5. Recurrence and Ergodicity

§5.1 Introduction

The aim of this lecture is to state Birkhoff's Ergodic Theorem and to study some of its applications in the context of ergodic transformations of a probability space. In particular we will apply it to the doubling map and to the continued fraction map and deduce some results of a number-theoretic nature. Birkhoff's Ergodic Theorem can be viewed as giving us information about rates of recurrence. Indeed, one application that we shall study is Kac's Lemma: given a set A of measure $\mu(A)$, the orbit of almost every point of A eventually returns to A and the expected time of the first recurrence is $1/\mu(A)$. We begin by discussing Poincaré's theorem: this gives us information about the recurrence properties of an arbitrary measure-preserving transformation.

§5.2 Poincaré's Recurrence Theorem

Theorem 5.1 (Poincaré's Recurrence Theorem)
Let $T : X \to X$ be a measure-preserving transformation of $(X, \mathcal{B}, \mu)$ and let $A \in \mathcal{B}$ have $\mu(A) > 0$. Then for μ-a.e. $x \in A$, the orbit $\{T^n x\}_{n=0}^{\infty}$ returns to A infinitely often.

Proof. Let

$$E = \{x \in A \mid T^n x \in A \text{ for infinitely many } n\},$$

then we have to show that $\mu(A \backslash E) = 0$.

If we write

$$F = \{x \in A \mid T^n x \notin A \ \forall n \geq 1\}$$

then we have the identity

$$A \setminus E = \bigcup_{k=0}^{\infty} (T^{-k} F \cap A).$$

Thus we have the estimate

$$\begin{aligned} \mu(A \backslash E) &= \mu\left(\bigcup_{k=0}^{\infty} (T^{-k} F \cap A)\right) \\ &\leq \mu\left(\bigcup_{k=0}^{\infty} T^{-k} F\right) \\ &\leq \sum_{k=0}^{\infty} \mu(T^{-k} F). \end{aligned}$$

Since $\mu(T^{-k}F) = \mu(F)\ \forall k \geq 0$ (because the measure is preserved), it suffices to show that $\mu(F) = 0$.

First suppose that $n > m$ and that $T^{-m}F \cap T^{-n}F \neq \emptyset$. If y lies in this intersection then $T^m y \in F$ and $T^{n-m}(T^m y) = T^n y \in F \subset A$, which contradicts the definition of F. Thus $T^{-m}F$ and $T^{-n}F$ are disjoint.

Since $\{T^{-k}F\}_{n=0}^{\infty}$ is a disjoint family, we have

$$\sum_{k=0}^{\infty} \mu(T^{-k}F) = \mu\left(\bigcup_{k=0}^{\infty} T^{-k}F\right) \leq \mu(X) = 1.$$

Since the terms in the summation have the constant value $\mu(F)$, we must have $\mu(F) = 0$. □

Remark There are many ways of proving Poincaré's Recurrence Theorem (see Petersen's book). The proof above avoids the more relaxed attitude towards sets of measure zero in the proof presented in the lecture (although that proof is fully rigorous); it also makes more explicit use of the fact that μ is a probability measure.

§5.3 Ergodic Theorems

An ergodic theorem is a result that describes the limiting behaviour of sequences of the form

$$\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j \tag{5.1}$$

as $n \to \infty$. The precise formulation of an ergodic theorem depends on the class of function f (for example, one could assume that f is integrable, L^2, or continuous), and the notion of convergence used (for example, the convergence could be pointwise, L^2, or uniform). Here we discuss von Neumann's (Mean) Ergodic Theorem and Birkhoff's Ergodic Theorem. Von Neumann's Ergodic Theorem is in the context of $f \in L^2$ and L^2-convergence of the ergodic averages (5.1); Birkhoff's Ergodic Theorem is in the context of $f \in L^1$ and almost everywhere pointwise convergence of (5.1). Note that L^2 convergence neither implies nor is implied by almost everywhere pointwise convergence.

Before stating these theorems, we first need to discuss conditional expectation.

§5.4 Conditional expectation

We will need the concepts of Radon-Nikodym derivates and conditional expectation.

