Application of Ergodic Theory to *Uniform distribution* mod 1

Oleg Ivrii

February 13, 2008

Chapter 1

Ergodic Theory

1.1 The Setting

Our setting will vary, but (X, μ) will be some measure space and T a measure preserving transformation, that is the measure of a set equals to the measure of its pre-image (for all measurable sets E, we have $\mu(E) = \mu(T^{-1}(E))$). All measures will be probability measures (have total mass 1) unless explicitly stated otherwise.

To make the reader more comfortable with the notation, we give a proof of a theorem due to Poincaré:

Theorem 1. *(Multiple Recurrence) Suppose $A \subset X$ be measurable, then for almost all points of A, there is some index $n > 0$ for which $T^n(x) \in A$ (consequently, infinitely many such values exist).*

Proof. Let B denote the set of points x in A which do not recur, that is $T^k(x) \notin A$ for all $k > 0$. The sets B, $T^{-1}(B)$, $T^{-2}(B)$, $\ldots$ have equal measure and are disjoint: if y is both in $T^{-p}(B)$ and $T^{-q}(B)$ $(q < p)$, then $T^p(y)$ is both in B and $T^{-(q-p)}(B)$, i.e $T^p(y)$ recurs in $q - p$ steps. It follows that each of the sets B, $T^{-1}(B)$, $T^{-2}(B)$, $\ldots$ have measure 0. $\square$

Suppose f is some function X to $\mathbb{R}$. We start at some initial point x and consider its T-orbit: $x, Tx, T^2x, \ldots$, we can apply the function f to these points to create a bunch of numbers and average their partial sums, and take limit to form the "time average" (LHS of the equation below) of point x.

We are interested on the dependence of the the time average on the point we start with; more specially, we are interested whether the time average should be equal to the "space average" (RHS):

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k<n} f(T^k(x)) =_? \int_X f(x)d\mu$$

The answer turns out to be so provided that T is *ergodic*, i.e if all T-invariants have measure 0 or 1 (a measurable set $E \subset X$ is called T-*invariant* if $T^{-1}(E) = E$).

Problem 1. *Equivalently the transformation T is ergodic with respect to (X, μ) if the only μ-invariant functions $f \in L^2$ are the constant functions (hint: consider sets of type $f(x) > a$).*

We will call $U = U_T$ the operator which sends function f to $f \circ T$. If f happens to be in L^1, then Uf is also in L^1 and

$$\int_X f(x)d\mu = \int_X f(T(x))d\mu.$$

Indeed, the above equation is true for characteristic functions; by linearity its true for simple functions; by monotone convergence, it is true for non-negative functions and so true for general functions as well. In particular, U_T is an isometry on L^1.

Problem 2. *Show that $U_T : L^p(X, \mu) \to L^p(X, \mu)$ is an isometry for $1 \le p \le \infty$.*

1.2 Birkhoff's Theorem

Theorem 2 (Birkhoff)**.** *Suppose T is measure preserving transformation of a σ-finite measure space X, and f is some function in L^1. Then $\overline{f} = \lim_{n\to\infty} \frac{1}{n}\sum_{k<n} f(T^k(x))$ exists almost everywhere, is μ-integrable and T-invariant. If additionally, $\mu(X)$ is finite, then $\int_X \overline{f}(x)d\mu = \int_X f(x)d\mu$.*

Suppose we have a sequence of real numbers $(a_i)_{i=0}^N$, we call index k an m-*leader* if there is a $j < m$ such that

$$a_k + a_{k+1} + \cdots + a_{k+j} \ge 0.$$

The smallest such j is called the *breakout point*. We claim that the sum of all m-leaders is always non-negative. It suffices to split m-leaders into blocks such that each block has non-negative sum. This is based on a very simple observation: suppose a negative number is an m-leader, then the next index is an m-leader as well. In fact all numbers between this negative number and its breakout point are m-leaders.

The idea is the following: consider the first number: if its non-negative, put it in a box; if it is negative and an m-leader, put a box around it and its breakout point and continue on with the rest of the sequence. Then all m-leaders will be put in boxes with non-negative total sum. Below is an example with $m = 4$:

$$\boxed{5} \quad \boxed{0} \quad \boxed{4} \quad -3 \quad \boxed{-2 \quad 4} \quad \boxed{-1 \quad 2} \quad -8 \quad \boxed{-4 \quad 6} \quad -2 \quad \boxed{-3 \quad 4} \quad -1$$

While Birkhoff's theorem is applicable to complex-valued functions as well, there is no loss of generality to assume the function is real-valued. We do so. Our proof of Birkhoff's theorem is based on the following:

Theorem 3 (Maximal Ergodic Theorem)**.** *Suppose f is real-valued, and A is the set for which some average $\frac{1}{n}\sum_{k<n} f(T^k(x))$ is non-negative, then $\int_A f d\mu \geq 0$.*

Proof. We will refer to A as the "maximal set" of f. We can describe A as the set of points x for which some initial average of $\{f_k(x)\}_{k=0}^{\infty}$ is non-negative (we will sometimes write $f_k(x)$ for $f(T^k x)$). Let $A_m \subset A$ be the set of points x, for which at most m-th initial average is non-negative. By the monotone convergence theorem, it suffices to show that the integral over each A_m is non-negative.

