The Pointwise Ergodic Theorem and its Applications

Peter Oberly
11/9/2018

Introduction

Algebra has homomorphisms and topology has continuous maps; in these notes we explore the structure preserving maps for measure theory known (somewhat unimaginatively) as measure preserving transformations. The first section contains some (but not all) of the necessary definitions for this talk and in the second we introduce some classical examples to illustrate these definitions. We then turn our attention to the dynamics of measure preserving maps which leads us to the pointwise ergodic theorem. In the final section we use the ergodic theorem to prove Borel's theorem on normal numbers.

Definitions

Definition. A *σ-algebra* $\mathcal{A}$ is a collection of subsets of a non-empty set X so that $X \in \mathcal{A}$ and $\mathcal{A}$ is closed under complementation and countable unions; The pair $(X, \mathcal{A})$ is called a *measurable space*, and elements of $\mathcal{A}$ are called *measurable sets*. A particularly important σ-algebra is the collection of *Borel sets*, defined to be the σ-algebra generated by the open subsets of a topological space X.

Definition. A *measure* $m : \mathcal{A} \to [0, \infty]$ is a function which satisfies the following:
1. $m(E) \geq 0$ for all $E \in \mathcal{A}$;
2. $m(\emptyset) = 0$;
3. If $\{E_n\}_{n=1}^{\infty} \subset \mathcal{A}$ is a sequence of pairwise disjoint sets in $\mathcal{A}$, then $m(\bigcup_n E_n) = \sum_n m(E_n)$.
A *measure space* is a triple $(X, \mathcal{A}, m)$ where $(X, \mathcal{A})$ is a measurable space and m is a measure defined on $\mathcal{A}$. The triple $(X, \mathcal{A}, m)$ is called a *probability space* if $m(X) = 1$.

Definition. Let $(X, \mathcal{A}, m)$ and $(Y, \mathcal{B}, n)$ be measure spaces, and let $T : X \to Y$ be a map from X into Y. T is said to be *measurable* if $T^{-1}(E) \in \mathcal{A}$ for each $E \in \mathcal{B}$; that is, if the pre-image of every measurable set is measurable.

Definition. A measurable transformation $T : (X, \mathcal{A}, m) \to (Y, \mathcal{B}, n)$ is said to be *measure-preserving* if $m(T^{-1}(E)) = n(E)$ for all $E \in \mathcal{B}$. If T is a bijection and T^{-1} is also measure preserving, then T is said to be *invertible*.
If $(X, \mathcal{A}, m)$ is a probability space, and if $T : X \to X$ is measure preserving, then the quadruple $(X, \mathcal{A}, m, T)$ is sometimes referred to as a *measurable dynamical system*.

Remarks:
(1) We should really write $T : (X, \mathcal{A}, m) \to (Y, \mathcal{B}, n)$ since the measure preserving property

depends on both the σ-algebras and the measures, but will often write $T : X \to Y$ instead.
(2) If $T : (X, \mathcal{A}, , m) \to (Y, \mathcal{B}, n)$ and $S : (Y, \mathcal{B}, n) \to (Z, \mathcal{C}, p)$ are measure preserving, then so is $S \circ T$.
(3) Measure preserving maps are the structure preserving transformations (morphisms) of measure spaces.
(4) As such, a measure preserving map $T : X \to X$ induces a morphism on the Banach space of m-integrable functions $\mathcal{L}^1(m)$. In detail, let $U_T : \mathcal{L}^1(m) \to \mathcal{L}^1(m)$ be defined by $U_T(f) = f \circ T$. It is evident that U_T is linear, and if $f \geq 0$ (and so is real valued), then $(U_T f)(x) = f(T(x)) \geq 0$ for $x \in X$. So U_T is positive. In fact, U_T is an isometry. For if s is a non-negative simple function $s = \sum_{k=1}^{n} a_k \chi_{A_k}$, where a_k are scalars and A_k are the measurable sets where $s > 0$, then

$$\int U_T(s)\, dm = \sum_{k=1}^{n} a_k \int \chi_{A_k} \circ T\, dm = \sum_{k=1}^{n} a_k m(T^{-1}(A_k)) = \sum_{k=1}^{n} a_k m(A_k) = \int s\, dm.$$

Therefore choosing a sequence of simple functions s_n which converges monotonically to $|f|$, where $f \in \mathcal{L}^1(m)$, shows $||U_T(f)||_1 = ||f||_1$. Note also that this shows U_T really does map into $\mathcal{L}^1(m)$.
(5) As we are interested in the dynamics of measure preserving maps, from now on we will restrict our attention to measurable functions $T : X \to X$. Additionally, unless other wise stated, we will assume that $(X, \mathcal{A}, m)$ is a probability space.