Definition. Let μ be a measure on $(X, \mathcal{B})$. We say that a measure ν is *absolutely continuous with respect to* μ and write $\nu \ll \mu$ if $\nu(B) = 0$ whenever $\mu(B) = 0$, $B \in \mathcal{B}$.

Remark Thus ν is absolutely continuous with respect to μ if sets of μ-measure zero also have ν-measure zero (but there may be more sets of ν-measure zero). For example, let $f \in L^1(X, \mathcal{B}, \mu)$ be non-negative and define a measure ν by

$$\nu(B) = \int_B f \, d\mu.$$

Then $\nu \ll \mu$.

The following theorem says that, essentially, all absolutely continuous measures occur in this way.

Theorem 5.2 (Radon-Nikodym)
Let $(X, \mathcal{B}, \mu)$ be a probability space. Let ν be a measure defined on $\mathcal{B}$ and suppose that $\nu \ll \mu$. Then there is a non-negative measurable function f such that

$$\nu(B) = \int_B f \, d\mu, \quad \text{for all } B \in \mathcal{B}.$$

Moreover, f is unique in the sense that if g is a measurable function with the same property then $f = g$ μ-a.e.

Remark If $\nu \ll \mu$ then it is customary to write $d\nu/d\mu$ for the function given by the Radon-Nikodym theorem, that is

$$\nu(B) = \int_B \frac{d\nu}{d\mu} \, d\mu.$$

The following relations are all easy to prove, and indicate why the notation was chosen in this way.

(i) If $\nu \ll \mu$ and f is a μ-integrable function then

$$\int f \, d\nu = \int f \frac{d\nu}{d\mu} \, d\mu.$$

(ii) If $\nu_1, \nu_2 \ll \mu$ then

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu}.$$

(iii) If $\lambda \ll \nu \ll \mu$ then

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu}.$$

Let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Note that μ defines a measure on $\mathcal{A}$ by restriction. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then we can define a measure ν on $\mathcal{A}$ by setting

$$\nu(A) = \int_A f \, d\mu.$$

Note that $\nu \ll \mu|_{\mathcal{A}}$. Hence by the Radon-Nikodym theorem, there is a unique $\mathcal{A}$-measurable function $E(f \mid \mathcal{A})$ such that

$$\nu(A) = \int E(f \mid \mathcal{A}) \, d\mu.$$

We call $E(f \mid \mathcal{A})$ the *conditional expectation* of f with respect to the σ-algebra $\mathcal{A}$.

So far, we have only defined $E(f \mid \mathcal{A})$ for non-negative f. To define $E(f \mid \mathcal{A})$ for an arbitrary f, we split f into positive and negative parts $f = f_+ - f_-$ where $f_+, f_- \geq 0$ and define

$$E(f \mid \mathcal{A}) = E(f_+ \mid \mathcal{A}) - E(f_- \mid \mathcal{A}).$$

Thus we can view conditional expectation as an operator

$$E(\cdot \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu).$$

Note that $E(f \mid \mathcal{A})$ is uniquely determined by the two requirements that

(i) $E(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable, and

(ii) $\int_A f \, d\mu = \int_A E(f \mid \mathcal{A}) \, d\mu$ for all $A \in \mathcal{A}$.

Intuitively, one can think of $E(f \mid \mathcal{A})$ as the best approximation to f in the smaller space of all $\mathcal{A}$-measurable functions.

To state von Neumann's and sBirkhoff's Ergodic Theorems precisely, we will need the sub-σ-algebra $\mathcal{I}$ of T-invariant subsets, namely:

$$\mathcal{I} = \{B \in \mathcal{B} \mid T^{-1}B = B \text{ a.e.}\}.$$

It is straightforward to check that $\mathcal{I}$ is a σ-algebra. Note that if T is ergodic then $\mathcal{I}$ is the trivial σ-algebra consisting of all sets in $\mathcal{B}$ of measure 0 or 1.

§5.5 Von Neumann's Mean Ergodic Theorem

Von Neumann's Ergodic Theorem deals with the L^2-limiting behaviour of $\frac{1}{n}\sum_{j=0}^{n-1} fT^j$ for $f \in L^2(X, \mathcal{B}, \mu)$.

