k)$. It is easy to see that this map is injective and continuous. Also $h(\mathbb{N}^{\mathbb{N}}) \subset C$ and $C \backslash h(\mathbb{N}^{\mathbb{N}})$ contains only the points removed in the construction of the sets $A(n_1, n_2, ..., n_k)$.

Corollary

Any uncountable Borel subset B of a Polish space X contains a homeomorphic copy of $\{0, 1\}^{\mathbb{N}}$.

Proof: There is a continuous injective map $f : \mathbb{N}^{\mathbb{N}} \to X$ whose image is contained in B. Restriction of f to $\{0, 1\}^{\mathbb{N}}$ gives the required homeomorphism.

Definition: let X be Polish. A subset A of $\mathbb{N}^{\mathbb{N}} \times X$ is *universal* for a family of subsets of X is every set in the family is a section of A.

Lemma

Let X be a separable metric space. There is an open A set in $\mathbb{N}^{\mathbb{N}} \times X$ which is universal for the class of all open sets in X. Also, there is a closed C set in $\mathbb{N}^{\mathbb{N}} \times X$ which is universal for the class of all closed sets in X.

Proof: consider $\{\emptyset, V_1, V_2, ...\}$ where $\{V_1, V_2, ...\}$ is a basis for X. Let $A = \bigcup_k \{\{\{n_1, n_2, ...\}, x\} : x \in V_{n_k}\}$

$= \bigcup_k \{A(k,n) \times V_n\}$ where $A(k,n) = \{\{\{n_1, n_2, ...\} \in \mathbb{N}^{\mathbb{N}} : n_k = n\}$. Since $A(k,n)$ is open it follows that A is open. If

$\{n_1, n_2, ...\} \in \mathbb{N}^{\mathbb{N}}$ then the section of A by $\{n_1, n_2, ...\}$ is $\bigcup_k V_{n_k}$ and these unions exhaust all open sets in X. A^c is a closed set whose sections exhaust all closed sets in X.

Theorem

If X is Polish then there is an analytic set A in $\mathbb{N}^{\mathbb{N}} \times X$ which is universal for the class of all analytic sets in X.

Proof: there is a closed subset C of $\mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}} \times X$ which is universal for the class of closed sets in $\mathbb{N}^{\mathbb{N}} \times X$. Let $A = \phi(C)$ where $\phi : \mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}} \times X \to \mathbb{N}^{\mathbb{N}} \times X$ is defined by $\phi(\{m_1, m_2, ...\}, \{n_1, n_2, ...\}, x) = (\{m_1, m_2, ...\}, x)$. Being a continuous image of a Polish space A is analytic. For each $\{m_1, m_2, ...\}$ the section of A by $\{m_1, m_2, ...\}$ is nothing but the projection on X of the section of C by $\{m_1, m_2, ...\}$. Recalling that analytic sets in X are precisely projections of closed sets in $\mathbb{N}^{\mathbb{N}} \times X$ we conclude that A is universal for analytic sets.

Theorem

There is an analytic in $\mathbb{N}^{\mathbb{N}}$ set which is not Borel.

Proof: let A be an analytic set in $\mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}}$ which is universal for the class of all analytic sets in X. Let $S = \{\{n_1, n_2, ...\} : (\{n_1, n_2, ..., \}, \{n_1, n_2, ...\}) \in A\}$. This is the projection on $\mathbb{N}^{\mathbb{N}}$ of $A \cap \Delta$ where Δ is the diagonal of $\mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}}$. Since Δ is closed it follows that S is analytic. Suppose S is Borel. Then so is S^c. Thus S^c is analytic and it must be a section of A. Let $\{m_1, m_2, ...\} \in \mathbb{N}^{\mathbb{N}}$ be such that

$S^c = \{\{n_1, n_2, ...\} : (\{m_1, m_2, ...\}, \{n_1, n_2, ...\}) \in A\}$. (*)

Either $\{m_1, m_2, ...\} \in S$ or $\{m_1, m_2, ...\} \in S^c$. In the first case $(\{m_1, m_2, ..., \}, \{m_1, m_2, ...\}) \in A$ which implies $\{m_1, m_2, ..., \} \in S^c$ by (*). In the second case $\{m_1, m_2, ...\} \in S^c = \{\{n_1, n_2, ...\} : (\{m_1, m_2, ...\}, \{n_1, n_2, ...\}) \in A\}$ and so $(\{m_1, m_2, ...\}, \{m_1, m_2, ...\}) \in A$ and the definition of S shows that $\{m_1, m_2, ..., \} \in S$, a contradiction again.

Theorem

Any uncountable Polish space has an analytic subset that is not Borel.

Proof: let A be an analytic set in $\mathbb{N}^{\mathbb{N}}$ which is not Borel. Suppose X is Polish, uncountable and every analytic subset of X is Borel. There is a continuous injective map $\phi : \mathbb{N}^{\mathbb{N}} \to X$ such that $X \backslash \phi(\mathbb{N}^{\mathbb{N}})$ is at most countable. $\phi(A)$ is analytic and hence Borel. It follows that $A = \phi^{-1}(\phi(A))$ is Borel too, a contradiction.

In particular not every analytic subset of $\mathbb{R}$ is a Borel set.

We now prove a separation theorem for analytic sets and use it prove a Borel isomorphism theorem.

Let X be a Polish space and $A_1, A_2 \subset X$. If there exist disjoint Borel sets B_1, B_2 such that $A_i \subset B_i . i = 1, 2$ we say A_1 and A_2 are separated by Borel sets.

Theorem

Disjoint analytic sets in Polish space can always be separated by Borel sets.

Proof: claim 1: if C_n and D can be separated by Borel sets for each n then $\bigcup_n C_n$ and D can be separated by Borel sets. Claim 2: if C_n and D_m can be separated by Borel sets for each n and m then $\bigcup_n C_n$ and $\bigcup_m D_m$ can be separated by Borel sets. Proofs of these two claims are elementary and we omit the details. Now let A_1 and A_2be disjoint analytic sets in a Polish space X. There exist continuous maps $f, g : \mathbb{N}^{\mathbb{N}} \to X$ such that $f(\mathbb{N}^{\mathbb{N}}) = A_1$ and $g(\mathbb{N}^{\mathbb{N}}) = A_2$. Suppose, if possible, we cannot separate A_1 and A_2 by Borel sets. Let $E_{n_1, n_2, ..., n_k} = \{\{m_1, m_2, ...\} \in \mathbb{N}^{\mathbb{N}} : m_i = n_i \text{ for } 1 \leq i \leq k\}$. Note that $\bigcup_n f(E_n) = f(\mathbb{N}^{\mathbb{N}}) = A_1$ and $\bigcup_n g(E_n) = g(\mathbb{N}^{\mathbb{N}}) = A_2$. Hence there exist n_1 and m_1 such that $f(E_{n_1})$ and $g(E_{m_1})$ cannot be separated by Borel sets. Since $f(E_{n_1}) = \bigcup_{n_2} f(E_{n_1, n_2})$ and $g(E_{m_1}) = \bigcup_{m_2} g(E_{m_1, m_2})$ we can find n_2 and m_2 such that $f(E_{n_1, n_2})$ and $g(E_{m_1, m_2})$ cannot be separated by Borel sets. By induction we can generate sequences $\{n_1, n_2, ...\}$ and $\{m_1, m_2, ...\}$ such that for any k $f(E_{n_1, n_2, .., n_k})$ and $g(E_{m_1, m_2, ..., m_k})$ cannot be separated by Borel sets. If $f(\{n_1, n_2, ...\}) \neq g(\{m_1, m_2, ...\})$ then these two points can be separated by disjoint open sets which implies that $f(\{n_1, n_2, ..., n_N\})$ and $g(\{m_1, m_2, ..., m_N\})$ can be separated by disjoint open sets for N sufficiently large. [If U is open and contains $f(\{n_1, n_2, ...\})$ then the neighbourhood $f^{-1}(U)$ of $\{n_1, n_2, ...\}$ contains all points $\{n'_1, n'_2, ...\}$ with $n'_i = n_i, 1 \leq i \leq N$ provided N is sufficiently large. Similar result holds for a neighbourhood of $\{m_1, m_2, ...\}$]. We have proved that $f(\{n_1, n_2, ...\}) = g(\{m_1, m_2, ...\})$ contradicting the fact that the left side belongs to A_1, the right side to A_2 [and A_1 and A_2 are disjoint].

Corollary

If $A_1, A_2, ..., A_n$ are disjoint analytic sets in a Polish space we can find disjoint Borel sets $B_1, B_2, ..., B_n$ such that $A_i \subset B_i$ for each i.

Proof: this is elementary.

Theorem

If A and A^c are both analytic then A is a Borel set.

Proof: there exist disjoint Borel sets B_1, B_2 such that $A \subset B_1$ and $A^c \subset B_2$. But then $A^c \subset B_2 \subset B_1^c$ and so $B_1 \subset A$ forcing A to be equal to the Borel set B_1.

Theorem

Let X and Y be Polish spaces and $f : X \to Y$. Then f is Borel measurable if and if only if its graph is a Borel set in $X \times Y$.

Proof: $G_f = g^{-1}(\Delta)$ where $g : A \times Y \to Y \times Y$ is defined by $g(a, y) = (f(a), y)$ and $\Delta = \{(y, y) : y \in Y\}$. Thus G_f is Borel if f is Borel measurable. Suppose G_f is Borel and B is any Borel set in Y. The disjoint analytic sets $G_f \cap (X \times B)$ and $G_f \cap (X \times B^c)$, when projected on the first coordinate, yield disjoint analytic sets and so we can find disjoint Borel sets E_1, E_2 containing these projections. However these two projections are nothing but $f^{-1}(B)$ and $f^{-1}(B^c)$. Thus $f^{-1}(B) \subset E_1$ and $f^{-1}(B^c) \subset E_2$. From the disjointness of E_1 and E_2 it follows easily that $f^{-1}(B) = E_1 \cap A$. Hence f is Borel measurable.

Theorem

Let X and Y be Polish, $A \subset X$ Borel and $f : A \to Y$ be Borel measurable. Suppose f is injective and $B = f(A)$ is a Borel set. Then f^{-1} is Borel measurable.

Proof: think of f^{-1} as a map from Y into X. $G_{f^{-1}} = \theta(G_f)$ where $\theta(x, y) = \theta(y, x)$. Since G_f is Borel and $\theta = \theta^{-1}$ is Borel measurable it follows that $G_{f^{-1}}$ is a Borel set and hence f^{-1} is Borel measurable.

Theorem

Let A and B be Borel sets in Polish spaces. Then A and B are Borel isomorphic if and only if they have the same cardinality.

Proof: suppose A and B are Borel sets in Polish spaces X and Y with the same cardinality. If these sets are countable they are clearly Borel isomorphic, so we assume that they are uncountable. There exist continuous injective maps $f : \mathbb{N}^{\mathbb{N}} \to A, g : \mathbb{N}^{\mathbb{N}} \to B$ such that $A \backslash f(\mathbb{N}^{\mathbb{N}})$ and $B \backslash g(\mathbb{N}^{\mathbb{N}})$ are at most countable. It is clear that these two countable sets are Borel sets (and so are their complements) and $f : \mathbb{N}^{\mathbb{N}} \to f(\mathbb{N}^{\mathbb{N}}), f : \mathbb{N}^{\mathbb{N}} \to f(\mathbb{N}^{\mathbb{N}})$ are Borel isomorphisms by previous theorem. Thus $g \circ f^{-1} : f(\mathbb{N}^{\mathbb{N}}) \to g(\mathbb{N}^{\mathbb{N}})$ is a Borel isomorphism. We can extend this to a Borel isomorphism of A onto B using an arbitrary bijection on their complements provided $A \backslash f(\mathbb{N}^{\mathbb{N}})$ and $B \backslash g(\mathbb{N}^{\mathbb{N}})$ are infinite. If they are finite we can remove countable infinite sets from the ranges of f and g and combine them with $A \backslash f(\mathbb{N}^{\mathbb{N}})$ and $B \backslash g(\mathbb{N}^{\mathbb{N}})$. We leave it to the reader to fill in the details of this argument.

Theorem

Let X and Y be Polish. Let $A \subset X$ be Borel and $f : A \to Y$ be Borel measurable and injective. Then $f(A)$ is a Borel set.

Proof: we claim that there is a Borel measurable function $g : Y \to X$ such that $g(Y) \subset A$ and $g \circ f$ is the identity on A. Once this claim is proved we can conclude that $f(A) = \{y \in Y : f(g(y)) = y\}$ is a Borel set since the identity map of Y and the map $f \circ g$ are Borel measurable on Y. We now prove the claim. Let d be a complete metric for X and $x_0 \in A$. We define a sequence $\{g_n\}$ of functions from Y to X as follows: for each n let $\{A_{n,1}, A_{n,2}, ...\}$ be a partition of A into non-empty Borel sets of diameter at most $\frac{1}{n}$. Let $x_{n,k} \in A_{n,k}$. Since $\{f(A_{n,k}) : k = 1, 2, ...\}$ are disjoint analytic sets we can separate them by disjoint Borel sets $\{B_{n,k} : k = 1, 2, ...\}$. Let $g_n(y) = \sum_k x_{n,k} I_{B_{n,k}} + x_0 I_{Y \setminus \bigcup_k B_{n,k}}$. Each g_n is a Borel measurable function from Y into A. If $x \in A$ then $d(x, g_n(f(x))) \leq 1/n$. In fact $x \in A_{n,k}$ for some k and $f(x) \in f(A_{n,k}) \subset B_{n,k}$ so $g_n(f(x)) = x_{n,k}$; thus $d(x, g_n(f(x))) \leq$ diameter of $A_{n,k} \leq 1/n$. Thus $\lim_n g_n(y)$ exists if $y \in f(A)$. Define $g(y)$ to be this limit if it exists and x_0 if it doesn't. Then $d(x, g(f(x))) = 0$ for all $x \in A$. This proves the claim.