Our last definition requires a bit of motivation. Let $(X, \mathcal{A}, m, T)$ be a measurable dynamical system. If $T^{-1}(E) = E$ for $E \in \mathcal{A}$, then $T^{-1}(X \setminus E) = X \setminus E$ and we could study our system by examining the two simpler systems $(E, \mathcal{A} \cap E, m|_{E\,\mathcal{A} \cap E}, T|_E)$ and $(X \setminus E, \mathcal{A} \cap (X \setminus E), m|_{\mathcal{A} \cap (X \setminus E)}, T|_{X \setminus E})$ (with the corresponding measures normalized appropriately). If $0 < m(E) < 1$, then we have actually decomposed our original system into two smaller ones. However, if $m(E) = 0$ or $m(X \setminus E) = 0$ (i.e. $m(E) = 1$), then one of our simpler systems is in fact trivial, and we are left with a system essentially the same as the one we started with. It follows that those measurable dynamical systems where $T^{-1}(E) = E$ implies $m(E) = 0$ or 1 are not usefully decomposable in this way. It makes sense therefore to study those systems where such decomposition is not possible, for understanding these will enable us to understand the ones which can be simplified. We call such systems *ergodic*.

Definition. A measurable dynamical system $(X, \mathcal{A}, m, T)$ is said to be *ergodic* if $E \in \mathcal{A}$ and $T^{-1}(E) = E$ implies that $m(E) = 0$ or 1. We will often have a specific probability space $(X, \mathcal{A}, m)$ in mind and refer to the measure preserving transformation T as ergodic.

There are many characterizations of ergodicity; one which will prove useful in this talk is the following.

Theorem 1. *$(X, \mathcal{A}, m, T)$ is ergodic if and only if $f \in \mathcal{L}^1(m)$ and $f \circ T = f$ ae implies that f is constant ae.*

Proof. Assume that for all $f \in \mathcal{L}^1(m)$ that if $f \circ T = f$ ae then f is constant ae. Let $E \in \mathcal{A}$ be so that $T^{-1}(E) = E$. Then $\chi_E \circ T = \chi_E$. As $\chi_E \in \mathcal{L}^1(m)$, then χ_E is constant ae. Therefore χ_E is either 0 or 1 ae and so $m(E) = 0$ or 1. The converse is more technical, and can be found in McDonald and Wiess on page 616. □

Examples

(1) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map and let m be the Lebesgue measure on the Borel sets of $\mathbb{R}^n$. If T is singular, then range T is a proper subspace of $\mathbb{R}^n$, and so T is not measure preserving. If instead T is non-singular, from linear algebra then $m(T^{-1}(E)) = m(E)/|\det T|$ for all Borel sets E. Therefore T is a measure preserving linear map if and only if $|\det T| = 1$.

(2) Let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and let $\mathcal{B}$ denote the Borel σ-algebra. Then with normalized circular Lebesgue measure m, the triple $(S^1, \mathcal{B}, m)$ is a probability space. For $a \in S^1$, define the rotation $T_a : S^1 \to S^1$ by $T_a(z) = az$. Then T_a is measure preserving and invertible for all a.
It is very instructive to show the following.