Theorem 5.3 (Von Neumann's Ergodic Theorem)
Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving transformation. Let $\mathcal{I}$ denote the σ-algebra of T-invariant sets. Then for every $f \in L^2(X, \mathcal{B}, \mu)$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} fT^j \to E(f \mid \mathcal{I})$$

in L^2.

Corollary 5.4
Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be an ergodic measure-preserving transformation. Let $f \in L^2(X, \mathcal{B}, \mu)$. Then

$$\frac{1}{n}\sum_{j=0}^{n-1} fT^j \to \int f\, d\mu, \qquad \text{as } n \to \infty,$$

in L^2.

Proof. If T is ergodic then $\mathcal{I}$ is the trivial σ-algebra $\mathcal{N}$ consisting of sets of measure 0 and 1. If $f \in L^2(X, \mathcal{B}, \mu)$ then $E(f \mid \mathcal{N}) = \int f\, d\mu$. □

In order to prove von Neumann's Ergodic Theorem, it is useful to recast it in terms of spectral theory.

Theorem 5.5 (von Neumann's Ergodic Theorem for Operators)
Let U be an unitary operator of a complex Hilbert space H. Let $I = \{v \in H \mid Uv = v\}$ be the subspace of U-invariant functions and let $P_I : H \to I$ be orthogonal projection onto I. Then for all $v \in H$ we have

$$\frac{1}{n}\sum_{j=0}^{n-1} U^j v \to P_I v \tag{5.2}$$

in the norm induced on H by the inner product.

Proof of Theorem 5.5. First note that if $v \in I$ then (5.2) holds, as

$$\frac{1}{n}\sum_{j=0}^{n-1} U^j v = v = P_I v.$$

If $v = Uw - w$ for some $w \in H$ then

$$\left\| \frac{1}{n}\sum_{j=0}^{n-1} U^j v \right\| = \frac{1}{n}\|U^n w - w\| \le \frac{1}{n} 2\|w\| \to 0.$$

If we let B denote the norm-closure of the subspace $\{Uw - w \mid w \in H\}$ then it follows that

$$\frac{1}{n}\sum_{j=0}^{n-1} U^j v \to 0$$

for all $v \in B$, by approximation.

We claim that $H = I \oplus B$, an orthogonal decomposition. Suppose that $v \perp B$. Then $\langle v, Uw - w\rangle = 0$ for all $w \in H$. Hence $\langle U^*v, w\rangle = \langle v, w\rangle$ for all $w \in H$. Hence $U^*v = v$. As U is unitary, we have that $U^* = U^{-1}$. Hence $v = Uv$, so that $v \in I$. Reversing each implication we see that $v \in I$ implies $v \perp B$, and the claim follows. □

Remark Note that an isometry of a Hilbert space H is a linear operator U such that $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in H$. We say that U is unitary if, in addition, it is invertible. Equivalently, U is unitary if the dual operator U^* is the inverse of U: $U^*U = UU^* = \mathrm{id}$.

We can prove von Neumann's Ergodic Theorem for an invertible measure-preserving transformation T of a probability space $(X, \mathcal{B}, \mu)$ as follows. Recall that $L^2(X, \mathcal{B}, \mu)$ is a Hilbert space with respect to the inner product

$$\langle f, g\rangle = \int f\bar{g}\, d\mu$$

and that T induces a linear operator $U : L^2 \to L^2$ by $Uf = f \circ T$. As T is measure-preserving, we have that U is an isometry; if T is invertible then U is unitary.

Hence, when T is invertible, Theorem 5.3 follows immediately from Theorem 5.5.

One can deduce from Theorem 5.5 that the result continues to hold when U is an isometry and is not assumed to be invertible (this makes a good exercise). Instead, we give an argument based around the construction of the natural extension of the dynamical system T.

Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Introduce a new space $\hat{X} = \{(x_j)_{j=0}^{\infty} \mid T(x_j) = x_{j-1}\}$ (essentially, we are just extending the space X so that it contains all the possible pasts for each point $x \in X$. Equip $\hat{X}$ with the smallest σ-algebra $\hat{B}$ which contains all sets of the form $\{(x_j)_{j=0}^{\infty} \mid x_0 \in B_0, \ldots, x_r \in B_r\}$ for each $B_j \in \mathcal{B}$ and any $r \in \mathbb{N}$. Define a measure $\hat{\mu}$ on such sets by

$$\hat{\mu}\{(x_j)_{j=0}^{\infty} \mid x_0 \in B_0, \ldots, x_r \in B_r\} = \mu(T^{-r}B_0 \cap \cdots \cap B_r)$$

and extend to $\hat{\mathcal{B}}$ by using the Kolmogorov extension theorem. Define a transformation $\hat{T} : \hat{X} \to \hat{X}$ by $T(x_0, x_1, x_2, \ldots) = (Tx_0, Tx_1, Tx_2, \ldots)$. Then $\hat{T}$ is invertible: $\hat{T}^{-1}(x_0, x_1, x_2, \ldots) = (x_1, x_2, \ldots)$.

One can check that $\hat{\mu}$ is a $\hat{T}$-invariant measure if and only if μ is a T-invariant measure. Indeed, $\hat{\mu}$ is ergodic for $\hat{T}$ if and only if μ is ergodic for T.

The transformation $\hat{T}$ is called the natural extension of T.

Let $f \in L^2(X, \mathcal{B}, \mu)$. Then we obtain a function $\hat{f} \in L^2(\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ by defining $\hat{f}(x_0, x_1, x_2, \ldots) = f(x_0)$. Then one can easily see that

$$\frac{1}{n}\sum_{j=0}^{n-1} \hat{f}\hat{T}^j = \frac{1}{n}\sum_{j=0}^{n-1} fT^j.$$

Hence von Neumann's Ergodic Theorem for T follows from von Neumann's Ergodic Theorem for $\hat{T}$.

§5.6 Birkhoff's Pointwise Ergodic Theorem

Birkhoff's Ergodic Theorem deals with the behaviour of $\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$ for μ-a.e. $x \in X$, and for $f \in L^1(X, \mathcal{B}, \mu)$.

Theorem 5.6 (Birkhoff's Ergodic Theorem)
Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving transformation. Let $\mathcal{I}$ denote the σ-algebra of T-invariant sets. Then for every $f \in L^1(X, \mathcal{B}, \mu)$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to E(f \mid \mathcal{I})$$

for μ-a.e. $x \in X$.

Corollary 5.7
Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be an ergodic measure-preserving transformation. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to \int f \, d\mu, \qquad \text{as } n \to \infty,$$

for μ-a.e. $x \in X$.

§5.7 Applications of Birkhoff's Ergodic Theorem

§5.7.1 Kac's Lemma

Poincaré's Recurrence Theorem tells us that, under a measure-preserving transformation, almost every point of a subset A of positive measure will return to A. However, it does not tell us how long we should have to wait for this to happen. One would expect that return times to sets of large measure

are small, and that return times to sets of small measure are large. This is indeed the case, and forms the content of Kac's Lemma.

Let $T : X \to X$ be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$ and let $A \subset X$ be a measurable subset with $\mu(A) > 0$. By Poincaré's Recurrence Theorem, the integer

$$n_A(x) = \inf\{n \geq 1 \mid T^n(x) \in A\}$$

is defined for a.e. $x \in A$.

Theorem 5.8 (Kac's Lemma)
Let T be an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}$ be such that $\mu(A) > 0$. Then

$$\int_A n_A \, d\mu = 1.$$

Proof. Let

$$A_n = A \cap T^{-1}A^c \cap \cdots \cap T^{-(n-1)}A^c \cap T^{-n}A.$$

Then A_n consists of those points in A that return to A after exactly n iterations of T, i.e. $A_n = \{x \in A \mid n_A(x) = n\}$.