Fix m. Let n be an arbitrary positive integer. For each point x, we consider the m-leaders of $T_0(x), T_1(x), \dots T_{n+m-1}(x)$, and denote their sum by $s(x)$. If E_k is the set of points x for which $f_k(x)$ is an m-leader of $T_0(x), T_1(x), \dots T_{n+m-1}(x)$, then $s(x) = \sum_{k=0}^{n+m-1} f_k(x) \cdot \chi_{E_k}$. Then,

$$\int_X s(x) d\mu = \sum_{k=0}^{n+m-1} \int_{E_k} f_k(x) d\mu \geq 0.$$

We observe that "$Tx \in E_{k-1}$" is equivalent to "$x \in E_k$" (for $1 \leq k \leq n-1$ anyway), thus $E_k = T^{-1}E_{k-1}$, and so $E_k = T^{-k}E_0$. Therefore for $1 \leq k \leq n-1$,

$$\int_{E_k} f_k(x) d\mu = \int_{T^{-k}E_0} f(T^k x) d\mu =_? \int_{E_0} f(x) d\mu.$$

The change of variables is justified by the discussion preceding Problem 2. Hence

$$\int_X s(x) d\mu \leq n \int_{E_0} f(x) d\mu + m||f||_1.$$

As this works for any n, $\int_{E_0} f(x) = \int_{A_m} f(x)$ must be non-negative. $\square$

Proof of Birkhoff's Theorem

Proof. For $a < b$, let $Y = Y(a, b)$ be the set of points x for which

$$\liminf \frac{1}{n} \sum_{j=0}^{n-1} f_j(x) < a < b < \limsup \frac{1}{n} \sum_{j=0}^{n-1} f_j(x).$$

Clearly, Y is measurable and T-invariant; we want to show it has measure 0. Without loss of generality, we may suppose $b > 0$, otherwise $a < 0$ and we can work with $-f$ and $-a$. We want to apply the maximal ergodic theorem to $f - b \cdot \chi_Y$ and $a \cdot \chi_Y - f$ to obtain

$$\int_Y (f(x) - b)dx \geq 0, \qquad \int_Y (a - f(x))dx \geq 0.$$

Adding the two inequalities, we see that Y must have measure 0. But we do not yet know that these are integrable, so we must first show that Y has *finite* measure.

As Y is σ-finite, there exists measurable sets Y_j which increase to Y and have finite measure. It suffices to prove that each Y_j have uniformly bounded measure. We may apply the maximal ergodic theorem to $f - b \cdot \chi_{Y_j}$. We obtain that $\int_F (f - b\chi_{Y_j})dx \geq 0$ where F is the set where some average of $f - b\chi_{Y_j}$ is non-negative. If $x \in Y$, then $b < \limsup \frac{1}{n} \sum_{j=0}^{n-1} f_j(x)$, this condition is satisfied and so $Y \subset F$. Hence

$$||f||_1 = \int_X |f(x)| \geq \int_F f(x)dx \geq \int_F b\chi_{Y_j}(x)dx = b \cdot \mu(Y_j \cap F) = b \cdot \mu(Y_j).$$

Applying the result to all rational pairs (a, b), we see that for almost every x, the time average actually tends to some limit function. Moreover, as

$$\int_X \Big| \frac{1}{n} \sum_{j<n} f_j(x) \Big| dx \leq \frac{1}{n} \int_X \Big| \sum_{j<n} f_j(x) \Big| dx \leq \frac{1}{n} \sum_{j<n} \int_X \Big| f_j(x) \Big| dx = ||f||_1,$$

the limit function $\overline{f}$ is integrable. Also $\overline{f}$ is invariant (the Césaro limit does not depend on first few terms). To proceed further, we must assume that X has finite total measure. In this case, we can also say that the integrals of f and $\overline{f}$ are equal.

Suppose $\overline{f}(x) \geq a$ almost everywhere, then for each ε, some sum $\sum_{j<n}(f_j(x) - a + \varepsilon)$ is non-negative; hence $\int f(x)dx \geq (a - \varepsilon)m(X)$. As this happens for all ε, $\int f(x)dx \geq am(x)$. Similarly, if $\overline{f}(x) \leq b$ almost everywhere, then $\int f(x)dx \leq b\mu(X)$.