Our next aim is to show that analytic sets in Polish spaces are *universally measurable*,i.e. they belong to the completion of the Borel sigma field w.r.t. any finite measure μ.

Let A be an analytic subset of a Polish space X. Let μ be a finite positive Borel measure on X. Let $\mu^*(A) = \inf\{\mu(B) : A \subset B, B \text{ Borel}\}$. We claim that for any $\epsilon > 0$ we can find a compact set K such that $K \subset A$ and $\mu(K) \geq \mu^*(A) - \epsilon$. Suppose the claim is proved. We can find B Borel such that $A \subset B$ and $\mu^*(A) + \epsilon > \mu(B)$. Thus, taking $\epsilon = 1/n$ we get compact sets K_n and Borel sets B_n with $K_n \subset A \subset B_n$ and $\mu(B_n \setminus K_n) = \mu(B_n) - \mu(K_n) < \mu^*(A) + 1/n - \{\mu^*(A) - 1/n\} = 2/n$. Now $A = \{\bigcup K_n\} \cup \{A \setminus \bigcup K_n\}$ and $\{A \setminus \bigcup K_n\} \subset \bigcap \{B_n \setminus K_n\}$. Since $\mu(\bigcap \{B_n \setminus K_n\}) = 0$ it follows that A is $\mu-$ measurable. Thus it remains only to prove the claim. There is a continuous map $f : \mathbb{N}^{\mathbb{N}} \to A$ which is surjective. Let $I_{n_1, n_2, ..., n_k} = \{\{m_1, m_2, ...\} \in \mathbb{N}^{\mathbb{N}} : m_i \leq n_i$ for $1 \leq i \leq k\}$. We now show that there exist positive integers $n_1, n_2, ...$ with $\mu^*(f(I_{n_1, n_2, ..., n_k})) > \mu^*(A) - \epsilon$ for all k. [It is not obvious that this proves our claim!]. Since $\mu^*(f(I_n)) \uparrow \mu^*(A)$ we can find n_1 such that $\mu^*(f(I_{n_1})) > \mu^*(A) - \epsilon$. Since $I_{n_1} = \bigcup I_{n_1, n_2}$ we can find n_2 such that $\mu^*(f(I_{n_1, n_2})) > \mu^*(A) - \epsilon$. Induction produces $n_1, n_2, ...$ with $\mu^*(f(I_{n_1, n_2, ..., n_k})) > \mu^*(A) - \epsilon$ for all k. Let $K = f(I)$ where $I = \bigcap_k I_{n_1, n_2, ..., n_k}$. K is compact because I is compact in $\mathbb{N}^{\mathbb{N}}$. It remains to show that $\mu^*(K) \geq \mu^*(A) - \epsilon$. We claim that $K = \bigcap_k [f(I_{n_1, n_2, ..., n_k})]^-$. Let d be a complete separable metric for X. Let $x \in \bigcap_k [f(I_{n_1, n_2, ..., n_k})]^-$. For each k there exists $m^{(k)} \in I_{n_1, n_2, ..., n_k}$ such that $d(x, f(m^{(k)})) < 1/k$. By a diagonal procedure we can extract a subsequence $\{m^{(k_j)}\}$ of $\{m^k\}$ which converges to some $m \in \mathbb{N}^{\mathbb{N}}$. Clearly $m \in I$ and $f(m^{(k_j)}) \to f(m)$ so $d(x, f(m)) = 0$. Thus $x = f(m) \in K$. Thus K contains $\bigcap_k [f(I_{n_1, n_2, ..., n_k})]^-$. Obviously $K \subset \bigcap_k [f(I_{n_1, n_2, ..., n_k})]^-$ and we have proved that $K = \bigcap_k [f(I_{n_1,n_2,\ldots,n_k})]^-$. Now $[f(I_{n_1,n_2,\ldots,n_k})]^-$ is closed and contains $f(I_{n_1,n_2,\ldots,n_k})$. Hence $\mu([f(I_{n_1,n_2,\ldots,n_k})]^-) \geq \mu^*(f(I_{n_1,n_2,\ldots,n_k})) > \mu^*(A) - \epsilon$. Since $[f(I_{n_1,n_2,\ldots,n_k})]^-$ is decreasing we have $\mu(K) = \lim \mu[f(I_{n_1,n_2,\ldots,n_k})]^- \geq \mu^*(A) - \epsilon$ which proves our claim.

We have proved the following:

Theorem
Any analytic subset of a Polish space is universally measurable.

Remark: if A is a Borel set in $\mathbb{R}^2$ then the projection of A on the first (or second) coordinate is analytic, hence universally measurable, but it need not be a Borel set. To see that the projection need not be Borel we prove the following:

Theorem
If X is a Polish space and $A \subseteq X$ then the following are equivalent:

a) there is a Borel set in $X \times X$ whose (first) projection is A
b) there is a Polish space Y and a Borel set in $X \times Y$ whose (first) projection is A
c) there is a continuous map from $\mathbb{N}^{\mathbb{N}}$ into X with range A
d) there is a closed set in $X \times \mathbb{N}^{\mathbb{N}}$ whose projection is A
e) for any uncountable Polish space Y there exists a G_δ set in $X \times Y$ whose projection is A
f) A is analytic

Note that if A is an analytic set in $\mathbb{R}$ which is not Borel then A is the projection of a Borel set in $\mathbb{R} \times \mathbb{R}$ by f) implies a); hence *there is a Borel set in $\mathbb{R}^2$ whose projection on $\mathbb{R}$ is not Borel. In fact there is a Closed set in $\mathbb{R}^2$ whose projection on $\mathbb{R}$ is not Borel. A proof is given below.*

Proof: any analytic set is a continuous image of a Polish space and any Polish space is a continuous image of $\mathbb{N}^{\mathbb{N}}$. Hence f) implies c). Since $\mathbb{N}^{\mathbb{N}}$ is a Polish space c) implies f). a) implies b) is obvious. Let b) hold, say $A = p(B)$, B Borel in $X \times Y$, p being the projection map of $X \times Y$ onto X. Since B is analytic it is a continuous image of a Polish space and any Polish space is a continuous image of $\mathbb{N}^{\mathbb{N}}$ (by an earlier theorem). Hence there is a continuous map $\phi : \mathbb{N}^{\mathbb{N}} \to X \times Y$ with range B. Now $A = p(\phi(\mathbb{N}^{\mathbb{N}}))$ and this proves c) since $p \circ \phi$ is continuous. c) implies d): let $f : \mathbb{N}^{\mathbb{N}} \to X$ be continuous with range A. Then G_f is closed in $\mathbb{N}^{\mathbb{N}} \times X$. Its projection on X is A. Applying the homeomorphism $(a, x) \to (x, a)$ to G_f and then projecting it will give us a closed set in $X \times \mathbb{N}^{\mathbb{N}}$ whose projection is A. d) implies e): let $\phi : \mathbb{N}^{\mathbb{N}} \to Y$ be a homeomorphism into Y. [This follows from a theorem of Mazurkiewicz which is proved below.]. The range E of ϕ is a G_δ in Y because it is Polish. Let C be a closed set in $X \times \mathbb{N}^{\mathbb{N}}$ whose projection is A. $\{(x, \phi(a)) : (x, a) \in C\}$ is a

G_δ set in $X \times Y$ whose projection is A. [If $C = \bigcap_{n=1}^{\infty} W_n$ and $E = \bigcap_{n=1}^{\infty} V_n$ with W_n and V_n for each n then $\{(x, \phi(a)) : (x, a) \in C\}$ is the intersection of the sets $\{(x, \phi(a)) : (x, a) \in W_n, \phi(a) \in V_n\}$ which is open in E. Since E is itself a G_δ in Y it follows that $\{(x, \phi(a)) : (x, a) \in C\}$ is a G_δ set in $X \times Y$].Since e) implies a) is obvious the proof is complete.

Theorem [Mazurkiewicz]

$\mathbb{N}^{\mathbb{N}}$ is homeomorphic to any G_δ subset S of a zero dimensional (z.d.) Polish space X such that S and $X \backslash S$ are both dense in X.

Corollary

$\mathbb{N}^{\mathbb{N}}$ is homeomorphic to a subset of $\mathbb{R}$. In fact every uncountable Polish space contains a homeomorphic copy of $\mathbb{N}^{\mathbb{N}}$.

Proof of the corollary: let X be the set of all irrational numbers. Since X is a G_δ in $\mathbb{R}$ is it Ploish. It is z.d. because $\{(\alpha, \beta) \cap X : \alpha < \beta, \alpha \in \mathbb{Q}, \beta \in \mathbb{Q}\}$ is a clopen basis for X. Let $S = X \backslash C$ where C is a countable dense subset of X. Then S and $X \backslash S$ are dense in X and S is a G_δ. The theorem implies that $\mathbb{N}^{\mathbb{N}}$ is homeomorphic to S.

For the second part note that every uncountable Polish space contains a homeomorphic copy of $\{0,1\}^{\mathbb{N}}$.

The set S is homeomorphic to a subset of $\{0,1\}^{\mathbb{N}}$ via dyadic expansion.

Proof of the theorem: we can write S as $G_1 \cap G_2 \cap ...$ where each G_n is open and $G_{n+1} \subseteq G_n \ \forall n$. Since S is not closed there exists j_1 such that G_{j_1} is not closed. Since G_{j_1} is open we can write it as $A_1 \cup A_2 ...$ where each A_k is a non-empty clopen set, the diameter of each A_k does not exceed 1 and $A_k's$ are disjoint. [The fact that there are infinitely many non-empty $A_k's$ follows from the fact that their union G_{j_1} is not closed]. Now fix k and consider the clopen set A_k. We claim that $A_k \cap G_j$ is not closed for some $j \geq 2$. Suppose this is false. Then $S \cap A_k = A_k \cap G_2 \cap G_3 \cap ...$ is closed. If $x \in A_k \backslash S$ then there exists a sequence in $S \cap A_k$ converging to x. [This follows from the hypothesis since A_k is a non-empty open set]. This is a contradiction. Hence $A_k \cap G_j$ is not closed for some $j = j_2 \geq 2$. We can write $A_k \cap G_{j_2}$ as a disjoint union of sets $A_{k1}, A_{k2}, ...$ each of which is clopen and non-empty with diameter not exceeding $\frac{1}{2}$. Now for each A_{kj} there exists $j_2 \geq 3$ such that $A_{kj} \cap G_{j_2}$ is not closed. We get a collection $\{A_{kjl}\}$ of non-empty clopen sets with diamter of A_{kjl} not exceeding $\frac{1}{3}$ whose union is $A_{kj} \cap G_{j_2}$. By induction we construct copen sets $A_{k_1 k_2 ... k_n}$. Let $x \in S$. Then $x \in G_{j_n}$ for each n. Hence there exist $k_1, k_2, ...$ such that $x \in \bigcap_n A_{k_1 k_2 ... k_n}$. This intersection is a singleton set. The map $\phi : x \to \{k_1, k_2, ...\}$ is a bijective map from S into $\mathbb{N}^{\mathbb{N}}$. [If $x \in \bigcap_n A_{k_1 k_2 ... k_n}$ then $x \in G_{j_r}$ for each r, hence $x \in S$]. Suppose $x_j \to x$. Let k_1 be such that $x \in A_{k_1}$. Since A_{k_1} is open $x_j \in A_{k_1}$ for j sufficiently large. Hence the first coordite of $\phi(x_j)$ converfges to the first coordinate of $\phi(x)$. Clearly the same is true of all coordinates which proves continuity of ϕ. Continuity of ϕ^{-1} follows easily from the fact that diameter of $A_{k_1 k_2 \ldots k_n}$ tends to 0 as $n \to \infty$.

Corollary

There exists a closed set C in $\mathbb{R}^2$ whose projection on the first coordinate is not Borel.

Proof: there exists an analytic set A in $\mathbb{R}$ which is not Borel. A is the image of $\mathbb{N}^{\mathbb{N}}$ under a continuous real . valued map f . Thje graph G_f of f is a closed set in $\mathbb{N}^{\mathbb{N}} \times \mathbb{R}$. Regarding $\mathbb{N}^{\mathbb{N}}$ as a subset of $\mathbb{R}$ (possible by previous corollary) we get a closed set C in $\mathbb{R}^2$ whose projection on the first coordinate is A which is not Borel).

An isomorphism theorem for measure algebras

Theorem

Any separable non-atomic measure algebra is isomorphic to the measure algebra of Lebesgue measure on $(0, 1)$.