Consider the illustration in figure 5.7.1. As T is ergodic, almost every

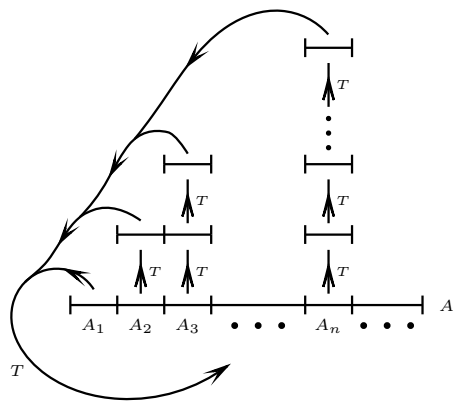

Figure 5.1: The return times to A

point of X eventually enters A. Hence the diagram represent almost all of X. Note that the column above A_n in the diagram consists of n sets, $A_{n,0}, \ldots, A_{n,n-1}$ say, with $A_{n,0} = A_n$. Note that $T^{-k}A_{n,k} = A_n$. As T is measure-preserving, it follows that $\mu(A_{n,k}) = \mu(A_n)$ for $k = 0, \ldots, n-1$. Hence

$$1 = \mu(X)$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} \sum k = 1^n \mu(A_{n,k}) \\
&= \sum_{n=1}^{\infty} n\mu(A_n) \\
&= \sum_{n=1}^{\infty} \int_{A_n} n_A \, d\mu \\
&= \int_A n_A \, d\mu.
\end{aligned}
$$

□

§5.7.2 Ehrenfests' example

The following example, due to P. and T. Enhrenfest, demonstrates that the return times in Poincaré's Recurrence Theorem may be extremely large.

Consider two urns. One urn contains 100 balls, numbered 1 to 100, and the other urn is empty. We also have a random number generator: this could be a bag containing 100 slips of paper, numbered 1 to 100.

Each second, a slip of paper is drawn from the bag, the number is noted, and the slip of paper is returned to the bag. The ball bearing that number is then moved from whichever urn it is currently in to the other urn.

Naively, we would expect that the system will settle into an equilibrium state in which there are 50 balls in each urn. Of course, there will continue to be small random fluctuations about the 50-50 distribution. However, it would appear highly unlikely for the system to return to the state in which 100 balls are in the first urn. Nevertheless, the Poincaré Recurrence Theorem tells us that this situation will occur almost surely (although we will have to wait a long time for this to happen).

To see this, we represent the system as a full shift on 101 symbols with an appropriate measure. Regard $x_j \in \{0, \ldots, 100\}$ as being the number of balls in the first urn after j seconds. Hence a sequence $(x_j)_{j=0}^{\infty}$ records the number of balls in the first urn at each time. As the number of balls in the first urn increases or decreases by 1 each second, such sequences determine the shift of finite type

$$\Sigma = \{\{(x_j)_{j=0}^{\infty} \mid x_j \in \{0, \ldots, 100\}, |x_j - x_{j+1}| = 1 \text{ for } j = 0, 1, 2, \ldots\}$$

(this corresponds to a transition matrix $A = (a_{ij})$ where $a_{i,i+1} = a_{i+1,i} = 1$ for $i = 0, \ldots, 100$ and $a_{ij} = 0$ otherwise).

Let p_i denote the probability of there being i balls in the first n. This probability is independent of time, and is equal to

$$p_i = \frac{1}{2^{100}} \begin{pmatrix} 100 \\ i \end{pmatrix}.$$

If we have i balls in the first urn then at the next stage we must have either $i-1$ or $i+1$ balls in the first urn. The number of balls becomes $i-1$ if the random number chosen is equal to the number of one of the balls in the first urn. As there are currently i such balls, the probability of this happening is $i/100$. Hence the conditional probability $P_{i,i-1}$ that there are $i-1$ balls remaining given that we started with i balls in the first urn is $i/100$. Similarly, the conditional probability $P_{i,i+1}$ that there are $i+1$ balls in the first urn given that we started with i balls is $(100-i)/100$. This defines a stochastic matrix:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ \frac{1}{100} & 0 & \frac{99}{100} & 0 & 0 & \cdots \\ 0 & \frac{2}{100} & 0 & \frac{98}{100} & 0 & \cdots \\ 0 & 0 & \frac{3}{100} & 0 & \frac{97}{100} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Note that $P_{i,j} \neq 0$ if and only if $A_{ij} = 1$ and so P and A are compatible. It is straightforward to check that $pP = p$. Hence we have a Markov probability measure μ_P defined on Σ. The matrix A is irreducible (but is not aperiodic); this ensures that μ_P is ergodic.