Let $X(k, n)$ be the set of points x for which $\overline{f}(x)$ is between $k/2^n$ and $(k+1)/2^n$. We obtain

$$\frac{k}{2^n} \mu(X(k, n)) \leq \int_{X(k,n)} \overline{f}(x)dx \leq \frac{k+1}{2^n} \mu(X(k, n)).$$

By the previous paragraph, the last statement holds for $\int_{X(k,n)} f(x)dx$ as well. It follows that

$$-\frac{1}{2^n}\mu(X(k,n)) \leq \int_{X(k,n)} f - \overline{f} \leq \frac{1}{2^n}\mu(X(k,n)).$$

Summing over k, we obtain that

$$\Big|\int f(x)dx - \int \overline{f}(x)dx| \leq \frac{1}{2^n}\mu(X).$$

But as n is arbitrary, the RHS side can be made as small as we want. $\square$

Remark. *The condition that $\mu(X)$ is finite is necessary. For instance, consider the interval $(0,1]$ with measure $\mu = x^{-1}dm$ where m is the Lebesgue measure and $T : x \to x/2$. Here, the partial averages of $\chi_{[1/2,1]}$ tend pointwise to 0, but their integrals are equal to* $\ln 2$*.*

The following is an easy corollary (and useful form) of Birkhoff:

Theorem 4 (Space Average = Time Average)**.** *Now, we return back to our initial setting with $\mu(X) = 1$, we claim that T is ergodic if and only*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} f(T^k(x)) = \int_X f(x)d\mu.$$

for almost all points $x \in X$ (such points are called generic points*) and all integrable functions f.*

Proof. For an ergodic transformation T, $\overline{f} = C$ must be a constant almost-everywhere. Birkhoff's theorem tells us that $\int_X f(x)d\mu = \int_X \overline{f}d\mu = C$ which is exactly what we wanted to show. Conversely, if T wasn't ergodic, we would have an invariant set E of measure strictly between 0 and 1 and χ_E would fail the above formula. $\square$

Remark. *The problem with ergodicity is that it gives results of an "almost everywhere" nature. However, nothing can be said about any specific point $x \in X$. To obtain results for* all *rather than almost all points, we would need to place much stronger condition: namely, that of* unique ergodicity*. More on this next chapter.*

Chapter 2

Weyl's Lemma

2.1 Ergodic Measures

Previously, we had a space with a measure and were interested whether a transformation was ergodic with respect to that measure. Now we change the setting: the transformation T is given and we interested to find measures for which it is measure-preserving.

We will denote the set of all Borel probability measures on X by $\mathcal{M}$ endowed with the weak-$*$ topology. We would need to know that $\mathcal{M}$ is compact, this happens when X is a compact metric space:

Theorem 5. *If X is a compact metric space, then $\mathcal{M}$ is compact.*

Proof. The space of continuous functions on X is separable, let $\mathcal{F}$ be a countable dense set. Given a sequence of probability measures, $\mu_1, \mu_2, \dots$ and $f \in \mathcal{F}$, the integrals $\int_X f d\mu_n$ are bounded (say by $||f||_\infty$), so there is a subsequence μ_{n_j} for which $\int_X f d\mu_{n_j}$ converge. By a diagonalization argument, there is a subsequence u_{n_j} for which $\int_X f d\mu_{n_j}$ converge for all $f \in \mathcal{F}$. Then automatically $\lim_{j\to\infty} \int_X g d\mu_{n_j}$ exists for all $g \in C(X)$ (if $||f-g||_\infty < \epsilon$ then $|\int_X f d\mu_{n_j} - \int_X g d\mu_{n_j}| < \epsilon$ as well).

Thus $\lim_{j\to\infty} \int_X g d\mu_{n_j}$ defines a positive linear function on $C(X)$ with norm 1: obviously the norm is at most 1 for $\lim_{j\to\infty} \int_X g d\mu_{n_j} \leq ||g||_\infty$. Whats important here is the other direction: for that set $g \equiv C$ (a constant function). Thus by the Riesz representation theorem, this functional is given by a positive probability measure μ, which means that $\mu_{n_j} \to \mu$ in the weak-$*$ topology. $\square$

The next problem shows that we can't do away with the compactness of X:

Problem 3. *Take $X = \mathbb{R}^n$. Construct a sequence of probability measures μ_n which tend to the zero measure in the weak-$*$ topology.*

The subset of T-invariant measures will be denoted $\mathcal{M}_T$. By problem 2, it is easy to see that $\mathcal{M}_T$ is closed (and therefore compact). It is also non-empty. Indeed, the sequence $\mu_n = \frac{1}{n}\sum_{k<n}\delta_{T^k x}$ must have an accumulation point which necessarily must be a T-invariant measure.

The theorem on compact-convex sets states that a compact convex set is the convex hull of its extreme points, i.e the ones which cannot be represented as non-trivial linear combinations of (two) other points. We have a nice characterization of ergodic measures:

Theorem 6. *Ergodic T-invariant measures are precisely the extreme points of $\mathcal{M}_T$.*

Proof. If μ is not ergodic, then there is a invariant set A with $0 < \mu(A) < 1$. We can define measure μ_A by restricting μ to A and normalizing so the total measure would be 1. Similarly, we can define the measure $\mu_{X\setminus A}$. Then, μ_A and $\mu_{X\setminus A}$ are T-invariant and $\mu = \mu(A)\cdot\mu_A + \mu(X\setminus A)\cdot\mu_{X\setminus A}$ so μ is not an extreme point.