Here are the definitions of the terms used in this theorem: let $(\Omega, \mathcal{F}, P)$ be a probability space and Z the metric space of equivalence classes of sets in $\mathcal{F}$ under the equivalence relation $A\tilde{}B$ if $P(A\Delta B) = 0$ with the metric $d(A, B) = P(A\Delta B)(= \int |I_A - I_B| \, dP$. We define set theoretic operations of unions, intersections and complements in Z in the obvious way. (e.g. $[A]^c = [A^c]$ where $[A]$ stands for the equivalence class of A). This space is the measure algebra associated with the probability space $(\Omega, \mathcal{F}, P)$. We note that if $\mathcal{F}$ is countable generated then the metric space Z is separable: if $A_1, A_2, \ldots$ generate $\mathcal{F}$ then the field generated by these sets is countable and dense. Conversely if there is a countable dense set $\{[A_n]\}$ in Z then, given $A \in \mathcal{F}$ there exist integers $n_1 < n_2 < \ldots$ such that $\int \left| I_A - I_{A_{n_k}} \right| dP < 1/k$. Thus $I_{A_{n_k}} \to I_A$ in L^1 and we also have a.e. convergence along a subsequence. It follows that $I_A = I_B$ a.e. for some B belonging to the sigma field generated by $A_1, A_2, \ldots$. We call the measure algebra separable when this last condition holds. Non-atomic means that $P(A) > 0$ implies $0 < P(B) < P(A)$ for some measurable set B contained in A. In this case $0 < a < \mu(A)$ implies $\mu(B) = a$ for some measurable set B contained in A. An isomorphism of measure algebras is a bijection that preserves complements and countable unions as well as measures.

Lemma 1

Let $\{A_{n,1}, A_{n,2}, \ldots, A_{n,k_n}\}, n = 1, 2, \ldots$ be a decreasing sequence of partitions of Ω by measurable sets and suppose this sequence is *dense* in the sense A measurable and $\epsilon > 0$ imply there exists a positive integer N and a set B which is a union of some of the sets $A_{N,1}, A_{N,2}, \ldots, A_{N,k_N}$ such that $P(A\Delta B) < \epsilon$. If P is non-atomic then $\max\limits_{1 \le j \le k_n} P(A_{n,j}) \to 0$.

[The partitions $\{A_{n,1}, A_{n,2}, ..., A_{n,k_n}\}, n = 1, 2, ...$ are said to be decreasing if each $A_{n,j}$ is a union of some of the sets $\{A_{n+1,1}, A_{n+1,2}, ..., A_{n+1,k_{n+1}}\}$. Note that $\{\max\limits_{1\leq j\leq k_n} P(A_{n,j})\}$ is decreasing in this case].

Proof: suppose, if possible, $\max\limits_{1\leq j\leq k_n} P(A_{n,j}) \downarrow \delta > 0$. There exists $j_1 \leq k_1$ such that $\max\limits_{1\leq j\leq k_n} P(A_{n,j} \cap A_{1,j_1}) \geq \delta$. [If this is false then for $1 \leq j \leq k_n$ we can choose k such that $A_{n,j} \subset A_1, j_1$ and hence $P(A_{n,j}) = P(A_{n,j} \cap A_{1,j_1}) < \delta$. This contradicts the fact that $\max\limits_{1\leq j\leq k_n} P(A_{n,j}) \geq \delta$]. Now, if $n \geq 2$ there exists $j_2 \leq k_2$ such that $\max\limits_{1\leq j\leq k_n} P(A_{n,j} \cap A_{1,j_1} \cap A_{1,j_2}) \geq \delta$. [The sets $A_{n,j} \cap A_{1,j_1}, 1 \leq j \leq k_n$ form a partition of A_{1,j_1} and $\max\limits_{1\leq j\leq k_n} P(A_{n,j} \cap A_{1,j_1}) \geq \delta$. The partition $A_{n,j} \cap A_{1,j_1}, 1 \leq j \leq k_n$ is finer than $A_{2,j} \cap A_{1,j_1}, 1 \leq j \leq k_2$ and hence the first step can be applied to the space A_{1,j_1} with the restriction of P to this set to conclude that there exists $j_2 \leq k_2$ with $\max\limits_{1\leq j\leq k_n} P(A_{n,j} \cap A_{1,j_1} \cap A_{1,j_2}) \geq \delta$. (The fact that $P(\Omega) = 1$ we not used in the first step)]. Repeating this we get integers $j_i \leq k_i, i = 1, 2, ...$ such that for any positive integer p $\max\limits_{1\leq j\leq k_n} P(A_{n,j} \cap A_{1,j_1} \cap A_{1,j_2} ... \cap A_{1,j_p}) \geq \delta$ if $n \geq p$. In particular $P(A_{1,j_1} \cap A_{1,j_2} ... \cap A_{1,j_p}) \geq \delta$ for all p and so $P(A_{1,j_1} \cap A_{1,j_2} \cap ...) \geq \delta$. Let $E = A_{1,j_1} \cap A_{1,j_2} \cap ...$ so that $P(E) \geq \delta$. By hypothesis there exists a measurable set $F \subset E$ such that $0 < P(F) < P(E)$. Let $0 < \epsilon < \min\{P(F), P(E) - P(F)\}$. By hypothesis there exists a positive integer N and a set B which is a union of some of the sets $A_{N,1}, A_{N,2}, ..., A_{N,k_N}$ such that $P(F \Delta B) < \epsilon$. However $E \subset B$ or $E \cap B = \emptyset$. [$E \subset A_{N,j_N}$ so for each $j \leq k_N$ either $E \subset A_{N,j}$ or $E \cap A_{N,j} = \emptyset$. Since B is a union of some of the sets $A_{N,1}, A_{N,2}, ..., A_{N,k_N}$ we get $E \subset B$ or $E \cap B = \emptyset$]. If $E \subset B$ we get $(F \subset B)$ and $P(B \backslash F) < \epsilon$ and $P(E) \leq P(B) < \epsilon + P(F) < \{P(E) - P(F)\} + P(F) = P(E)$ a contradiction. If $E \cap B = \emptyset$ then $F \cap B = \emptyset$ and we get $P(F) + P(B) = P(F \Delta B) < \epsilon \leq P(F)$ a contradiction again. This completes the proof.

Lemma 2

Let $X = [0, 1], \mathcal{B}$ the Borel sigma field and m the Lebesgue measure. If $\{0, x_{n,1}, x_{n,2}, ..., x_{n,k_n}\}$ is a sequence of partitions of $[0, 1]$ each finer than the previous one with $\max\limits_{2\leq j\leq k_n} \{x_{n,j} - x_{n,j-1}\} \to 0$ as $n \to \infty$ then the sequence of partitions $\{\{[x_{n,j-1}, x_{n,j}) : 1 \leq j \leq k_n\} : n = 1, 2, ...\}$ is dense.

[In order that we really have partitions of X we have to modify the interval $[x_{n,j-1}, x_{n,j})$ to $[x_{n,j-1}, x_{n,j}]$ when $j = k_n$. We shall ignore this (insignificant) point].

Proof: let $\epsilon > 0$ and choose N such that $\max\limits_{2\leq j\leq k_N} \{x_{N,j} - x_{N,j-1}\} < \epsilon/2$. Let (a, b) be any open interval contained in X. Let j_1 be such that $a \in [x_{N,j_1-1}, x_{N,j_1})$. Call $[x_{N,j_1-1}, x_{N,j_1})$ as E_1. If $b \notin E_1$ consider the intervals $[x_{N,j_1}, x_{N,j_1+1}), [x_{N,j_1+1}, x_{N,j_1+2}), ..., [x_{N,j_1+l-1}, x_{N,j_1+l})$ where l is such that $b \in [x_{N,j_1+l-1}, x_{N,j_1+l}) \backslash [x_{N,j_1+l}, x_{N,j_1+l+1})$. Let B be the union of the intervals

$[x_{N,j_1}, x_{N,j_1+1}), [x_{N,j_1+1}, x_{N,j_1+2}), ..., [x_{N,j_1+l-1}, x_{N,j_1+l})$,i.e. $B = [x_{N,j_1}, x_{N,j_1+l})$. Clearly $m((a,b)\Delta B) \leq 2 \max_{2 \leq j \leq k_N} \{x_{N,j} - x_{N,j-1}\} < \epsilon$. It follows from this that the conclusion holds if (a,b) is replaced by a finite disjoint union of half-closed intervals or, more generally, by a Borel set.

We are now ready to prove the isomorphism theorem. Let $\{E_n\}$ be dense in the measure algebra of a non-atomic separable probability space $(\Omega, \mathcal{F}, P)$. For fixed N the sets $\{A_1 \cap A_2 \cap ... \cap A_N : A_i = E_i \text{ or } A_i = E_i^c \text{ for each } i\}$ is partition of Ω and these partitions become finer and finer as N increases. Also, this sequence is dense. [The sigma algebras generated by these partitions are increasing, their union is an algebra and this algebra generates $\mathcal{F}$]. By Lemma1 we conclude that $\max_{1 \leq j \leq k_n} P(A_{n,j}) \to 0$ where the $N-th$ partition has been denoted by $\{A_{N,1}, A_{N,2}, ..., A_{N,k_N}\}$. We begin by partitioning $[0,1]$ by points $0 < x_{1,1} < x_{1,2} < ..., x_{1,k_1}$ with $x_{1,j} - x_{1,j-1} = P(A_{1,j})$ for each j. We then form a finer partition whose subintervals have lengths $P(A_{2,j}), 1 \leq j \leq k_2$ and so on. Lemma 2 can be applied to these partitions of [0,1]. We get an isometry between the measure algebras of $(\Omega, \mathcal{F}, P)$ and $([0,1], \mathcal{B}, m)$ by mapping partition elements as well as their unions to corresponding partition elements and their unions. Since these union form dense subsets of the appropriate spaces it is clear that the map extends to an isometry of the measure algebras which preserves complements and countable unions. It also preserves measures and the proof is complete.

An isomorphism theorem for measure spaces

Let μ be a Borel probability measure on $[0,1]$ such that $\mu\{x\} = 0 \; \forall x$. Then $([0,1], \mathcal{B}, \mu)$ is isomorphic to $([0,1], \mathcal{B}, m)$ where m is the Lebesgue measure.

Proof: let $F(x) = \mu(-\infty, x]), G(x) = \inf\{t \in \mathbb{R} : F(t) \geq x\}$ and $H(x) = \sup\{t \in \mathbb{R} : F(t) = x\}$. Then G is measurable and $m\{x : G(x) \leq t\} = F(t)$ $\forall t$. [This is standard; just verify that $G(x) \leq t$ iff $F(t) \geq x$]. We claim that G becomes a Borel isomorphism if we remove suitable null sets from $[0,1]$ and $\mathbb{R}$. First observe that $H(x) < G(y)$ if $0 < x < y < 1$. Indeed, F is continuous by hypothesis and the supremum and the infimum if the definitions of G and H are attained. If $G(y) = \alpha$ and $H(x) = \beta$ then $F(\alpha) = y$ (why?) and $F(\beta) = x < y = F(\alpha)$ which implies $\beta < \alpha$ as stated. Claim: $J \equiv \{x : G(x) < H(x)\}$ is at most countable. Note that $G \leq H$ (because $G(x) = \inf\{t \in \mathbb{R} : F(t) = x\}$) and if $x < y$ then the intervals $(G(x), H(x))$ and $(G(y), H(y))$ are disjoint because $H(x) < G(y)$. Since there can only be a countable number of disjoint open intervals in $\mathbb{R}$ the claim follows. Let $E = \bigcup_{x \in J} (G(x), H(x)]$. Note that $\mu((G(x), H(x)]) = F(H(x)) - F(G(x)) = x - x = 0$. It follows that $\mu(E) = 0$. Consider the map G from $[0,1] \backslash \{0,1\}$ to $\mathbb{R} \backslash E$. We claim that this is a Borel isomorphism from $([0,1], \mathcal{B}, m)$ to $(\mathbb{R}, \mathcal{B}, \mu)$. If $x < y$ then $G(x) \leq H(x) < G(y)$ so G is strictly increasing, hence one-to-one. If $G(a) \in (G(x), H(x)]$ then

$G(x) < G(a)$, hence $x < a$. But then $H(x) < G(a)$ a contradiction. Hence G maps $[0,1]\backslash\{0,1\}$ into $\mathbb{R}\backslash E$. Let $y \in \mathbb{R}\backslash E$. We shall show that $G(F(y)) = y$ proving that G is onto. $G(F(y)) \leq y$ by definition of G. Suppose $G(F(y)) < y$. Then $y \in (G(x), H(x)]$ where $x = F(y)$. [because $H(F(y)) = \sup\{t : F(t) = F(y)\} \geq y$]. But then $y \in E$, a contradiction. We have proved that G is a measurable bijection from $[0,1]\backslash\{0,1\}$ onto $\mathbb{R}\backslash E$. Measurability of its inverse follows from monotonicity. All that remains is to show that $m \circ G^{-1} = \mu$. It suffices to note that $F(x) = m\{t : G(t) \leq x\}$.

Corollary
If X is an uncountable Polish space and μ is a Borel probability measure on it such that $\mu\{x\} = 0\ \forall x$ then $(X, \mathcal{B}(X), \mu)$ is isomorphic to $([0,1], \mathcal{B}, m)$.

Proof: $(X, \mathcal{B}(X))$ is Borel isomorphic to $([0,1], \mathcal{B})$ and hence $(X, \mathcal{B}(X), \mu)$ is isomorphic as a measure space to $([0,1], \mathcal{B}, \nu)$ for some measure ν. Further $\nu\{x\} = 0\ \forall x$ and hence $([0,1], \mathcal{B}, \nu)$ is isomorphic to $([0,1], \mathcal{B}, m)$.

END OF APPENDIX

APPENDIX ON CHARACTER THEORY

FOURIER ANALYSIS ON GROUPS

Throughout G denotes a locally compact abelian (LCA) group and m denotes a Haar measure on G which is assumed to be a probability measure when G is compact.