Theorem 2. *The rotation $T = T_a$ is ergodic if and only if a is not a root of unity*

Proof. Suppose that a is a root of unity. Then $a^p = 1$ for some $p \neq 0$. Let $f : S^1 \to S^1$ be defined by $f(z) = z^p$. Then $(f \circ T)(z) = f(az) = a^p z^p = f(z)$ for all $z \in S^1$. Therefore $f \circ T = f$ but f is non-constant. So T is not ergodic by theorem 1.
Conversely let A be a measurable subset of S^1 so that $T^{-1}(A) = A$. Notice that the functions $e_n : S^1 \to S^1$ defined by $e_n(z) = z^n, n \in \mathbb{Z}$ form an orthonormal basis for $\mathcal{L}^2(m)$. Let the Fourier series for χ_A be $\chi_A \sim \sum_n b_n e_n$. Since $e_n(Tz) = a^n e_n(z)$, it follows by a change of variable that

$$b_n = \int \chi_A e_{-n} \, dm = a^{-n} \int_{T^{-1}(A)} e_{-n} \, dm,$$

and so $\chi_{T^{-1}(A)} \sim \sum_n a^n b_n e_n$. As $T^{-1}(A) = A$, then $\chi_A = \chi_{T^{-1}(A)}$ and thus have the same Fourier coefficients. Therefore $b_n = a^n b_n$ for all n. If a is not a root of unity, the only way this can hold is if $b_n = 0$ for all $n \neq 0$. By the uniqueness of Fourier coefficients, χ_A is a constant almost everywhere and so $m(A) = 0$ or 1. Therefore T_a is ergodic when a is not a root of unity. □

(3) Let $([0,1), \mathcal{B}, m)$ be the probability space consisting of the half open unit interval with Borel sets $\mathcal{B}$ and m the Lebesgue measure. Define $T : [0,1) \to [0,1)$ by

$$T(x) = 2x \mod 1 = \begin{cases} 2x, & \text{if } 0 \leq x < 1/2; \\ 2x - 1, & \text{if } 1/2 \leq x < 1. \end{cases}$$

This map is referred to as the *dyadic transformation*. Notice that if x has binary expansion $x = 0.x_1x_2x_3...(2)$ then $T(x) = 0.x_2x_3...(2)$. It is worth showing that T is measure preserving. From measure theory [Billingsly, p 4], it suffices to prove that T preserves measure on a semi-algebra which generates the Borel σ-algebra. The collection of half open intervals with rational dyadic endpoints is such a semi-algebra. So let $E = [\frac{k}{2^n}, \frac{j}{2^n})$ where $n \geq 0$ and

$0 \le k \le j \le 2^n$. Then

$$\begin{aligned} T^{-1}(E) &= \{x \in [0,1/2) : \frac{k}{2^n} \le 2x < \frac{j}{2^n}\} \cup \{x \in [1/2,1) : \frac{k}{2^n} \le 2x - 1 < \frac{j}{2^n}\} \\ &= [\frac{k}{2^{n+1}}, \frac{j}{2^{n+1}}) \cup [1/2 + \frac{k}{2^{n+1}}, \frac{1}{2} + \frac{j}{2^{n+1}}) \\ &= \frac{1}{2}E \cup (\frac{1}{2} + \frac{1}{2}E) \end{aligned}$$

and the translation invariance of the Lebesgue measure implies

$$m(T^{-1}(E)) = \frac{1}{2}m(E) + \frac{1}{2}m(E) = m(E).$$

So T is measure preserving. We sketch the proof that T is in fact ergodic. Let A be a measurable subset of $[0,1)$ with $T^{-1}(A) = A$. Let $x = 0.0x_2x_3...(2)$ and $x' = 0.1x_2x_3...(2)$ and assume that these are unique expansions. Then $T(x) = T(x') = 0.x_2x_3...(2)$. Now $x \in A$ is equivalent to $Tx \in A$ and similarly $x' \in A$ exactly when $Tx' \in A$. So $T(x) = T(x')$ implies $x \in A$ if and only if $x' \in A$. Then it follows that $A \cap [1/2,1) = 1/2 + A \cap [0,1/2)$. So $m(A \cap [0,1/2)) = m(A \cap [1/2,1))$, and hence

$$\begin{aligned} m(A) &= m(A \cap [0,1/2)) + m(A \cap [1/2,1)) \\ &= 2m(A \cap [0,1/2)) \\ &= m(A \cap [0,1/2))/m([0,1/2)). \end{aligned}$$