Conversely, suppose that μ is ergodic and μ is non-trivial convex combination of invariant measures ν and κ, say $\mu = t\nu + (1-t)\kappa$. Then ν and κ are absolutely continuous with respect to μ. We may write $\nu = gd\mu$. A priori, g is in $L^1(\mu)$. As $\mu \geq t\nu$, $g \leq \frac{1}{t}$ almost everywhere, so g is also in $L^2(\mu)$. For any measurable set E, $\int_E gd\nu = \mu(E) = \mu(T^{-1}E) = \int_{T^{-1}E} gd\nu = \int_E g\circ Td\nu$, hence g is essentially T-invariant.

We remember that the operator $U : f \to f\circ T$ is an isometry. Suppose f is a function in $L^2(\mu)$. Then, $\langle Uf, g\rangle = \int(f\circ T)gd\mu = \int(f\circ T)(g\circ T)d\mu = \int(fg)\circ Td\mu = \int fgd\mu = \langle f, g\rangle$. It follows that $\langle f, U^*g\rangle = \langle Uf, g\rangle = \langle f, g\rangle$. As this is true for all functions $f \in L^2(\mu)$, $U^*g = g$. But then, $Ug = g$ as well. As μ is ergodic, g is constant almost everywhere, so μ is a constant multiple of ν. Same reasoning shows that μ is a constant multiple of κ as well. This forces the three measures to be equal. $\square$

Problem 4. *Ergodic measures are mutually singular (hint: use Theorem 4).*

Very interesting to us will be the case when $\mathcal{M}_T$ is just a single point, that is there is a unique T-invariant Borel probability measure on X. Obviously, it is an extreme point of $\mathcal{M}_T$ and hence ergodic. As promised, this is precisely what we need to remove the irritating almost-everywhere conclusion of Theorem 4:

Theorem 7. *T is uniquely ergodic if and only if $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$ converges to a constant C for all continuous f.*

Proof. Suppose T is uniquely ergodic and let μ be the unique T-invariant measure. Let $x \in X$ be any point. Consider the sequence $\mu_n = \frac{1}{n}\sum_{k<n} \delta_{T^k x}$. It can only accumulate to a T-invariant measure, but as it must accumulate somewhere, it converges to μ. Then $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \to C = \int f(x)d\mu$. Conversely, if μ_1, μ_2 were two ergodic measures, then exists a continuous function f such that $\int f(x)d\mu_1 \neq \int f(x)d\mu_2$. Let x_1 be a point generic for μ_1 and x_2 generic for μ_2. Then $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x_1))$ and $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x_2))$ converge to different values. $\square$

Remark. *In the course of the proof, it has also been shown that all ergodic measures are weak-$*$ limits of $\frac{1}{n}\sum_{k<n} \delta_{T^k x}$.*

Problem 5. *With the hypothesis of Theorem 7, convergence is actually uniform in x.*

2.2 Uniform Distribution

Imagine a "centaur" making jumps of constant length α on the circle of circumference 1. The centaur locations will form the sequence of fractional parts $\{k\alpha\}$.

Theorem 8. *If α is rational, this motion will be periodic, but if α is irrational, the set of centaur locations will be everywhere dense and in fact, uniformly distributed within the circle.*

A sequence of points $\{x_n\}$ is *uniformly distributed* (in the circle) if for all intervals I, the number of points x_n, $0 \leq n < N$ that belong to I is roughly $N|I|$. More precisely, we want the equation

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} f(x_k) = \int_0^1 f(x)dx. \tag{$*$}$$

to hold for all characteristic functions of intervals.

Theorem 9 (Weyl's Criterion). *It is equivalent to ask that equation* $(*)$ *holds for all continuous functions, or more specifically, for the trigonometric polynomials* $e(jx) = e^{2\pi ijx}$ *for all integers* $j \neq 0$.

Observe that if $(*)$ holds for functions f, g, it also holds for their linear combinations. Additionally, if $(*)$ holds for a sequence of functions $f_1, f_2, \ldots$ uniformly converging to f, it also holds for f. Also note that if $j = 0$ (i.e $f \equiv 1$ identically), then $(*)$ automatically holds.

These observations allow us to pass from intervals to continuous functions and back and forth from continuous functions to trigonometric polynomials, but a little more thought is required to go from continuous functions to intervals.