$C_c(G)$ stands for the space of continuous complex valued functions on G with compact support and $C_0(G)$ stands for the space of continuous complex valued functions on G which vanish at ∞. $f \in C(G)$ vanishes at ∞ for every $\epsilon > 0$ there is a compact set K in G such that $|f(x)| < \epsilon$ if $x \notin K$. Equivalently f can be extended to the one-point compactification of G by defining it to be 0 at the "point at infinity".

We recall that $C_c(G)$ is dense in $L^p (= L^p(m))$ for $1 \leq p < \infty$.

Theorem
If $1 \leq p < \infty$ and $f_x(y) = f(yx^{-1})$ then $x \to f_x$ is a uniformly continuous map from G into L^p.
[For the definition of uniform continuity refer to the proof of the lemma below].

Proof: this follows by a standard argument using the following:

Lemma
Every function in $C_0(G)$ is unformly continuous.

Proof: let $f \in C_0(G)$. We have to show that given $\epsilon > 0$ there is a neighbourhood U of e in G such that $|f(x) - f(y)| < \epsilon$ whenever $x, y \in G$ and $xy^{-1} \in U$. Let K be a compact set such that $|f(x)| < \epsilon$ if $x \notin K$. For each $x \in K$ there is a neighbourhood U_x of e such that $|f(x) - f(y)| < \epsilon$ whenever $y \in xU_x$. Let V_x be a symmetric neighbourhood of e such that $V_x V_x \subset U_x$. There is a finite set $\{x_1, x_2, ..., x_N\}$ of K such that $K \subset x_1 V_{x_1} \cup x_2 V_{x_2} \cup ... \cup x_N V_{x_N}$. Let V be the intersection of $V_{x_1}, V_{x_2}, ..., V_{x_N}$. Then V is a symmetric neighbourhood of e. Let $xy^{-1} \in V$. Suppose $x \in K$. Then $x \in x_i V_{x_i}$ for some i. Hence $y \in Vx \subset V x_i V_{x_i} \subset x_i V_{x_i} V_{x_i} \subset x_i U_{x_i}$ and hence $|f(x_i) - f(y)| < \epsilon$. Since $x \in x_i V_{x_i} \subset x_i U_{x_i}$ we also have $|f(x_i) - f(x)| < \epsilon$. Thus $|f(x) - f(y)| < 2\epsilon$. Since V is symmetric we get the same conlcusion if $y \in K$. Now suppose $x \notin K$ and $y \notin K$. Then $|f(x) - f(y)| < \epsilon + \epsilon$.

Definition: if f and g are measurable and $\int \left|f(xy^{-1})g(y)\right| dy < \infty$ (where $dy$is an abbreviation for $dm(y)$) we define $(f * g)(x) = \int f(xy^{-1})g(y)dy$. Thus $(f * g)(x) = \int f_y(x)g(y)dy$. $f * g$ is called the convolution of f and g.

Theorem
If $\int \left|f(xy^{-1})g(y)\right| dy < \infty$ then $(f * g)(x) = (g * f)(x)$
If $f, g \in L^1$ then $f * g \in L^1$ too and $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$
If $f, g, h \in L^1$ then $(f * g) * h = f * (g * h)$
If $p, q \in (1, \infty), \frac{1}{p} + \frac{1}{q} = 1, f \in L^p$ and $g \in L^q$ then $f * g$ is defined at every point, it is continuous and vanishes at ∞.
For $f, g \in C_c(G)$ we have $f * g \in C_c(G)$ and $S_{f*g} \subset S_f S_g$ where $S_f . S_g, S_{f*g}$ are the supports of f, g and $f * g$ respectively.
If $f \in L^1$ then $T_f : L^2 \to L^2$ defined by $T_f(g) = f * g$ is well-defined and $\|T_f\| \leq \|f_1\|$.

Proof: the proof is standard. We just mention that continuity of $f * g$ uses previous theorem.

Remark: L^1 is a Banach algebra with convolution as multiplication. In general it does not have a multiplicative unit. Also there is no involution operation * such that $\|f^*\| = \|f\|$ and $\|f^* f\| = \|f\|^2$. Hence Gelfand -Naimark Theorem cannot be applied directly to this algebra. The map $f \to T_f$ defined above can be used to get a Banch algebra with involution by taking the closure in operator norm of $\{T_f : f \in L^1\}$.

Note that if G is discrete (i.e. the toplogy of G is the discrete topology) then $\frac{1}{m(\{e\})}I_{\{e\}}$ is a multiplicative unit. (Since $m(\{x\})$ is independent of x this number cannot be 0). It can be shown that if L^1 has a unit then G is necessarily discrete.

Definition: if $f \in L^1$ and $\gamma \in \hat{G}$ (i.e. $\gamma : G \to S^1$ is a continuous homomorphism) then the Fourier transform of f at γ is the number $\hat{f}(\gamma) = \int f(x)\bar{\gamma(x)}dx$.

Theorem [Characterization of algebra homomorphisms of L^1]

A function $\phi : L^1 \to \mathbb{C}$ is a non-zero algebra homomorphism (i.e. it is linear and $\phi(f * g) = \phi(f)(g)$ for all $f, g \in L^1$)

if and only if there exists $\gamma \in \hat{G}$ such that $\phi(f) = \hat{f}(\gamma)$ for all $f \in L^1$.

Proof: if ϕ has this form it is obviously linear and the fact that $\phi(f * g) = \phi(f)(g)$ for all $f, g \in L^1$ is proved easily using Fubini's Theorem. Note that if $\phi(f) = 0$ for all $f \in L^1$ then $\int f(x)\bar{\gamma(x)}dx = 0$ for all $f, g \in L^1$ which implies $\gamma(x) = 0$ a.e. However $\gamma(x) \neq 0$ for any x. Thus ϕ is non-zero. Conversely let ϕ be any non-zero algebra homomorphism of L^1. We claim that the linear map ϕ is bounded on L^1. For homomorphisms on a Banach algebra with unit this is a standard argument (ref. Rudin's Functional Analysis). For the general case we can adjoin a unit to L^1 and ϕ becomes the restriction to L^1 of a homomorphism on the new Banach algebra. The conclusion is $\|\phi\| \leq 1$. There exists, therefore, a function γ in L^∞ such that $\phi(f) = \int f\gamma$ for all $f \in L^1$. The equation $\phi(f * g) = \phi(f)(g)$ gives $\int\int f(xy^{-1})g(y)dy\gamma(x)dx = \int\int f(x)\gamma(x)g(y)\gamma(y)dydx$ or all $f, g \in L^1$. This implies $\int f(xy^{-1})\gamma(x)dx = \int f(x)\gamma(x)\gamma(y)dx$ a.e. for all $f \in L^1$. In particular $\int f_y(x)\gamma(x)dx = (\int f(x)\gamma(x)dx)\gamma(y)$ for some x. The fact that $y \to f_y$ is continuous (and $\|\phi\| \leq 1$ so $|\gamma(x)| \leq 1$ a.e.) shows γ is continuous on G. Now $\int f(z)\gamma(zy)dz = \int f(xy^{-1})\gamma(x)dx = \int f(x)\gamma(x)\gamma(y)dx$ for all $f \in L^1$ which implies (by Fubini's Theorem) that $\gamma(zy) = \gamma(z)\gamma(y)$ a.e. on $G \times G$. By continuity of γ this implies $\gamma(zy) = \gamma(z)\gamma(y)$ for all $z, y \in G$. Hence $\gamma(e) = 0$ or 1. If $\gamma(e) = 0$ then γ (and hence ϕ) is zero contradicting the hypothesis. Hence $\gamma(e) = 1$ and $\gamma(x^{-1}) = \frac{1}{\gamma(x)}$. Since $|\gamma(x)| \leq 1$ we get $|\gamma(x)| = 1$ for all x. Finally we note that $\gamma_0 = \bar{\gamma}$ has similar properties and $\phi(f) = \hat{f}(\gamma_0)$ for all f.

On $\hat{G}$ we consider the topology of uniform convergence on compact sets. A base for this topology consists of sets of the type $\{\gamma : \gamma(K) \subset U\}$ where K is compact in G and U is open in $\mathbb{C}$.

Theorem

The topology defined above is the smallest one that makes the functions $\gamma \to \hat{f}(\gamma)$ (f varying over L^1) continuous.

Proof: if a net $\{\gamma_i\}$ converges to γ uniformly on compact sets then $\hat{f}(\gamma_i) \to \hat{f}(\gamma)$ whenver $f \in C_c(G)$ because $\left|\hat{f}(\gamma_i) - \hat{f}(\gamma)\right| \leq \int |f(x)| \, |\gamma_i(x) - \gamma(x)| \, dx$ and $|\gamma_i(x) - \gamma(x)| \to 0$ uniformly on the support of f. If $f \in L^1$ is arbitrary and $\epsilon > 0$ there exists $g \in C_c(G)$ such that $\int |f - g| < \epsilon$. Now $\left|\hat{f}(\gamma_i) - \hat{f}(\gamma)\right| \leq \left|\hat{f}(\gamma_i) - \hat{g}(\gamma_i)\right| + |\hat{g}(\gamma_i) - \hat{g}(\gamma)| + \left|\hat{g}(\gamma) - \hat{f}(\gamma)\right| < \epsilon + |\hat{g}(\gamma_i) - \hat{g}(\gamma)| + \epsilon$ and $|\hat{g}(\gamma_i) - \hat{g}(\gamma)| \to 0$. We have proved that $\gamma \to \hat{f}(\gamma)$ continuous for all $f \in L^1$. Let τ be the smallest one that makes the functions $\gamma \to \hat{f}(\gamma)$ (f varying over L^1) continuous. We have to show that $\{\gamma : \gamma(K) \subset U\} \in \tau$ for any compact set K in G and any open set U in $\mathbb{C}$. Let $\gamma_0 \in \{\gamma : \gamma(K) \subset U\}$. We claim that $(\gamma, x) \to \gamma(x)$ is continuous on $(\hat{G}, \tau) \times G$: fix $f \in L^1$ and let $\gamma_0 \in \hat{G}, x_0 \in G$ and $\epsilon > 0$. There exists a neighbourhood U of x_0 and a τ-neighbourhood V of γ_0 such that $\|f_x - f_{x_0}\|_1 < \epsilon$ for $x \in U$ and $\left|(f_{x_0})\hat{}\,(\gamma) - (f_{x_0})\hat{}\,(\gamma_0)\right| < \epsilon$ for $\gamma \in V$. Now $\left|(f_x)\hat{}\,(\gamma) - (f_{x_0})\hat{}\,(\gamma_0)\right| \leq \left|(f_x)\hat{}\,(\gamma) - (f_{x_0})\hat{}\,(\gamma)\right| + \left|(f_{x_0})\hat{}\,(\gamma) - (f_{x_0})\hat{}\,(\gamma_0)\right| \leq \|f_x - f_{x_0}\|_1 + \left|(f_{x_0})\hat{}\,(\gamma) - (f_{x_0})\hat{}\,(\gamma_0)\right| < 2\epsilon$ for $x \in U$ and $\gamma \in V$. Now recall that $(f_x)\hat{}\,(\gamma) = \gamma(\overline{x})\hat{f}(\gamma)$. Hence $\left|\gamma(\overline{x})\hat{f}(\gamma) - \gamma_0(\overline{x_0})\hat{f}(\gamma_0)\right| < 2\epsilon$ if $x \in U$ and $\gamma \in V$. Choose f such that $\hat{f}(\gamma_0) = 1$ and then shrink V so that $\left|\hat{f}(\gamma) - \hat{f}(\gamma_0)\right| < \epsilon$ for $\gamma \in V$. Then $|\gamma(x) - \gamma_0(x_0)| \leq \left|\gamma(\overline{x})\hat{f}(\gamma) - \gamma(\overline{x})\right| + \left|\gamma(\overline{x})\hat{f}(\gamma) - \gamma_0(\overline{x_0})\hat{f}(\gamma_0)\right| \leq \left|\hat{f}(\gamma) - 1\right| + 2\epsilon < 3\epsilon$. This proves the claim.

If $x \in K$ then $\gamma_0(x) \in U$ and hence there exists a neighbourhood W_x of x and a $\tau-$neighbourhood V_x of γ_0 such that $\gamma(y) \in U$ whenever $\gamma \in V_x$ and $y \in W_x$. Let $K \subset W_{x_1} \cup W_{x_2} \cup ... \cup W_{x_k}$. Let $V = V_{x_1} \cap V_{x_2} \cap ... \cap V_{x_k}$. Then $V \subset \{\gamma : \gamma(K) \subset U\}$. [$\gamma \in V, x \in K$ imply there exists i such that $x \in W_{x_i}$; since $\gamma \in V_{x_i}$ we have $\gamma(x) \in U$]. We have proved that $\{\gamma : \gamma(K) \subset U\}$ is open in τ. The proof is now complete.

Theorem

For each $f \in L^1$ the function $f\hat{} \in C_0(\hat{G})$. $A \equiv \{\hat{f} : f \in L^1\}$ is dense in $C_0(G\hat{}\,)$ (for the sup norm). $g \in A$ and $x \in G$ implies $\gamma \to g(\gamma)\gamma(x)$ is in A and $\gamma \to g(\gamma\gamma_1)$ is in A. If $f \in L^1$ and $\gamma \in G\hat{}$ then $(f * \gamma)(x) = \gamma(\overline{x})\hat{f}(\gamma)$.