Thus $m(A)m([0,1/2)) = m(A \cap [0,1/2))$. Now this argument can be elaborated to show that this is true of any half open interval with rational dyadic endpoints, or any disjoint union of such intervals. Now given $\epsilon > 0$, choose such a disjoint union E so that $m(A \triangle E) < \epsilon$, where $\triangle$ denotes the symmetric difference (which we can do as A is measurable and the half open dyadic intervals generate the Borel sets). Then $|m(A) - m(E)| < \epsilon$ and $|m(A) - m(A \cap E)| = |m(A) - m(A)m(E)| < \epsilon$. Hence $|m(A) - m(A)^2| < 2\epsilon$ and as ϵ is arbitrary, then $m(A) = m(A)^2$. So $m(A) = 0$ or 1 and T is ergodic.

The Ergodic Theorem

To motivate the pointwise ergodic theorem, we first show that all measure preserving transformations on a finite measure space enjoy the property of *recurrence*:

Theorem 3 (The Poincaré Recurrence Theorem)**.** *Let $T : X \to X$ be a measure preserving transformation of a probability space $(X, \mathcal{A}, m)$. Let $E \in \mathcal{A}$ with $m(E) > 0$. Then almost all points of E return to E infinitely often under iteration by T; that is, $T^n(x) \in E$ for almost all $x \in E$ and for infinitely many n.*

Proof. Given $N \ge 0$, set $E_N = \bigcup_{n=N}^{\infty} T^{-n}(E)$ and set $F = E \cap \bigcap_{N=0}^{\infty} E_N$. Then $x \in F$ if and only if $x \in E$ and for all $N \ge 0$, there is an $n \ge N$ so that $T^n(x) \in E$. So F is the set

of points of E which return to E infinitely often under iteration by T. Note that if $x \in F$, then there is a subsequence $n_1 < n_2 < ... < n_j < ...$ of natural numbers so that $T^{n_j}(x) \in E$ for all j; therefore for each j we have $T^{n_j}(x) \in F$ since $T^{n_j - n_i}(T^{n_i}(x)) \in E$ for all i. Thus every point of F returns to F infinitely often under iteration by T.
It remains to show that $m(F) = m(E)$. Note that $T^{-1}(E_N) = \bigcup_{n=N}^{\infty} T^{-(n+1)}(E) = E_{N+1}$ and so $m(E_N) = m(E_{N+1})$ for all N. Therefore $m(E_N) = m(E_0)$ for all N and since $E_0 \supset E_1 \supset ...$ then $m(\bigcap_{N=0}^{\infty} E_N) = m(E_0)$. Therefore $m(F) = m(E \cap E_0) = m(E)$ as $E \subset E_0$. □

This begs the natural question: how often, or with what frequency, do the iterates of $T(x)$ return to a set? There is a very big difference between $T^{2n}(x) \in E$ and $T^{n!}(x) \in E$ for all n (and almost all $x \in E$) even though both return to E infinitely often. It makes sense then to consider the long term behavior of the average number of times $T^n(x)$ returns to E; that is to consider the limit of the ratios

$$\frac{1}{n}\sum_{k=0}^{n-1} \chi_E(T^k(x))$$

as $n \to \infty$. It is not obvious in what sense, if any at all, this limit exists. It is also quite restrictive to consider just characteristic functions; in a wide variety of applications both in theoretical math and the sciences, it is impossible to calculate or observe the orbit of a point directly. Instead, we rely on numerical data. We are therefore lead to consider the convergence of the ratios

$$\frac{1}{n}\sum_{k=0}^{n-1} (f \circ T^k)(x)$$

where $f : X \to \mathbb{C}$ is now a measurable function. It is even less clear in what sense this limit may exist, or with what restrictions we may require to ensure convergence. Birkhoff's celebrated pointwise ergodic theorem provides an answer to these questions.