Problem 6. *Show that equation* $(*)$ *holds for intervals if it does so for continuous functions (hint: given an interval* I*, we can choose continuous functions* f, g, h *so that* $g < |f - \chi_I| < h$ *with* $||g||_1, ||h||_1 < \epsilon$*).*

Problem 7. *Prove Theorem 8 directly and by using Weyl's criterion.*

Now we study the multi-dimensional analogue:

Theorem 10 (Kronecker). *The set of points* $(\alpha_1 n, \alpha_2 n, \ldots \alpha_k n)$ *where* $n = 1, 2, 3, \ldots$ *are uniformly distributed in the* k*-dimensional torus* T^k *if and only if* $\{1, \alpha_1, \alpha_2, \ldots \alpha_k\}$ *are linearly independent over* $\mathbb{Q}$.

Proof. We are led to investigate the transformation

$$T : (x_1, x_2, \ldots x_k) \to (x_1 + \alpha_1, x_2 + \alpha_2, \ldots x_k + \alpha_k).$$

Or perhaps its easier to write it like this:

$$X_1 : +\alpha_1, \qquad X_2 : +\alpha_2, \qquad X_3 : +\alpha_3, \qquad \ldots \qquad X_k : +\alpha_k.$$

To prove Kronecker's theorem, by Theorem 4, we simply need to decide whether T is ergodic: if we show that for any *one point* $x = (x_1, \ldots, x_k)$ that $x, Tx, T^2x, \ldots$ are uniformly distributed, by the homogeneity of T, this will in fact hold for *all* points (including the origin for which the theorem is stated).

If $\{1, \alpha_1, \alpha_2, \ldots \alpha_k\}$ are linearly dependent, then exist integers n_j such that $\sum n_j \alpha_j = 0$ in $\mathbb{R}/\mathbb{Z}$; so the function $f(x_1, x_2, \ldots x_n) = e(\sum n_j x_j)$ is in L^2, invariant under T but not constant. So T is not ergodic by Problem 1.

For the converse, to be explicit, we argue for $k = 1$ (but proof holds with $k > 1$ just as well). Let $f(x) \in L^2$ be a T-invariant function. We claim that it is actually constant. We can develop f into a Fourier series: $f(x) = \sum_n a_n e^{2\pi i n x}$ with $\sum_n |a_n|^2 < \infty$. So $f(T(x)) = \sum_n a_n e^{2\pi i n(\alpha + x)} = \sum_n a_n e^{2\pi i n\alpha} e^{2\pi i n x}$. Thus we must have $a_n = a_n e^{2\pi i n \alpha}$ for all n, but if α is not rational, the $e^{2\pi i n\alpha} \neq 1$ telling us that $\alpha_n = 0$ for all $n \neq 0$, so indeed $f(x)$ is constant. $\square$

Remark. *Showing ergodicity is often easy. In this instance, we got away without doing the hard work of proving unique ergodicity by means of a trick.*

2.3 Weyl's Lemma

Definition. *Polynomials $P_1, P_2, \ldots, P_k$ are* $\mathbb{Q}$–linearly independent *if $a_1P_1 + a_2P_2 + \cdots + a_kP_k = Q$ (where Q is a polynomial with rational coefficients) with rational a_i implies that in fact all $a_i = 0$ (and $Q = 0$).*

Theorem 11 (Weyl)**.** *For a set of $\mathbb{Q}$–linearly independent polynomials without constant term $P_1, P_2, \ldots, P_k$, the locations $\{P_1(n), P_2(n), \ldots, P_k(n)\}$ are uniformly distributed.*

We make two reductions. First we notice that rational terms, i.e $a_l x^l$ with a_l rational can be dropped. This is because $\{a_l x^l\}$ are periodic, call the period q. Then by the *reduced* statement, for every $0 \leq r < q$, we see that $\{P_s(nq + r)\}$ are uniformly distributed. By a trivial generalization of the following problem, it easily follows that $\{P_s(n)\}$ themselves are uniformly distributed:

Problem 8. *Suppose that $\{x_n\}$ are $\{y_n\}$ are two sequences uniformly distributed in the unit circle. Then the mixed sequence $x_1, y_1, x_2, y_2, \ldots$ is also uniformly distributed.*

As polynomials are linear combinations of monomials, we could assume that $\{P_s(n)\}$ are monomials with irrational coefficient. Usually we consider monomials of type

$$n\alpha, n^2\alpha, n^3\alpha, \ldots$$

For our purposes; however, it is more convenient to represent the polynomials by means of *binomials*, e.g by linear combinations of

$$\alpha\binom{n}{1}, \alpha\binom{n}{2}, \alpha\binom{n}{3}, \ldots$$

The binomial $\alpha\binom{n}{k}$ can be obtained by iterating the following transformation T on a k-dimensional torus:

$$X_1 : +\alpha, \qquad X_2 : +X_1, \qquad X_3 : +X_2, \qquad \ldots \qquad X_k : +X_{k-1}.$$

Now we see why introduced binomials: for usual monomials, the forward difference of x^k, $\Delta(x^k) = x^{k+1} - x^k$ depends not only on x^{k-1} but also on lower powers of x as well.