Proof: the fact that $\hat{f} \in C_0(\hat{G})$ is proved using Banach algebra theory. [In my notes on GelfandN.tex the following fact is proved: let Δ be the set of all non-zero complex homomorphisms on L^1 with the topology from $(L^1)^*$, i.e. $\phi_i \to \phi$

iff $\phi_i(f) \to \phi(f)$ for all $f \in L^1$. Then for each $f \in L^1$ the map $\phi \in \Delta \to \phi(f)$ belongs to $C_0(\Delta)$. This is proved by identifying (via Banach Alaoglu Theorem) the set $\Delta \cup \{0\}$ with the weak*-topology as the one-point compactification of Δ and noting that $\phi \to \phi(f)$ is continuous on $\Delta \cup \{0\}$. Define $\xi : \Delta \to \hat{G}$ by $\hat{f}(\xi(\phi)) = \phi(f)$. ξ is a bijection. Also ξ is continuous when $\hat{G}$ is given the topology of uniform convergence on compact sets and Δ is given the weak* topology: if $\phi_i(f) \to \phi(f)$ for all $f \in L^1$ and $\phi(f) \equiv \hat{f}(\gamma), \phi_i(f) \equiv \hat{f}(\gamma_i)$ then $\gamma_i \to \gamma$ uniformly on compact sets. This follows from the previous theorem. If $f \in L^1$ and $\epsilon > 0$ there exists a compact set K in Δ such that $|\phi(f)| < \epsilon$ if $\phi \notin K$. The set $C = \xi(K)$ is compact in $\hat{G}$ and $\gamma \notin C$ implies $\left|\hat{f}(\gamma)\right| = |\phi(f)| < \epsilon$ where ϕ is such that $\xi(\phi) = \gamma$ (so that $\hat{f}(\gamma) = \hat{f}(\xi(\phi)) = \phi(f)$). We have proved that the Fourier transform of any L^1 function vanishes at ∞]. To show that A is dense we note that A is a sub-algebra of $C_0(\hat{G})$. If $\gamma_1 \neq \gamma_2$ then there exists $f \in L^1$ such that $\hat{f}(\gamma_1) \neq \hat{f}(\gamma_2)$. Also if $f \in L^1$ and $g(x) = \bar{f}(x^{-1})$ then $\hat{g}(\gamma) = [\int f(x^{-1})\gamma(x)dx]^- = [\int f(y)\gamma(y^{-1})dy]^- = [\int f(y)\bar{\gamma}(y)dy]^- = [\hat{f}(\gamma)]^-$. It now follows by Stone-Weierstrass Theorem that A is dense on $C_0(G)$. Next $\hat{f}(\gamma)\gamma(x) = [\bar{\gamma}(x)f]^{\hat{}} \in A$. Also $\hat{f}(\gamma\gamma_1) = [f(\bar{\gamma}_1)]^{\hat{}}$. It remains to show that $(f * \gamma)(x) = \gamma(x)\hat{f}(\gamma)$. The left side is $\int f(xy^{-1})\gamma(y)dy = \int f(z)\gamma(z^{-1}x)dz = \int f(z)\gamma(z^{-1})\gamma(x)dz$

$$= \int f(z)\bar{\gamma}(z)dz\gamma(x) = \gamma(x)\hat{f}(\gamma).$$

Theorem

$\hat{G}$ is an LCA group (under pointwise multiplication) and sets of the type $\{\gamma : |1 - \gamma(x)| < \delta \text{ for all } x \in K\}$ where K is a compact subset of G and $\delta > 0$ form a neighbourhood base at 1.

Proof: only thing that requires a proof is local compactness. Let U be a neighborhood of e whose closure is compact. Consider $V = \{\gamma : |1 - \gamma(x)| < \delta$ for all $x \in \bar{U}\}$. We prove that this neighbourhood of 1 is relatively compact.

Lemma

Let $\epsilon > 0$. There exists a positive integer N such that $c \in S^1$ and $c, c^2, .., c^N$ have real parts strictly positive imply $|c - 1| < \epsilon$.

Suppose this is false for some $\epsilon > 0$. Then there exists a sequence $\{c_n\}$ in S^1 such that $c_n, c_n^2, .., c_n^n$ have real parts strictly positive but $|c_n - 1| \geq \epsilon$. If c is a limit point of this sequence then $\operatorname{Re} c^k \geq 0$ for every positive integer k and $|c - 1| \geq \epsilon$. If c is not a root of unity then $\{c, c^2, ...\}$ is dense in S^1 and hence $\operatorname{Re} c^{k_j} \to \operatorname{Re}(-1) = -1$ for some $k_j \uparrow \infty$ which is a contradiction. Hence there is a least $N \geq 2$ such that $c^N = 1$. The numbers $c, c^2, ..., c^N$ are distinct and they are all $N - th$ roots of 1. Hence every $N - th$ root of 1 has positive real part. This is a contradiction because Re $e^{2\pi i \frac{N//2}{N}} < 0$ if N is even and Re $e^{2\pi i \frac{(N-1)//2}{N}} < 0$ if N is odd. [Note that $\pi - \pi/N \in (\pi/2, \pi)$].

Back to the proof of the theorem: we assume that δ is so small that $c \in S^1, |c-1| < \delta \Rightarrow \operatorname{Re} c > 0$. Let $\epsilon > 0$ and choose a positive integer N as in the lemma. Let W be a neighbourhood of e such that $W^N \subset U$. If $\gamma \in V$ and $x, y \in W$ then $x^j, y^j \in U$ for $1 \le j \le N$. Hence $\left|1 - \gamma^j(x)\right| < \delta$ and $\left|1 - \gamma^j(y)\right| < \delta$ for $1 \le j \le N$. This implies $|1 - \gamma(x)| < \epsilon$ and $|1 - \gamma(y)| < \epsilon$ so $|\gamma(x) - \gamma(y)| < 2\epsilon$. Let $\{\gamma_i\}$ be a net in V. If $D = \{c \in \mathbb{C} : |c-1| \le \delta\}$ then $V \subset D^{\bar{U}}$ so by Tychonoff's Theorem there is a subnet $\{\gamma_{i_j}\}$ converging pointwise (say to ϕ) on $\bar{U}$.

Claim: $\{\gamma_{i_j}\}$ converges uniformly on any compact set K in G. If this is false we may suppose (by going to a subnet) $\left|\gamma_{i_j}(x_j) - \phi(x_j)\right| \ge 6\epsilon > 0$ for some net $\{x_j\}$ in K and some $\epsilon > 0$. Let W be as above for this ϵ. By going to a further subnet we may suppose $x_j \to x$ (say). Now $\left|\gamma_{i_j}(x_j) - \gamma_{i_j}(x)\right| < 2\epsilon$ for $j \ge$ some j_0 (because $x x_j^{-1} \in W$). Observe that since $|\gamma(x) - \gamma(y)| < 2\epsilon$ whenever $xy^{-1} \in W$ and $\gamma \in V$ it follows that $|\phi(x) - \phi(y)| < 2\epsilon$ whenever $xy^{-1} \in W$. Hence $|\phi(x_j) - \phi(x)| < 2\epsilon$ for $j \ge$ some j_1. Now $\left|\gamma_{i_j}(x_j) - \phi(x_j)\right| \le \left|\gamma_{i_j}(x_j) - \gamma_{i_j}(x)\right| + \left|\gamma_{i_j}(x) - \phi(x)\right| + |\phi(x_j) - \phi(x)| < 5\epsilon$ for $j \ge$ some j_3 contradicting the fact that $\left|\gamma_{i_j}(x_j) - \phi(x_j)\right| \ge 6\epsilon$. We have proved that every net in V has a subnet thatg converges uniformly on compact sets. [It may be noted that ϕ is a continuous homorphism into S^1, i.e. $\phi \in \hat{G}$].

Theorem

In G sets of the form $\{x : |1 - \gamma(x)| < r \ \forall \gamma \in C\}$ where C is compact in $\hat{G}$ and $r > 0$ are neighbourhoods of e. In $\hat{G}$ sets of the form $\{\gamma : |1 - \gamma(x)| < r \ \forall \gamma \in K\}$ where K is compact in G and $r > 0$ form a neighbourhood base at e.

Proof: the second part is obvious. Let $x_0 \in V \equiv \{x : |1 - \gamma(x)| < r \ \forall \gamma \in C\}$. The map $\gamma \to \gamma(x_0)$ is continuous on $\hat{G}$ and C is compact in $\hat{G}$ and so $\sup\{|1 - \gamma(x_0)| : \gamma \in C\}$ is attained. Hence $\sup\{|1 - \gamma(x_0)| : \gamma \in C\} < r$. Let $0 < \epsilon < r - \sup\{|1 - \gamma(x_0)| : \gamma \in C\}$. For each $\gamma_0 \in C$ we have $|1 - \gamma_0(x_0)| < r$. By the continuity of the map $(x, \gamma) \to \gamma(x)$ on $\hat{G} \times G$ established above we see that there exists neighbourhoods V, W_{γ_0} of x_0 and γ_0 respectively such that $x \in V, \gamma \in W_{\gamma_0} \Rightarrow |\gamma(x) - \gamma_0(x_0)| < \epsilon$. The neighbourhods $W_{\gamma_0}, \gamma_0 \in C$ form an open cover of C. Hence there exists a finite set $\{\gamma_1, \gamma_2, ..., \gamma_k\}$ such that C is covered by $W_{\gamma_i}, 1 \le i \le k$. Now $\gamma \in C$ implies $\gamma \in W_{\gamma_i}$ for some i and hence $|\gamma(x) - \gamma_i(x_0)| < \epsilon$. This gives $|\gamma(x) - 1| \le \epsilon + |1 - \gamma_i(x_0)| \le \epsilon + \sup\{|1 - \gamma(x_0)| : \gamma \in C\} < r$. Hence $V \subset \{x : |1 - \gamma(x)| < r \ \forall \gamma \in C\}$ and the proof is complete.

Theorem [Bochner]

Let $\phi : G \to \mathbb{C}$ be continuous at e and positive definite. Then there exist a positive regular Borel measure μ on $\hat{G}$ such that $\phi(x) = \int\limits_{G\hat{}} \gamma(x)d\mu(\gamma)$ for all $x \in G$.

Proof: standard arguments give the following: $\phi(x^{-1}) = \bar{\phi}(x), |\phi(x)| \le \phi(e)$ and $|\phi(x) - \phi(y)|^2 \le 2\phi(e)\operatorname{Re}[\phi(e) - \phi(xy^{-1})]$. Thus ϕ is bounded, uniformly continuous and $\phi(e) > 0$ (unless $\phi \equiv 0$ in which case we can take $\mu = 0$). Without loss of generality we may suppose $\phi(e) = 1$. Let $f \in C_c(G)$ and let K be its support and $\epsilon > 0$. Then $f(x)\bar{f}(y)\phi(xy^{-1})$ is uniformly continuous on $K \times K$. It is easy to see that there exists a partition $\{A_1, A_2, ..., A_k\}$ of K such that $\sum\limits_{i,j=1}^{k} f(x_i)\bar{f}(x_j)\phi(x_i - x_j)m(E_i)m(E_j)$ differs from $\int\int f(x)\bar{f}(y)\phi(xy^{-1})dxdy$ by at most ϵ. It follows that the double integral here is non-negative for any $f \in C_c(G)$, hence for any $f \in L^1(G)$. Let $T_\phi(f) = \int\limits_G f\phi$ $(f \in L^1(G))$. Recall that for each $f \in L^1$ the function $\hat{f} \in C_0(\hat{G})$ and $A \equiv \{\hat{f} : f \in L^1\}$ is dense in $C_0(\hat{G})$. Define $S_\phi : A \to \mathbb{C}$ by $S_\phi(f\hat{}) = T_\phi(f)$. We are going to show that $|T_\phi(f)| \le \left\|f\hat{}\right\|_\infty$ which shows that $\hat{f} = \hat{g}$ implies $T_\phi(f) = T_\phi(g)$. Thus S_ϕ is well-defined. Also $\left|S_\phi(\hat{f})\right| = |T_\phi(f)| \le \left\|\hat{f}\right\|_\infty$ so S_ϕ extends to a bounded linear functional on $C_0(\hat{G})$ (with the sup norm). Hence there exists a regular Borel measure μ on $\hat{G}$ such that $S_\phi(\hat{f}) = \int\limits_{G\hat{}} \hat{f}(\gamma)d\mu(\gamma)$ for all $f \in L^1$. This means $\int\limits_{G\hat{}} \hat{f}(\gamma)d\mu(\gamma) = \int\limits_G f\phi$ for all $f \in L^1$. Hence $\int\limits_{G\hat{}}\int f(x)\bar{\gamma}(x)dxd\mu(\gamma) = \int\limits_G f\phi$ and since $f \in L^1$ is arbitrary this gives $\int\limits_{G\hat{}} \bar{\gamma}(x)d\mu(\gamma) = \phi(x)$ almost everywhere. Clearly both sides are continuous and hence the equality holds at every point x. To complete the proof replace μ by $\mu \circ \tau^{-1}$ where $\tau : \hat{G} \to \hat{G}$ is the map $\gamma \to \frac{1}{\gamma}$.