Theorem 4 (Birkhoff's Pointwise Ergodic Theorem)**.** *Let $(X, \mathcal{A}, m)$ be a (possibly σ-finite) measure space and let $T : X \to X$ be measure preserving. If $f \in \mathcal{L}^1(m)$, then the limit*

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

converges pointwise almost everywhere to a function $f^* \in \mathcal{L}^1(m)$. Furthermore, $f^* \circ T = f^*$ ae (f^* is invariant), and if $m(X) < \infty$ then

$$\int f^* \, dm = \int f \, dm.$$

Remark. If T is also ergodic, then f^* is constant ae by theorem 1. So if $m(X) < \infty$, then $\int f^* \, dm = f^* m(X) = \int f \, dm$ ae and thus $f^* = \frac{1}{m(X)} \int f \, dm$. In particular, if T is ergodic and $(X, \mathcal{A}, m)$ is a probability space then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (f \circ T^k)(x) = \int f \, dm$$

for almost all $x \in X$ and all $f \in \mathcal{L}^1(m)$. This is the form of the ergodic theorem that may be the most familiar; that the time average tends to the space average for almost every point. This answers our question on the asymptotic frequency with which the orbit of a point x lies in a given measurable set E. For if T is ergodic, then

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k)(x) = m(E)$$

for almost every x in the probability space X.

We will only outline the proof. A detailed exposition can be found in Walters, Halmos, or Billingsly. The form of this proof is from Walters.

(1) The first step is to prove the maximal ergodic theorem, or rather the following corollary of it. The maximal ergodic theorem, along with the convergence theorems of Lebesgue theory, is what drives the proof of the pointwise ergodic theorem.

Theorem 5 (Maximal Ergodic Theorem)**.** *Let $(X, \mathcal{A}, m)$ be a finite measure space and $T : X \to X$ be measure preserving. If f is real-valued and integrable, then*

$$\int_A f \, dm \geq 0,$$

where

$$A = \{x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) > 0\}$$

Proof. As noted in the introduction, the map $U_T : \mathcal{L}^1_{\mathbb{R}}(m) \to \mathcal{L}^1_{\mathbb{R}}(m)$ defined by $U_T(f) = f \circ T$ is a positive linear isometry. Let $f_0 = 0$ and $f_n = f + U_T f + ... U_T^{n-1} f$ for $n \geq 1$. Set $F_N = \max_{0 \leq n \leq N} f_n$ and note that $F_N \geq 0$ for all $N \in \mathbb{N}$. Also observe that F_N is integrable since f_n is. We have $F_N \geq f_n$ for $0 \leq n \leq N$, and so $U_T(F_N) \geq U_T(f_n)$ by positivity. Hence $U_T(F_N) + f \geq f_{n+1}$, and therefore $U_T(F_N) + f \geq \max_{1 \leq n \leq N} f_n$. Thus if $x \in X$ and $F_N(x) > 0$, then

$$(U_T F_N)(x) + f(x) \geq \max_{0 \leq n \leq N} f_n(x) = F_N(x).$$

So $f \geq F_N - U_T F_N$ on $A_N = \{x \in X : F_N(x) > 0\}$. As $F_N(x) = 0$ on $X \setminus A_N$, then

$$\begin{aligned}
\int_{A_N} f \, dm &\geq \int_{A_N} F_N \, dm - \int_{A_N} U_T(F_N) \, dm \\
&= \int_X F_N \, dm - \int_{A_N} U_T(F_N) \, dm \\
&\geq \int_X F_N \, dm - \int_X F_N \, dm \\
&= ||F_N||_1 - ||U_T(F_N)||_1 \\
&= 0,
\end{aligned}$$

where we have used the fact that $\int_{A_N} U_T(F_N)\ dm \leq \int_X U_T(F_N)\ dm$ and that U_T is an isometry. Given $x \in X$, we see that $\sup_{n\geq 1} \frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f) > 0$ if and only if there is an N so that $\max_{0\leq n\leq N} f_n(x) = F_N(x) > 0$; hence $A = \bigcup_{n=0}^{\infty} A_N$. As $F_N \leq F_{N+1}$, then $A_N \subset A_{N+1}$ and so applying the monotone convergence theorem to $f \cdot \chi_{A_N}$ yields the desired claim. $\square$