To be explicit, we will not prove Weyl's lemma in full generality, but only for the pair of polynomials $(n\alpha, \frac{n(n-1)}{2}\alpha)$ in two different ways. Both of these approaches can be easily adapted to prove Weyl's lemma in its entirety.

Theorem 12. *The transformation $T : T^k \to T^k$ just constructed is ergodic.*

Proof. We will give the explicit proof for $k = 2$ (the case of general k is very similar and is left to the reader). Arguing as in the proof of Theorem 10, we express an L^2 function f as a Fourier series $\sum_{m,n} a_{m,n} e^{2\pi i(mx+ny)}$ with $\sum_{m,n} |a_{m,n}|^2 < \infty$. This gives the invariance relation on the Fourier coefficients $a_{m,n} = e^{-2\pi i n\alpha} a_{m,m+n}$. From this one can easily see that $a_{mn} = 0$ for all $(m, n) \neq (0, 0)$: if $a_{m,n} \neq 0$ for some $m \neq 0$, we have infinitely many Fourier coefficients of the same magnitude contradicting $\sum_{m,n} |a_{m,n}|^2 < \infty$. Also $a_{0,n} = e^{-2\pi i n\alpha} a_{0,n}$ forcing all $a_{0,n}, n \neq 0$ to be 0 as well. This shows that f must be constant. □

2.4 Furstenburg's Method

Theorem 12 really is interesting; however, to argue by Theorem 7, what we need is unique ergodicity. The first approach is inductive and uses the following general construction:

Suppose $T : X \to X$ is a homeomorphism, G is a compact topological group and ϕ a continuous function from X to G. We can define a map S from $X \times G$ to itself by taking (x, g) to $(T(x), \phi(x)g)$.

It is well known that a compact group possesses a natural measure called the *Haar measure* which has total mass 1 and is invariant under multiplication from both left and right. (We will only need to use $G = S^1$ where the Haar measure is the usual Lebesgue measure). Denote the Haar measure on G by m. Endow the product space $X \times G$ with the product measure $\nu = \mu \times m$. We need a result which concludes the unique ergodicity of $X \times G$ under sufficient hypothesis.

Theorem 13 (Furstenberg)**.** *If T is uniquely ergodic and S is ergodic under ν, then S is actually uniquely ergodic.*

$$\begin{tikzcd}
X \times G \arrow[r, "S\,:\,(x{,}g)\,\to\,(Tx{,}\phi(x)g)", "\text{Ergodic}"'] \arrow[d, "\pi"] & X \times G \arrow[d, "\pi"] \\
X \arrow[r, "T", "\text{Uniquely Ergodic}"'] & X
\end{tikzcd}$$

Proof. As S acts on the G-coordinate from the left, it commutes with the right action of G, which we shall denote by $(\cdot h)$. Suppose (x, g) was ν generic. This means that for all continuous functions f,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} f(S^k(x,g)) = \int_{X\times G} f(x,g)d(\mu \times m).$$

Writing this out for the continuous function $f \circ (\cdot h)$ (the action of $(\cdot h)$ on the set of continuous functions takes the continuous functions *onto* themselves), we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} f(S^k(x,gh)) = \lim_{n\to\infty} \frac{1}{n} \sum_{k<n} f(S^k(x,g)\cdot h) = \int_{X\times G} f(x,gh)d(\mu \times m).$$

This is precisely the condition for (x, gh) being generic. It follows that the set of ν generic points consists of complete fibers. Since ν almost every point is ν generic, by Fubini's theorem, the set of base points of generic fibers A has full μ measure.

Suppose $\tilde{\nu}$ was another S-ergodic measure on $X \times G$ (if we don't have other ergodic measures, we don't have any other invariant measures as well). The measure $\tilde{\mu}$ on X given by $\tilde{\mu}(E) = \tilde{\nu}(E\times G)$ is T-invariant (and has total measure 1), but as T is uniquely ergodic, $\tilde{\mu}$ and μ must coincide. However, as ergodic measures are mutually singular, the measure of ν is essentially contained in $(X \setminus A)\times G$, hence $\tilde{\mu}$ is null outside A which is impossible. □

2.5 Wiener's Method

Now we give a more direct way of showing the unique ergodicity of T. Suppose μ was an invariant probability measure with Fourier coefficients $a_{m,n}$. It would have to satisify the invariance relation $a_{m,n} = e^{-2\pi i n\alpha} a_{m,m+n}$. We want to show that all $a_{m,n}$ except $a_{0,0}$ are zero: this will force μ to be a multiple of the Lebesgue measure. But now the problem is *harder*, for it is quite possible that μ may have infinite many Fourier coefficients of the same magnitude. To the rescue comes the famous theorem of Wiener (which shall not be proved here but can be found in any standard text on harmonic analysis):

Theorem 14 (Wiener)**.** *Suppose measure μ on S^1 has point masses a_k at points x_k, then*

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N}^{N} |\hat{\mu}(n)|^2 = \sum_k |a_k|^2.$$

In particular this implies that a measure having evenly-spaced Fourier coefficients of the same magnitude must have point masses.