It remains to prove that $|T_\phi(f)| \le \left\|\hat{f}\right\|_\infty$ for all $f \in L^1$. Define $< f, g >' = \int\int f(x)\bar{g}(y)\phi(xy^{-1})dxdy$. It is easy to verify that $< f, g >' = T_\phi(f * g\tilde{})$ where $g\tilde{}(x) = \bar{f}(x^{-1})$. Note that $< f, g >'$ is linear in f conjugate linear in g and $< f, f >' \ge 0$. Cauchy - Schwartz inequality holds under these conditions and so $|< f, g >'| \le \sqrt{< f, f >'}\sqrt{< g, g >'}$. Let U be an open set such that $e \in U, \bar{U}$ is compact and $U^{-1} = U$. Let $g = \frac{1}{m(U)} I_U$. We have $< f, g >' = \frac{1}{m(U)} \int_U \int f(x)\phi(xy^{-1})dxdy = \frac{1}{m(U)} \int_U \int f(x)\{\phi(xy^{-1}) - \phi(x)\}dxdy + \int f(x)\phi(x)dx$. Hence $|< f, f >' - T_\phi(f)| < \epsilon$ if U is suitable chosen. Also

$< g,g >'= \frac{1}{m^2(U)}\int_U\int_U \phi(xy^{-1})dxdy = 1 + \frac{1}{m^2(U)}\int_U\int_U \{\phi(xy^{-1})-1\}dxdy$ so $|< g,g >'-1| < \epsilon$ if U is appropriately chosen. Applying $|< f,g >'| \leq \sqrt{< f,f >'}\sqrt{< g,g >'}$ and letting $\epsilon \to 0$ yields $|T_\phi(f)| \leq \sqrt{< f,f >'} = \sqrt{T_\phi(f * f^{\tilde{}})}$. Now T_ϕ is a bounded linear functional on L^1 corresponding to the element ϕ of L^∞ and $\|\phi\|_\infty \leq 1$ so $\|T_\phi\| \leq 1$. If $g = f * f^{\tilde{}}, g_2 = g * g^{\tilde{}}, g_{n+1} = g_n * g_n^{\tilde{}} (n \geq 2)$ then $|T_\phi(f)|^2 \leq T_\phi(g) \leq \sqrt{T_\phi(g_2)}$

$\leq .... \leq \{T_\phi(g_n)\}^{1/2^{n-1}}$. Hence $|T_\phi(f)|^2 \leq \|g_n\|_1^{1/2^{n-1}}$. A simple exercise is to show that $(f * f^{\tilde{}})^{\tilde{}} = f * f^{\tilde{}}$. This shows that g_n is simply the "2^{n-1} -th power" of g w.r.t. convolution operation. By the spectral radius formula we get $|T_\phi(f)|^2 \leq \rho(g)$ where $\rho(g)$ is the spectral radius of g. By Banach algebra theory we have $\rho(g) \leq \left\|\hat{f}\right\|_\infty$. [We use two theorems from Rudin's Functional Analysis: let $X = L^1(G) \times \mathbb{C}$ with the multiplication operation $(f,c)(g,d) = (fg + cg + df, cd)$ and the norm $\|(f,c)\| = \|f\|_1 + |c|$. Then X is a commutative Banach algebra with unit $(0,1)$. Fix $f \in L^1$ and consider the element $(f,0)$ of X. By Theorem 10.13, page 235 of Rudin we have $\|(f,0)^n\|^{1/n} \to \rho((f,0))$ (the spectral radius of $(f,0)$. This gives $\lim \|f^n\|_1^{1/n} = \rho(f,0)$. Now if $c \in \sigma((f,0))\backslash\{0\}$ then, by Theorem 11.5 page 265 of Rudin, there exists a complex homomorphism Φ of X such that $c = \Phi((f,0))$. Define $\phi : L^1 \to \mathbb{C}$ by $\phi(g) = \Phi((g,0))$ for any $g \in L^1$. Then ϕ is a non-zero complex homomorphism on L^1 and $c = \Phi((f,0)) = \phi(f)$. But any non-zero complex homomorphism ϕ of L^1 is of the type $\phi(g) = g^{\hat{}}(\gamma)$ for some $\gamma \in \hat{G}$ and, conversely, any γ yields a complex homomorphism of L^1. Hence, if $\sigma((f,0))\backslash\{0\}$ is non-empty then $\rho((f,0)) = \sup\{|c| : c \in \sigma((f,0)) \leq \sup\{|\phi(f)| : \phi$ is a complex homomorphism of $L^1\} = \sup\{\left|f^{\hat{}}(\gamma)\right| : \gamma \in \hat{G}\} = \left\|f^{\hat{}}\right\|_\infty$. This inequality also holds if $\sigma((f,0))\backslash\{0\}$ is empty.]. Identifying elements ξ of $\Delta_{L^1(G)}$ with the functions $f \to \hat{f}(\gamma)$ with $\gamma \in \hat{G}$ we get $\rho(f) = \sup\{\left|\hat{f}(\gamma)\right| : \gamma \in \hat{G}\} = \left\|f^{\hat{}}\right\|_\infty$. Finally we get $|T_\phi(f)|^2 \leq \|\hat{g}\|_\infty$ and since $(f^{\tilde{}})^{\hat{}} = [\hat{f}]^-$ we get $|T_\phi(f)|^2 \leq \|\hat{g}\|_\infty \leq \left\|\hat{f}\right\|_\infty \left\|(f^{\tilde{}})^{\hat{}}\right\|_\infty = \left\|\hat{f}\right\|_\infty^2$. It remains to show that μ is a positive measure. We have $1 = \phi(e) = \int_{G^{\hat{}}} \gamma(e)d\mu(\gamma) = \mu(\hat{G})$ and $|\mu|(\hat{G}) = \|T_\phi\| \leq 1$. We can write $\mu(E) = \int_E f d|\mu|$ with $|f| = 1$ a.e. $[|\mu|]$. We have $1 = \mu(\hat{G}) = \int_{G^{\hat{}}} f d|\mu|$ which implies $1 = \int_{G^{\hat{}}} \operatorname{Re} f d|\mu|$. Hence $1 = \int_{G^{\hat{}}} \operatorname{Re} f d|\mu| \leq \int |f| d|\mu| = |\mu|(\hat{G}) = 1$ which gives $\operatorname{Re} f = 1$ a.e. and hence $\operatorname{Im} f = 0$ a.e.. Thus $f = 1$ a.e. and $\mu = |\mu|$.

Theorem

The measure μ in Bochner's Theorem is unique.

Proof: we claim that $\int\limits_{G^\wedge} \gamma(x)d\mu(\gamma) = 0$ for all x (where μ is a complex measure) implies $\mu = 0$. If $f \in L^1$ then $\int\limits_{G^\wedge} \hat{f}(x)d\mu(\gamma) = \int\limits_{G^\wedge}\int\limits_{G} f(x)\bar{\gamma}(x)dxd\mu(\gamma) \int\limits_{G}\int\limits_{G^\wedge} \bar{\gamma}(x)d\mu(\gamma)f(x)dx = 0$ (because $\bar{\gamma}(x) = \gamma(x^{-1})$). But $\{\hat{f} : f \in L^1\}$ is dense in $C_0(\hat{G})$ so $\mu = 0$.

Theorem [Fourier Inversion Theorem]

Let $M = \{f \in \mathbb{C}^G : f(x) = \int\limits_{G^\wedge} \gamma(x)d\mu(\gamma)\ \forall x \in G$ for some regular Borel measure μ on $\hat{G}\}$. Then

1) $\hat{f} \in L^1$ for all $f \in L^1 \cap M$

2) if $f \in L^1 \cap M$ then we can write $f(x) = \int\limits_{G^\wedge} \hat{f}(\gamma)\gamma(x)d\gamma$ for all x provided Haar measure on $\hat{G}$is suitably normalized.

Proof : for $f \in L^1 \cap M$ we write μ_f for a measure which satisfies the equation $f(x) = \int\limits_{G^\wedge} \gamma(x)d\mu_f(\gamma)\ \forall x \in G$. Note that $(g * f)(e) = \int g(x^{-1})f(x)dx = \int g(x^{-1})\int\limits_{G^\wedge} \gamma(x)d\mu_f(\gamma)dx = \int\limits_{G^\wedge}\int g(x^{-1})\gamma(x)dxd\mu_f(\gamma) = \int\limits_{G^\wedge}\int g(z)\bar{\gamma}(z)dzd\mu_f(\gamma)$

$= \int\limits_{G^\wedge} \hat{g}(\gamma)d\mu_f(\gamma)\ \forall g \in L^1$ and, changing g to $g * h$ (where $h \in L^1 \cap M$)

$\int\limits_{G^\wedge} \hat{h}(\gamma)\hat{g}(\gamma)d\mu_f(\gamma) = ((h*g)*f)(e) = ((g*f)*h)(e) = \int\limits_{G^\wedge} \hat{f}(\gamma)\hat{g}(\gamma)d\mu_h(\gamma)$. From this we conclude that $\hat{h}(\gamma)d\mu_f(\gamma) = \hat{f}(\gamma)d\mu_h(\gamma)$. [$\{\hat{g} : g \in L^1\}$ is dense in $C_0(\hat{G})$ as proved earlier]. This equation holds for all $f, h \in L^1 \cap M$. Now let $h \in C_c(\hat{G})$ and K be its support. For each $\gamma \in K$ there exists $f \in C_c(G)$ with $\hat{f}(\gamma) \neq 0$. Now $(f * f^\sim)^\wedge(\gamma) = \left|\hat{f}(\gamma)\right|^2 > 0$. Also $(f * f^\sim)^\wedge(\gamma_1) = \left|\hat{f}(\gamma_1)\right|^2 \geq 0$ for all γ_1. Since $\gamma_1 \to f^\wedge(\gamma_1)$ is continuous (because convergence in $\hat{G}$ implies uniform convergence on the support of f) each $\gamma \in K$ has a neighbourhood on which $(f * f^\sim)^\wedge$ is positive. These neighbourhoods form an open cover. Extracting a finite subcover we get $f_1, f_2, ..., f_k \in C_c(G)$ and $\gamma_1, \gamma_2, ..., \gamma_k \in K$ such that the Fourier transform of $\psi = \sum_{i=1}^{k} f_i * f_i^\sim$ is positive on K. Note that $\psi \in C_c(G)$. We claim that $\psi \in M$. For this we apply Bochner's Theorem. First note that $f * f^\sim$ is positive definite for any $f \in L^1$. [In fact $\sum_{i=1}^{k} c_i\bar{c_j}(f * f^\sim)(x_i x_j^{-1}) =$

$\int \left| \sum_{i=1}^{k} c_i f(x_i y) \right|^2 dy]$. Hence ψ is continuous and positive definite. Bochner's Theorem implies that $\psi \in M$. Let $Th = \int\limits_{G\hat{}} \frac{h}{\hat{\psi}} d\mu_\psi$ where μ_ψ is defined by $\psi(x) = \int\limits_{G\hat{}} \gamma(x) d\mu_\psi(\gamma)$. We claim that T is well defined on $C_c(G)$: $\psi_1 \in C_c(G), \hat{\psi}_1 > 0$ on K_1 and $\psi_1(x) = \int\limits_{G\hat{}} \gamma(x) d\mu_{\psi_1}(\gamma)$. We have to show that $\int\limits_{G\hat{}} \frac{h}{\hat{\psi}} d\mu_\psi = \int\limits_{G\hat{}} \frac{h}{\hat{\psi}_1} d\mu_{\psi_1}$. Since $\psi, \psi_1 \in M$ we have $\hat{\psi}(\gamma) d\mu_{\psi_1}(\gamma) = \hat{\psi}_1(\gamma) d\mu_\psi(\gamma)$. Multiplying both sides by $\frac{h}{\hat{\psi}\hat{\psi}_1}$ and integrating we get $\int\limits_{G\hat{}} \frac{h}{\hat{\psi}} d\mu_\psi = \int\limits_{G\hat{}} \frac{h}{\hat{\psi}_1} d\mu_{\psi_1}$. Thus T is a well-defined linear map on $C_c(G)$. Note that μ_ψ is a positive measure. Hence T is a positive linear functional on $C_c(G)$. Claim: T is not identically 0. If $f \in L^1 \cap M$ then $T(hf\hat{}) = \int \frac{h\hat{f}}{\hat{\psi}} d\mu_\psi = \int \frac{h\hat{\psi}}{\hat{\psi}} d\mu_f = \int h d\mu_f$. To prove the claim it suffices to show that $\int h d\mu_f \neq 0$ for some $h \in C_c(G)$ and some $f \in L^1 \cap M$. If this is false then $\mu_f = 0$ for every $f \in L^1 \cap M$. The equation $(g * f)(e) = \int\limits_{G\hat{}} \hat{g}(\gamma) d\mu_f(\gamma)$ $\forall g \in L^1, \forall f \in L^1 \cap M$ shows that $(g*f)(e) = 0$, i.e. $\int g(y) f(y^{-1}) dy = 0 \; \forall g \in L^1$ so $L^1 \cap M = \{0\}$ in which case there is nothing to prove. Hence T is not the zero functional. Now fix $h \in C_c(\hat{G})$ and $\gamma \in \hat{G}$. Let $\psi \in C_c(G)$ be constructed as above with $\psi \in K$ replaced by $K \cup (K\gamma)$ (so $\hat{\psi} > 0$ on K as well as on $K\gamma$). Let $f(x) = \frac{\psi(x)}{\gamma(x)}$. Then $\hat{f}(\gamma_1) = \int \frac{\psi}{\gamma(x)} \bar{\gamma}_1(x) dx = \hat{\psi}(\gamma\gamma_1)$ and $\mu_\psi(A) = \mu_f(\frac{1}{\gamma} E)$. [Define $\tau : \hat{G} \to \hat{G}$ by $\tau(\gamma_1) = \gamma\gamma_1$. Then $\int \gamma_1(x) d\mu_\psi(\gamma_1) = \psi(x) = \gamma(x) f(x) = \gamma(x) \int \gamma_1(x) d\mu_f(x) = \int \tau(\gamma_1)(x) d\mu_f(x) = \int \gamma_1(x) d\mu_f \circ \tau^{-1}(x)$. Uniqueness of the measure in Bochner's theorem gives $\mu_\psi = \mu_f \circ \tau^{-1}]$. Let $h_0(\gamma_1) = h(\frac{\gamma_1}{\gamma})$. Then $Th_0 = \int\limits_{G\hat{}} \frac{h_0}{\hat{\psi}} d\mu_\psi = \int\limits_{G\hat{}} h(\frac{\gamma_1}{\gamma}) \frac{1}{\hat{\psi}(\gamma_1)} d\mu_\psi(\gamma) = \int\limits_{G\hat{}} h(\gamma_1) \frac{1}{\hat{\psi}(\gamma\gamma_1)} d\mu_f(\gamma) = \int\limits_{G\hat{}} \frac{h(\gamma_1)}{f\hat{}\,\backslash(\gamma_1)} d\mu_f(\gamma) = Th$. In other words $Th = Th_\gamma$ where h_γ is the translate of h by γ. T is, therefore, a translation invariant positive linear functionsal on $C_c(G)$ and hence $Th = \int h d\nu$ for an appropriate choice of Haar measure ν on

$\hat{G}$. If $f \in L^1 \cap M$ and $h \in C_c(\hat{G})$ then $\int_{\hat{G}} h d\mu_f = T(h\hat{f}) = \int_{\hat{G}} h\hat{f} d\nu(\gamma)$. This yields $d\mu_f = \hat{f} d\nu$ which in turn proves that $\hat{f} \in L^1(\hat{G})$.