(2) We make some simplifying assumptions and introduce notation.
Assume first that $m(X) < \infty$ and that f is real valued. Given $x \in X$, define

$$a_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)),$$

and

$$f^*(x) = \limsup_n a_n(x), \quad f_*(x) = \liminf_n (x).$$

As a_n is measurable for all n, then so are f^* and f_*. Notice that

$$a_n(Tx) = \left(\frac{n+1}{n}\right)a_{n+1}(x) - \frac{f(x)}{n}$$

for all n. Since $f \in \mathcal{L}^1(X)$, we can assume that $f(x) < \infty$ by redefining f on a set of measure zero if necessary. Therefore $f(x)/n \to 0$ as $n \to \infty$ and so

$$f^*(T(x)) = \limsup_n a_n(Tx) = \limsup_n (\frac{n+1}{n} a_{n+1}(x) - f(x)/n) = \limsup_n a_{n+1}(x) = f^*(x).$$

A similar argument shows that $f_* \circ T = f_*$ ae.

(3) We show that $f^* = f_*$ ae; that is, that the set $E = \{x \in X : f_*(x) < f^*(x)\}$ has measure zero. For real numbers a and b with $a < b$, let $E(a,b) = \{x \in X : f_*(x) < a < b < f^*(x)\}$. Then $E = \bigcup\{E(a,b) : a, b \in \mathbb{Q}\}$, so we show $m(E(a,b)) = 0$. As f_* and f_* are measurable, then so is $E(a,b)$ and therefore so is E. As $f^* \circ T = f^*$ and $f_* \circ T = f_*$ ae, then

$$T^{-1}(E(a,b)) = \{x \in X : f_*(Tx) < a < b < f^*(Tx)\} = E(a,b).$$

It is here that we need to use the maximal ergodic theorem.

(4) Notice that $E(a,b) \cap \{x \in X : \sup_{n\geq 1} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) > b\} = E(a,b)$. So apply the maximal ergodic theorem to the function $f - b$ to conclude

$$\int_{E(a,b)} f - b\ dm \geq 0, \ \text{ so } \int_{E(a,b)} f\ dm \geq bm(E(a,b))$$

and similarly

$$\int_{E(a,b)} a - f\ dm \geq 0, \ \text{ so } \int_{E(a,b)} f\ dm \leq am(E(a,b)).$$

Therefore $aE(a,b) \geq bE(a,b)$; since $b > a$, this can be true only if $m(E(a,b)) = 0$. Hence $f_* = f^*$ ae.

(5) To show that f^* is integrable, note that

$$\int |a_n|\ dm \leq \frac{1}{n}\int |\sum_{k=0}^{n-1} f \circ T^k|\ dm = \int |f(x)|\ dm < \infty,$$

where we have used a change of variables and the fact that T is m-invariant. Fatou's lemma implies then

$$\int \liminf |a_n|\ dm \leq \liminf \int |f|\ dm < \infty.$$

So $f^* \in \mathcal{L}^1(m)$.

(6) The last part is to show that $\int f\ dm = \int f^*\ dm$. Notice that

$$\int a_n\ dm = \frac{1}{n}\sum_{k=0}^{n-1}\int f \circ T^k\ dm = \int f\ dm$$

by changing variables and since T preserves measure. Therefore if we show that the interchange of limit and integral

$$\int f^*\ dm = \int \lim_n a_n\ dm = \lim_n \int a_n\ dm = \int f\ dm$$

is valid then the proof for the case $m(X) < \infty$ is complete. This is accomplished by another application of the maximal ergodic theorem and the dominated convergence theorem.

(7) For the case when X is σ-finite, the above will work so long as $m(E(a,b)) < \infty$ so that we can apply the maximal ergodic theorem. This is done by choosing a subset $C \subset E(a,b)$ with finite measure (which exists by σ-finiteness) and applying the maximal ergodic theorem to the function $f - b\chi_C$ to conclude (after a few more steps) that

$$\int |f|\ dm \geq bm(C).$$

Therefore if $C \subset E(a,b)$ has $m(C) < \infty$, then $m(C) \leq \frac{1}{b} \leq \int |f|\ dm$; it follows from σ-finiteness that $m(E(a,b)) < \infty$ as well.

Consequences of the Ergodic Theorem

A real number x is *normal* to base r if the expansion of x in base r contains each digit in the same proportion.