Recall that we had the relation $a_{m,n} = e^{-2\pi i n\alpha} a_{m,m+n}$. By induction, we see that

$$a_{m,lm+n} = e^{2\pi i\alpha(\frac{l(l-1)}{2}m+ln)} a_{m,n}$$

for $l \in \mathbb{Z}$. A sequence of numbers $\{b_l\}$ is called *Fourier-Stieltjes* if they are the Fourier coefficients of some finite (possibly complex-valued) measure on S^1. Similarly a double sequence $\{a_{m,n}\}$ is Fourier-Stieltjes if they are the Fourier coefficients of some finite measure on T^2.

Problem 9. *Some operations on Fourier-Stieltjes sequences:*

1. *If $\{a_{m,n}\}$ is Fourier-Stieltjes, then so are $\{a_{m-m_0,n}\}, \{a_{0,n}\}, \{a_{m_0,n}\}$.*
2. *If $\{b_l\}$ is Fourier-Stieltjes, then so is $\{b_{k_0 l}\}$.*
3. *If $\{a_{m,n}\}$ is Fourier-Stieltjes then for fixed m, n, so is $\{b_l\} = \{a_{m,lm+n}\}$.*
4. *If $\{b_l\}, \{c_l\}$ are Fourier-Stieltjes, then so is $\{b_l \cdot c_l\}$ (hint: take convolution).*
5. *For all $\theta \in [0,1]$, $\{e^{2\pi i\theta l}\}$ is Fourier-Stieltjes (hint: take a point mass).*

Problem 9-3 tells us that $\{e^{2\pi i\alpha(\frac{l(l-1)}{2}m+ln)}\}$ is Fourier-Stieltjes in l and by 9-4, 9-5, we see that so is $\{e^{2\pi i m\alpha(\frac{l(l-1)}{2})}\}$. It suffices to show this is possible only when $m = 0$. Assuming the contrary, suppose they were the Fourier coefficients of measure ν. Then

$$\widehat{e^{-2\pi i x}\nu}(l) = \hat{\nu}(l+1) = e^{2\pi i m l\alpha}\hat{\nu}(l) = \widehat{\tau_{m\alpha}\nu}(l)$$

where $\tau_{m\alpha}$ means "translate $m\alpha$ units to the left". By uniqueness of Fourier coefficients, the measures $e^{-2\pi i x}\nu$ and $\tau_{m\alpha}\nu$ are equal. Wiener's theorem tells us that ν has a point mass, say at x, but then ν must have point masses at $x + m\alpha, x + 2m\alpha, x + 3m\alpha, \ldots$ of the same magnitude. But α is irrational, so these points are not congruent modulo 1. This contradicts the finiteness of ν.

Appendices

A.1 A result of a.e nature

I want to convince the reader that results of the "almost-everywhere" kind are very abundant, i.e that the strong hypothesis needed to show that a property holds everywhere are justified. Below is a most remarkable result:

Theorem 15. *Suppose that $\{n_k\}$ is a sequence of distinct (e.g increasing) integers. Then for almost every α, $\{n_k\alpha\}$ are uniformly distributed in the unit circle.*

Proof. Fix $j \neq 0$ and consider $A_m(\alpha) = \frac{1}{m}\sum_{k<m} e(jn_k\alpha)$. To apply Weyl's criterion, we would need to show that $A_m(\alpha) \to 0$ for almost every α. As the n_k are distinct, it follows that $\int_0^1 |A_p|^2 d\alpha = \frac{1}{p}$ and so

$$\int_0^1 \sum_{p=1}^{\infty} |A_{p^2}(\alpha)|^2 d\alpha = \sum_{p=1}^{\infty} \int_0^1 |A_{p^2}(\alpha)|^2 d\alpha < \infty.$$

This means that the sum $\sum_{p=1}^{\infty} |A_{p^2}(\alpha)|^2$ is finite for almost every α. In particular, $A_{p^2}(\alpha) \to 0$ for these α. Now, for an integer $m \geq 0$, choose p so that $p^2 \leq m < (p+1)^2$. But as $|mA_m - p^2 A_{p^2}| \leq m - p^2$, we see that

$$\left| A_m - \frac{p^2}{m} A_{p^2} \right| \leq 1 - \frac{p^2}{m}.$$

For large m, $\frac{p^2}{m}$ is very close to 1, so $A_m(\alpha)$ goes to 0 through all integers m and not just through squares. $\square$

A.2 Alternative Approach

For variety, I give another method to Weyl's lemma which does not use ergodic theory, but instead similar to techniques originally used by Weyl.