Finally we have $f(x) = \int \gamma(x) d\mu_f(\gamma) = \int \gamma(x)\hat{f}(\gamma) d\nu(\gamma)$ if $f \in L^1 \cap M$. This completes the proof.

Corollary

Sets of the type $\{x \in G : |1-\gamma(x)| < \delta \ \forall\gamma \in K\}$ where $\delta > 0$ and K is compact in $\hat{G}$ form a local base at e. Also $\hat{G}$ separates points of G.

Proof: we already know that the sets here are open. Let $x \neq e$. There exists an open set U such that $x \notin U, e \in U$. There exists a compact neighborhood W of e such that $W^{-1} = W$ and $WW \subset U$. Let $f = \frac{1}{\sqrt{m(W)}} I_W$ and $g = f * f^{\tilde{}}$. Then g is continuous with support in WW. Also g is positive definite. By Bochner's Theorem it follows that $g \in L^1 \cap M$. Hence $g(x) = \int_{\hat{G}} \hat{g}(\gamma)\gamma(x) d\gamma$ for all x. In particular $g(e) = \int_{\hat{G}} \hat{g}(\gamma) d\gamma$. However $g(e) = \int f(y^{-1})\bar{f}(y) dy = \frac{1}{m(W)} m(W) = 1$. Thus $\int_{\hat{G}} \hat{g}(\gamma) d\gamma = 1$ and $\hat{g} = \left|\hat{f}\right|^2 \geq 0$. There is a compact set K in $\hat{G}$ such that $\int_K \hat{g}(\gamma) d\gamma > 2/3$. If $y \in G$ and $|1-\gamma(y)| < 1/3 \ \forall\gamma \in K$ we claim that $y \in U$. This would certainly show that every neighbourhhod of e contains $\{x \in G : |1-\gamma(x)| < \delta \ \forall\gamma \in K\}$ for some $\delta > 0$ and some compact set K in $\hat{G}$ proving the first part. It also shows that $\gamma(x) \neq 1$ for some $\gamma \in K$. [Otherwise $|1-\gamma(x)| < 1/3 \ \forall\gamma \in K$ so $x \in U$ which is a contradiction. Thus $x \neq e$ implies $\gamma(x) \neq 1$ for some $\gamma \in \hat{G}$. If $x \neq y$ then there exists $\gamma \in \hat{G}$ such that $\gamma(xy^{-1}) \neq 1$ which implies $\gamma(x) \neq \gamma(y)$. It remains to show that if $y \in G$ and $|1-\gamma(y)| < 1/3 \ \forall\gamma \in K$ then $y \in U$. We have $g(y) = \int_{\hat{G}} \hat{g}(\gamma)\gamma(y) d\gamma = \int_{\hat{G}\setminus K} \hat{g}(\gamma)\gamma(y) d\gamma + \int_K \hat{g}(\gamma)\gamma(y) d\gamma = c + d$ (say). Note that $|c| < 1/3$ because $\int_{K^c} \hat{g}(\gamma) d\gamma = 1 - \int_K \hat{g}(\gamma) d\gamma < 1/3$. Also $\operatorname{Re} d = \operatorname{Re} \int_K \hat{g}(\gamma)\gamma(y) d\gamma = \int_K \hat{g}(\gamma) \operatorname{Re}\gamma(y) d\gamma$ and $\operatorname{Re}\gamma(y) > 2/3$ so $\operatorname{Re} d > (2/3)^2$. It follows that $\operatorname{Re}(c+d) = \operatorname{Re} d + \operatorname{Re} c > (2/3)^2 - 1/3 = 1/9$ so $\operatorname{Re} g(y) > 1/9$. In particular $g(y) \neq 0$. Since g has support in $WW \subset U$ we have proved that $y \in U$.

Remarks: let G be an LCA group and H a closed subgroup of G. Give G/H the quotient topology (namely the smallest topology that makes the quotient map $\pi(x) = xH$ continuous). Then G/H is also an LCA group. If $x \in G\backslash H$ then xH is a non-zero element of G/H. Hence there exists a character ξ of G/H such that $\xi(xH) \neq 1$. Defining $\gamma \in G\hat{}$ by $\gamma = \xi \circ \pi$ we get $\gamma(y) = 1$ for all $y \in H$ but $\gamma(x) \neq 1$. We prove later that any character on a closed subgroup of G can be extended to a character on G.

Corollary

If G is a compact abelian group then the set of all trigonometric polynomials (i.e. functions of the type $\sum_{i=1}^{n} c_i\gamma_i$ where $n \in \mathbb{N}, c_i's \in \mathbb{C}$ and $\gamma_i's \in \hat{G}$) is dense in $C(G)$ with respect to the supremum norm.

Proof: this follows immediately from previous corollary and Stone - Weierstrass Theorem.

Remark: for any $p \in [1, \infty)$ $C(G)$ is dense in L^p so trigonometric polynomials are dense in L^p.

In the following theorem we write L^p for $L^p(G)$.

Theorem [Plancherel]

The map $f \to \hat{f}$ maps $L^1 \cap L^2$ into $L^2(\hat{G})$ and it is an isometry. It extends to an isometric isomorphism of L^2 onto $L^2(\hat{G})$.

Proof: let $f \in L^1 \cap L^2$ and $g = f * f\tilde{}$. Then g is positive definite. It is continuous and integrable. [Convolution of any two L^2 functions is continuous]. By Bochner's Theorem $g \in L^1 \cap M$. Hence the inversion formula $g(x) = \int \hat{g}(\gamma)\gamma(x)d\gamma$ holds for all x. Hence $g(x) = \int \left|\hat{f}(\gamma)\right|^2 \gamma(x)d\gamma$. Put $x = e$ to get $\int |f|^2 = (f * f\tilde{})(e) = g(e) = \int \left|\hat{f}(\gamma)\right|^2 d\gamma$. Hence $f \to \hat{f}$ maps $L^1 \cap L^2$ into $L^2(\hat{G})$ and it is an isometry. Since $C_c(G)$ is dense in L^2 it follows that $L^1 \cap L^2$ is dense in L^2. Hence $f \to \hat{f}$ extends to an isometric isomorphism of L^2 onto its range. Let $S = \{\hat{f} : f \in L^1 \cap L^2\}$. To complete the proof we have to show that S is dense in $L^2(\hat{G})$. Suppose $h \in L^2(\hat{G})$ is orthogonal to S. Note that $f \in L^1 \cap L^2 \Rightarrow \bar{\gamma}(x)\hat{f}(\gamma) \in S$ for each $x \in G$. [This is because $f_x \in L^1 \cap L^2$ and $(f_x)\hat{}(\gamma) = \bar{\gamma}(x)\hat{f}(\gamma)$]. Hence $f \in L^1 \cap L^2$ implies $\int h(\gamma)\bar{\gamma}(x)\hat{f}(\gamma)d\gamma = 0$ for all x. Since $h \in L^2(\hat{G})$ and $\hat{f} \in L^2(\hat{G})$ we have $h\hat{f} \in kL^1(\hat{G})$. However if μ is a complex Borel measure on $\hat{G}$ such that $\int \gamma(x)d\mu(\gamma) = 0$ for all x then $\mu = 0$. Hence $h\hat{f} = 0$ a.e. If $\gamma \in \hat{G}$ then there exists $f \in L^1 \cap L^2$ such that $\hat{f}(\gamma_1) \neq 0$ for all γ_1 in a neighbourhood of γ. [There exists $\phi \in C_c(\hat{G})$ such that $\phi = 1$ in a neighbourhood U of γ. Since $\{f\hat{\ } : f \in L^1\}$ is dense in $C_0(\hat{G})$ there exists a sequence $\{f_n\} \subset L^1$ such that $\hat{f}_n \to \phi$ uniformly. There exists $\{g_n\} \subset L^1 \cap L^2$ such that $\|f_n - g_n\|_1 < 1/n$. Now $\left|\hat{f}_n - \hat{g}_n\right|(\gamma) \leq \|f_n - g_n\|_1 < 1/n$. It follows that $\hat{g}_n \to \phi$ uniformly. Hence $|\hat{g}_n| > 1/2$ in U if n is sufficiently large. In particular $\hat{g}_n \neq 0$ on U. We can take $f = g_n$]. Hence $h = 0$ a.e. in a neighbourhood of each point γ. This implies $h = 0$ a.e.: if $K \subset \hat{G}$ is compact then $h = 0$ a.e. on K as seen by a straightforward compactness argument. If $E = \{\gamma : h(\gamma) \neq 0\}$ then $m(E \cap K) = 0$ for each compact set K. Regularity of Haar measure now shows that $m(E) = 0$. This completes the proof.

Theorem
$\{\hat{f} : f \in L^1\} = \{\phi * \psi : \phi, \psi \in L^2(\hat{G})\} \qquad (*)$

Proof: If $f \in L^1$ we can write $f = gh$ where $g, h \in L^2$. [Take $g = \sqrt{|f|}, h(x) = f(x)/g(x)$ if $g(x) \neq 0, 0$ if $g(x) = 0$]. By above theorem $\hat{g}$ and $\hat{h} \in L^2$. Claim: $\hat{f} = \hat{g} * \hat{h}$. If we prove this it would follow that the left side of $(*)$ is contained in the right side. To prove the claim note that an isometry preserves inner products. So we have the Parseval Formula $\int g(x)\bar{h}(x)dx = \int \hat{g}(\gamma_1)[\hat{h}(\gamma_1)]^- d\gamma_1$. Replace $h(x)$ by $\gamma(x)\bar{h}(x)$ to get $\int g(x)h(x)\bar{\gamma}(x)dx = \int \hat{g}(\gamma_1)\hat{h}(\gamma\gamma_1^{-1})d\gamma_1$ (because the Fourier transform of $\gamma(x)\bar{h}(x)$ at γ_1 is $[\hat{h}(\gamma\gamma_1^{-1})]^-$). Since $(\hat{g} * \hat{h})(\gamma) = \int \hat{g}(\gamma\gamma_1^{-1})\hat{h}(\gamma_1)d\gamma_1 = \int \hat{g}(\gamma_2)\hat{h}(\gamma\gamma_2^{-1})d\gamma_2$ we get $\int g(x)h(x)\bar{\gamma}(x)dx = (\hat{g} * \hat{h})(\gamma)$. The claim is proved.

It remains to show that the right side of $(*)$ is contained in the left side. If $\phi, \psi \in L^2(\hat{G})$ we can write $\phi = \hat{g}$ and $\psi = \hat{h}$ for some $g, h \in L^2$. Hence $\phi * \psi = \hat{g} * \hat{h} = f\hat{\ }$ where f is the L^1 function gh. This completes the proof.

Corollary
Let U be a non-empty open set in $\hat{G}$. There exists $f \in L^1$ such that $\hat{f} = 0$ on U^c but $\hat{f}$ is not identically 0.

Proof: there exists a compact set $K \subset U$ such that $m(K) > 0$. There exists an open set V containing e such that $\overline{V}$ is compact and $K + V \subset U$. By Plancherel Theorem there exist $g, h \in L^2$ such that $\hat{g} = I_K$ and $\hat{h} = I_V$. Since $g, h \in L^2$ we have (by the proof of above theorem) $\hat{g} * \hat{h} = \hat{f}$ where $f = gh$. Clearly, $f \in L^1, \hat{f} = 0$ on $(K+V)^c$ (because $\hat{g} = 0$ on K^c and $\hat{h} = 0$ on V^c) hence on U^c but $\int f\hat{\ }(\gamma)d\gamma = \int (\hat{g} * \hat{h})(\gamma)d\gamma = \int \hat{g}(\gamma)d\gamma \int \hat{h}(\gamma)d\gamma = m(K)m(V) > 0$ so $\hat{f}$ is not identically 0.