Theorem 6 (Borel's Theorem on Normal Numbers)**.** *Almost all numbers in* $[0,1)$ *are normal to base* r *for all integers* $r \geq 2$*; i.e. for almost all* $x \in [0,1)$ *the frequency of the digits* $0, 1, 2, ..., r-1$ *in the base* r *expansion of* x *occur with the same frequency* $1/r$*.*

Proof. Let $r \geq 2$ be an integer and define the r-adic transformation $T : [0,1) \to [0,1)$ by

$$T(x) = rx \mod 1 = \begin{cases} rx & 0 \leq x < 1; \\ rx - 1 & \frac{1}{r} \leq x < \frac{2}{r}; \\ \vdots & \\ rx - (r-1) & \frac{r-1}{r} \leq x < 1. \end{cases}$$

Just as for the dyadic transformation ($r = 2$), T is ergodic on $[0,1)$ with respect to the Lebesgue measure and Borel σ-algebra. Let X denote the set of points of $[0,1)$ which have unique base r expansion. Then $[0,1) \setminus X$ is countable so $m(X) = 1$. Let $x \in X$ and write x uniquely as $x = x_1 x_2 x_3 ...(r)$. Then

$$T(x) = T(0.x_1 x_2 ...) = 0.x_2 x_3 x_4 ...(r), \quad \text{and so } \; T^j(x) = 0.x_{j+1} x_{j+2} ...(r)$$

where $j \geq 0$. For ease of writing, let f denote the characteristic function $f = \chi_{[\frac{k}{r}, \frac{k+1}{r})}$, where $0 \leq k < r$ is an integer. Then

$$f(T^j(x)) = f(0.x_{j+1} x_{j+2} ...) = \begin{cases} 1, & \text{if } \; x_{j+1} = k; \\ 0, & \text{else.} \end{cases}$$

Therefore the number of times k appears in the first n digits of the r-adic expansion of x is $\sum_{j=0}^{n-1} f(T^j(x))$. Dividing by n and applying the ergodic theorem gives

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \to \int_{[0,1)} f \, dm = m([\frac{k}{r}, \frac{k+1}{r})) = \frac{1}{r}.$$

Hence the frequency with which $k \in \{0, 1, ..., r-1\}$ appears in the r-adic expansion of almost all numbers in $[0,1)$ is $1/r$. □

The pointwise ergodic theorem gives the following nice characterization of ergodicity.

Theorem 7. *A measurable dynamical system $(X, \mathcal{A}, m, T)$ is ergodic if and only if for all $A, B \in \mathcal{A}$*

$$\frac{1}{n} \sum_{k=0}^{n-1} m(T^{-k}(A) \cap B) \to m(A)m(B).$$

Proof. Suppose that T is ergodic. Applying the ergodic theorem to χ_A shows that

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k) \to m(A) \;\text{ a.e..}$$

Multiplying by χ_B gives $\frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k) \chi_B \to m(A) \chi_B$ a.e., and so the dominated convergence theorem implies

$$\frac{1}{n} \sum_{k=0}^{n-1} m(T^{-k}(A) \cap B) \to m(A)m(B) \;\text{ a.e.}$$

Conversely, suppose the convergence property holds. Suppose that $E \in \mathcal{A}$ with $T^{-1}(E) = E$. Set $A = B = E$; by assumption then

$$\frac{1}{n}\sum_{k=0}^{n-1} m(E) \to m(E)^2.$$

Since $\frac{1}{n}\sum_{k=0}^{n-1} m(E) = m(E)$ for all n then $m(E) = m(E)^2$ and so $m(E) = 0$ or 1. $\square$

This theorem provides a physical aid for understanding ergodic transformations; they are the maps which "stir" our space enough so that every measurable set will intersect every other measurable set in proportion to their relative size.

References

1. John McDonald, Neil Weiss. *A Course in Real Analysis*, 2nd edition, Academic Press 2012, chapter 16.
2. Paul R. Halmos. *Lectures on Ergodic Theory*, Martino Publishing, 2013
3. Peter Walters. *An Introduction to Ergodic Theory*, Springer-Verlag New York Inc., 1982
4. Patrick Billingsley. *Ergodic Theory and Information*, John Wiley and Sons Inc., 1965