To simplify arguments, we will again work in a special case. Explicitly I will show Weyl's lemma in one dimension which says that *given a polynomial $P(n)$ with a irrational non-constant coefficient, the sequence of fractional parts $\{P(n)\}$ is uniformly distributed.* This follows from the following fact:

Theorem 16 (van der Corput)**.** *For $\{x_k\}$ to be uniformly distributed, it suffices to check that for every integer $h > 0$, $\{x_{n+h} - x_n\}$ is uniformly distributed.*

This in turn is proved by means of the following remarkable inequality:

Theorem 17. *Suppose $u_1, u_2, \ldots, u_N$ be complex numbers and H, N be positive integers with $H < N$ (one must actually think that N is much larger than H). Then*

$$\Big|\sum_{n=1}^{N} u_n\Big|^2 \leq \frac{N+H}{H+1}\sum_{n=1}^{N}|u_n|^2 + \frac{2(N+H)}{H+1}\sum_{h=1}^{H}\Big(1 - \frac{h}{H+1}\Big)\Big|\sum_{n=1}^{N-h} u_n\overline{u_{n+h}}\Big|.$$

Assuming the inequality, van der Corput's theorem goes as follows: For a integer $j \neq 0$, plug in $u_n = e(jx_n)$ into the above inequality, divide by N^2 and estimate crudely to obtain

$$\Big|\frac{1}{N}\sum_{n=1}^{N} e(jx_n)\Big|^2 \leq \frac{2}{H} + \frac{4}{H}\sum_{h=1}^{H}\Big|\frac{1}{N-h}\sum_{n=1}^{N-h} e(j(x_n - x_{n+h}))\Big|.$$

To apply Weyl's criterion to conclude the equidistribution of $\{x_n\}$, we need to know that for large N, the LHS is small. But by the Weyl criterion applied to $\{x_n - x_{n+h}\}$, to make the RHS small, it suffices to take H large.

Proof of the Fundamental Inequality

Proof. To avoid the writing bounds of summation, set $u_n = 0$ for n outside of $[1, N]$. It is easily seen that

$$\sum_{n=1}^{N} u_n = \sum_{k=1}^{N+H}\Big(\frac{1}{H+1}\sum_{q=0}^{H} u_{k-q}\Big).$$

The Cauchy-Schwartz inequality tells us that

$$\Big|\sum_{n=1}^{N} u_n\Big|^2 \leq \frac{N+H}{(H+1)^2} \sum_{k=1}^{N+H} \Big|\sum_{q=0}^{H} u_{k-q}\Big|^2 = \frac{N+H}{(H+1)^2} \sum_{k=1}^{N+H} \Big(\sum_{r=0}^{H} u_{k-r}\Big)\Big(\sum_{s=0}^{H} \overline{u_{k-s}}\Big).$$

The diagonal terms amount to

$$\frac{N+H}{(H+1)^2} \sum_{k=1}^{N+H} \sum_{q=0}^{H} |u_{k-q}|^2 = \frac{N+H}{(H+1)^2} \cdot (H+1) \sum_{n=1}^{N} |u_n|^2.$$

For the off-diagonal terms, we group $u_{p-r}\overline{u_{k-s}}$ with $u_{k-s}\overline{u_{k-r}}$ to obtain

$$\frac{2(N+H)}{(H+1)^2} \operatorname{Re}\Big\{ \sum_{k=1}^{N+H} \sum_{0 \leq s < r \leq H} u_{k-r}\overline{u_{k-s}} \Big\}.$$

Next, we bound the real part by its absolute value. Changing variables $h = r - s, n = k - r$, we find that the contribution of the off-diagonal terms does not exceed

$$\frac{2(N+H)}{(H+1)^2} \cdot \Big|\sum_{h=1}^{H} (H+1-h) \sum_{n=1}^{N} u_n \overline{u_{n+h}}\Big|.$$

(The factor $H - h + 1$ comes from the fact that r varies between $h \leq r \leq H$ so that $s = r - h \geq 0$).

Putting the diagonal and off-diagonal terms back together, we get what we need. $\square$

Remark. *The theorem of van der Corput provides a sufficient condition for the a sequence to be uniformly distributed, but it is clearly not necessary, for instance take $x_n = n\alpha$ with α irrational. In this case, for any h, the expression $\{x_{n+h} - x_n\}$ is constant.*

Bibliography

[BS] Brin, M. and Stuck, G. *Introduction to Dynamical Systems*, 2003.

[C] Conway, J. *A Course in Functional Analysis*, 1985.

[DI] Dziabenko, D. and Ivrii, O. *Centaur Dynamics and Kronecker's Theorem (to be published in College Mathematical Journal).*

[Ha] Halmosh, P. *Lectures on Ergodic Theory*, 1956.

[He] Helson, H. *Harmonic Analysis*, 1983.

[Ka] Katznelson, Y. *An Introduction to Harmonic Analysis*, 3rd edition, 2004.

[KN] Kuipers, L. and Niederreiter, H. *Uniform Distribution of Sequences*, 2006.

[M] Montgomery, H. L. *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, 1990.

[W] Waters, P. *An Intro to Ergodic Theory*, 1982.