Theorem [Pontryagin Duality Theorem]
Let G be an LCA group and define $\theta : G \to G\hat{\ }\hat{\ }$ by $\theta(x)(\gamma) = \gamma(x)$. Then θ is a group isomorphism as well as a homeomorphism (onto $G\hat{\ }\hat{\ }$).

Proof: it is trivial to check that θ is a group isomorphism onto the range, say H, which is a subgroup of $G\hat{}\hat{}$. We first prove that this map is a homeomorphism onto its range. To complete the proof we then show that H is both dense and closed in $G\hat{}\hat{}$. Let $K \subset \hat{G}$ be compact, $r > 0$ and

$U = \{x \in G : |1 - \gamma(x)| < r \ \forall \gamma \in K\}, V = \{\Phi \in G\hat{}\hat{} : |1 - \Phi(\gamma)| < r \ \forall \gamma \in K\}$. We have proved that these are typical sets from neighbourhood bases at the identity in G and $G\hat{}\hat{}$ respectively. Note that $\theta^{-1}(V \cap \theta(G)) = U$. This proves that θ and θ^{-1} are continuous at the identity elements of G and $G\hat{}\hat{}$ respectively. Hence θ is a homeomorphism onto its range. As a consequence of this the range of θ is locally compact. The next lemma shows that a locally compact sugbroup of a locally compact abelian group is necessarily closed. Hence H is closed. All that remains is to show that H is dense in $G\hat{}\hat{}$. We apply previous corollary with G changed to $G\hat{}$ taking U to be $G\hat{}\hat{} \backslash \theta(G)$. Assuming that U is non-empty we shall arrive at a contradiction. There exists $h \in L^1(\hat{G})$ such that $\hat{h}$ is not identically 0 but $\hat{h}$ is 0 on $U^c = \theta(G)$. We have $\hat{h}(\theta(x)) = \displaystyle\int h(\gamma)[\theta(x)(\gamma)]^- d\gamma = \int h(\gamma)\bar{\gamma}(x) d\gamma = 0$ for all x. Changing x to x^{-1} gives $\displaystyle\int h(\gamma)\gamma(x) d\gamma = 0$ for all x. Since the only complex Borel mesure μ with $\displaystyle\int \gamma(x) d\mu(\gamma) = 0$ for all x is the zero measure we get $h = 0$. Hence $\hat{h} = 0$, a contradiction.

Lemma

Let G be a Hausdorff topological group. Let H be a subgroup of G which is also locally compact in the relative topology from G. Then H is closed in G.

Proof: let U be a neighbourhood of e in H whose closure (in H) is compact. Let $U = H \cap V$ with V open in G and $A = V \cup (Cl_H(U) \backslash U\}$ where $Cl_H(U)$ is the closure of U in H. Then $A \cap H = Cl_H(U)$ which is compact, hence closed in G. Let W be an open set in G such $W = W^{-1}$ and $WW \subset V$ and $e \in W$. [Possible because V is a neighbourhood of e in G]. If $x \in \bar{H}$ then $x^{-1} \in \bar{H}$ too because $\bar{H}$ is a subgroup. Hence $x^{-1}W \cap H \neq \emptyset$. Let $y \in x^{-1}W \cap H$. If $yx \notin A \cap H$ then there exists a neighbourhood U_1 of yx which does not intersect $A \cap H$. the neighbourhood $y^{-1}U_1 \cap xW$ of x must contain a point z of H. Thus $z \in y^{-1}U_1 \cap xW \cap H$ and so $yz \in yxW \subset WW$ (because $y \in x^{-1}W$) and $yz \in A$. Also y and $z \in H$. Thus $yz \in A \cap H \cap U_1$ a contradiction since U_1 does not intersect $A \cap H$. Thus $yx \in A \cap H$. Since $y \in H$ this gives $x \in H$. This completes the proof.

Corollary

If μ is a regular Borel measure on G such that $\displaystyle\int \gamma(x) d\mu(x) = 0$ for all $\gamma \in \hat{G}$ then $\mu = 0$.

Proof: this follows immediately from an earlier result if we think of μ as a measure on $G\hat{}\hat{}$.

Corollary [Inversion]

If $f \in L^1$ and $\hat{f} \in L^1(\hat{G})$ then $f^{\hat{}\hat{}}(\theta(x)) = f(x)$ a.e.

Proof: we prove that if μ is a regular Borel measure on G and $\hat{\mu} \in L^1(\hat{G})$ (where $\hat{\mu}(\gamma) = \int \bar{\gamma}(x) d\mu(x)$) then $\frac{d\mu}{dm} = \int \hat{\mu}(\gamma)\gamma(x) d\gamma$. The corollary follows from this by defining $\mu(A)$ as $\int\limits_A f dm$. Note that $\hat{\mu} \in L^1(\hat{G}) \cap M(\hat{G})$ because $\hat{\mu}$ is continuous and it is a linear combination of at most 4 positive definite functions. [$M(\hat{G})$ is defined the way M was defined earlier with G replaced by $\hat{G}$]. Let $f(x) = \int \hat{\mu}(\gamma)\gamma(x) d\gamma$. The inversion formula proved earlier states that $\mu^{\hat{}\hat{}}(x) = f(x^{-1}) \in L^1$ and $\hat{\mu}(\gamma) = \int\limits_G f(x^{-1})\gamma(x) dx$ for all γ. [We have used Pontryagin Theorem here]. The proof is complete.

Extending characters from closed subgroups to whole groups

If H is a closed subgroup of an LCA group G then G/H is an LCA group when it is given the qutient topology. Let $\pi : G \to G/H$ be the qutient map so that a set E is open in G/H iff $\pi^{-1}(E)$ is open in G.

z
Theorem
Any character on a closed subgroup H of G can be extended to a character on G.

Before we can prove this we need some preliminaries. The fact that G/H is an LCA group is easy to see. [Just use the fact that the quotient map $\pi : G \to G/H$ is continuous and open]. Let $A = \{\gamma \in \hat{G} : \gamma(x) = 1 \ \forall x \in H\}$. Note that A is a closed subgroup of $\hat{G}$ (hence an LCA group in the relative topology). Define $\phi_0 : (G/H)\hat{} \to A$ by $\phi_0(\xi)(x) = \xi(xH)$ for all $x \in G$. Clearly this is a well-defined map with range in A. We claim that ϕ_0 is a group isomorphism and a homeomorphism. It is obviously injective. If $\gamma \in A$ then $\xi(xH) = \gamma(x)$ gives a well-defined element of $(G/H)\hat{}$: continuity of ξ follows from the fact that $\xi \circ \pi (= \gamma)$ is continuous. Thus ϕ_0 is a group isomorphism. Basic neighbourhoods of the identity in $(G/H)\hat{}$ and A are of the type $V = \{\xi : |1 - \xi(xH)| < \delta \ \forall xH \in C\}$ and $W = \{\gamma \in A : |1 - \gamma(x)| < \delta \ \forall x \in D\}$ where C is compact in G/H , D is compact in G and $\delta > 0$. Observe that if $C = \pi(D)$ then $V = \{\xi : |1 - \xi \circ \pi(x)| < \delta \ \forall x \in D\} = \{\xi : |1 - \phi_0(\xi)(x)| < \delta \ \forall x \in D\} = \phi_0^{-1}(W)$. Hence, to show that ϕ_0 is a homeomorphism it suffices to show that C is compact in G/H if and only if there is a compact set D

in G such that $C = \pi(D)$. The "if" part is obvious. Suppose C is compact in G/H. For each $x \in \pi^{-1}(C)$ there is an open set U_x such that $x \in U_x$ and $\bar{U}_x$ is compact. Since $C \subset \bigcup_{x \in \pi^{-1}(C)} \pi(U_x)$ (and π is an open map) there exists a finite set $\{x_1, x_2, ..., x_k\} \subset \pi^{-1}(C)$ such that $C \subset \bigcup_{i=1}^{k} \pi(U_{x_i})$. Let $D = \pi^{-1}(C) \cap [\bar{U}_{x_1} \cup \bar{U}_{x_2} \cup ... \cup \bar{U}_{x_k}]$. Clearly D is compact. If $y \in C$ then $y \in \pi(U_{x_i})$ for some i so $y = \pi(x)$ for some $x \in U_{x_i}$. We then have $x \in D$ and $y = \pi(x)$ so $C \subset \pi(D)$. Of course $\pi(D) \subset \pi(\pi^{-1}(C)) \subset C$. We have proved that $\phi_0 : (G/H)\hat{} \to A$ is an isomorphism and a homeomorphism. By the next lemma $H = \{x : \gamma(x) = 1 \ \forall \gamma \in A\}$. Replacing G by $\hat{G}, H$ by A we conclude that $(\hat{G}/A)\hat{}$ is isomorphic and homeomorphic to $\theta(H)$ where θ is the Pontryagin map from G onto $G\hat{}\hat{}$. The isomorphism $\phi : (\hat{G}/A)\hat{} \to \theta(H)$ is defined by $(\phi(\tau))(\gamma) = \tau(\gamma A)$ for all $\tau \in (\hat{G}/A)\hat{}$.

Now let $\gamma_0 \in \hat{H}$. Consider the map $F : (\hat{G}/A)\hat{} \to \mathbb{C}$ defined by $F(\Phi) = \gamma_0[(\theta^{-1} \circ \phi)(\Phi)]$. This is well-defined because $\phi(\Phi) \in \theta(H)$ and $(\theta^{-1} \circ \phi)(\Phi) \in H$. Clearly $F \in (\hat{G}/A)\hat{}\hat{}$. Let $\theta_0 : (\hat{G}/A) \to (\hat{G}/A)\hat{}\hat{}$ be the Pontryagin map. Since θ_0 is surjective there exists $\gamma \in \hat{G}$ such that $\theta_0(\gamma A) = F$. We claim that γ is an extension of γ_0. Let $\tau = \phi^{-1}(\theta(x))$ where $x \in H$ is fixed. In the definition of F put $\Phi = \phi^{-1}(\theta(x))$. Then $\gamma_0[(\theta^{-1} \circ \phi)(\Phi)] = F(\Phi) = [\theta_0(\gamma A)](\Phi) = \Phi(\gamma A) = [\phi^{-1}(\theta(x))](\gamma A)$. The left side of this is $\gamma_0(\theta^{-1}(\theta(x)) = \gamma_0(x)$. The proof will be complete if we show that the right side of the equation, i.e. $[\phi^{-1}(\theta(x))](\gamma A)$ is $\gamma(x)$. Recalling that $\tau = \phi^{-1}(\theta(x))$ we get $\phi(\tau) = \theta(x)$ so $\phi(\tau)(\gamma) = \gamma(x)$ by the definition of θ. Now $\phi(\tau)(\gamma) = \tau(\gamma A)$ by the definition of ϕ. Hence $[\phi^{-1}(\theta(x))](\gamma A) = \tau(\gamma A) = \phi(\tau)(\gamma) = \gamma(x)$. The proof is now complete.

Lemma

If H is a closed subgroup of G and $A = \{\gamma \in \hat{G} : \gamma(x) = 1 \ \forall x \in H\}$ then $H = \{x \in G : \gamma(x) = 1 \ \forall \gamma \in A\}$.

Proof: G/H is an LCA group. If $x \notin H$ then $xH \neq eH$ so there exists $\xi \in (G/H)\hat{}$ such that $\xi(xH) \neq 1$ and $\xi \circ \pi$ (where $\pi : G \to G/H$ is the projection map) gives an element γ_0 of $\hat{G}$ such that $\gamma \equiv 1$ on H but $\gamma(x) \neq 1$. Thus $\gamma \in A$ but $\gamma(x) \neq 1$. This proves that $\{x \in G : \gamma(x) = 1 \ \forall \gamma \in A\} \subset H$. The reverse inclusion is obvious.

Theorem

If G is a compact abelian metric group then $\hat{G}$ is countable and $\hat{G}$ is an orthonormal basis for L^2.

Proof: if $\gamma_1, \gamma_2 \in \hat{G}$ then $\int \gamma_1(x)\bar{\gamma}_2(x)dx = 0$ if $\gamma_1 \neq \gamma_2$ and 1 if $\gamma_1 = \gamma_2$. [The Haar measure is normalized so as to make it a probability measure]. Indeed if γ is a character which is non-constant then $\int \gamma(x)dx = \int \gamma(yx)dx =$

$\gamma(y)\int\gamma(x)dx$ and we can choose y such that $\gamma(y)\neq 1$, so, taking $\gamma=\frac{\gamma_1}{\gamma_2}$ we get $\int\gamma_1(x)\bar{\gamma}_2(x)dx=0$]. Since $C(G)$ is separable (in sup norm) and dense in L^2 we see that L^2 is separable too. Hence the orthonormal set $\hat{G}$ must be at most countable. If $f\in L^2$ and $\int f(x)\bar{\gamma}(x)dx=0$ for every character γ then $f=0$ because $\|f\|_2=\left\|\hat{f}\right\|_2$. [Note that Haar measure is finite so $f\in L^2$ implies $f\in L^1$ and so $\hat{f}(\gamma)=\int f(x)\bar{\gamma}(x)dx$]. This proves the theorem.

END OF APPENDIX

TABLE OF CONTENTS

The number in brackets denotes the page number.