Consider the cylinder $A = [100]$ of length 1. The represents there being 100 balls in the first urn. By Poincaré's Recurrence Theorem, if we start in A then we return to A infinitely often. By Kac's lemma, the expected first return time to A is

$$\frac{1}{\mu_P(A)} = 2^{100} \text{ seconds,}$$

which is about 4×10^{22} years, or about 3×10^{12} times the length of time that the Universe has so far existed!

Remark This suggests that returning to a small set is a 'rare' event, and as such the return times are likely to have a Poisson distribution. This is normally formalised as follows. Let T be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$ and let $x \in X$. Let A_n be a decreasing sequence of subsets that decrease to x. Define the return time as $\tau(x) = \lim_{n\to\infty} \mu(A_n)\tau_{A_n}(x)$. Then $\int \tau \, d\mu = 1$. One would normally expect τ to have a Poisson distribution, but this is only known in particular cases; usually one needs some hyperbolicity of the dynamics T (such as being a shift of finite type). Additionally, there are normally restrictions on the sets A_n (such as requiring them to be cylinders).

§5.7.3 Normality of numbers

Let $r \geq 2$. Recall that any number $x \in [0,1]$ can be written as a base r 'decimal', i.e. there exist digits $x_j \in \{0, 1, \ldots, r-1\}$ for which

$$x = \sum_{j=1}^{\infty} \frac{x_j}{r^j}.$$

This r-adic expansion is unique, unless the sequence (x_j) ends in either infinitely repeated 0s or infinitely repeated $(r-1)$s.

Definition. Fix $r \geq 2$. A number $x \in [0,1]$ is said to be *(simply) normal (in base r)* if it has a unique expansion as an r-adic expansion, and for each $k = 0, 1, \ldots, r-1$, the frequency with which digit k occurs in its r-adic expansion is equal to $1/r$.

For each $r \geq 2$, define the map $T_r : [0,1] \to [0,1]$ by $T_r(x) = rx \bmod 1$. (The case $r = 2$ is the doubling map.) It is easy to see, by following the arguments for the doubling map, that Lebesgue measure μ on $[0,1]$ is an ergodic invariant measure for T_r.

The close connections between r-adic expansions and the map T_r can be used to prove the following result. A number is said to be *normal* if it is simultaneously simply normal in every base $r \geq 2$.

Proposition 5.9
Lebesgue almost every number in $[0,1]$ is normal.

Proof. Fix $r \geq 2$. Then clearly all but a countable set of points has a unique r-adic expansion. Fix $k \in \{0, 1, \ldots, r-1\}$. Then it is easy to see that $x_j = k$ if and only if $T^{j-1}x \in [k/r, (k+1)/r)$. Thus

$$\frac{1}{n}\text{card}\{1 \leq j \leq n \mid x_j = k\} = \frac{1}{n}\sum_{j=0}^{n-1} \chi_{[k/r,(k+1)/r)}(T^j x).$$

By Birkhoff's Ergodic Theorem for Lebesgue almost every point x the above expression converges to $\int \chi_{[k/r,(k+1)/r)}(x)\,dx = 1/r$. Let N_r denote the set of such points.

As N_r has measure 1 for each $r \geq 2$, it follows that $N = \cap_{r=2}^{/infty} N_r$ has measure 1. Hence Lebesgue almost every point is normal. □

Remark Given $r \geq 2$ it is easy to construct a number that is simply normal in base r. However, not a single example is known of a number that is simultaneously normal in every base $r \geq 2$.

One can easily use Birkhoff's Ergodic Theorem to prove the following result.

Proposition 5.10
For Lebesgue-almost every point $x \in [0,1]$, the arithmetic mean of the digits occurring in the base r expansion of x is $(r-1)/2$.

We leave this as an exercise.

§5.7.4 Continued fractions

We can prove similar results for the distribution of digits in the continued fraction expansion of real numbers.

Proposition 5.11
For Lebesgue-almost every $x \in [0,1]$, the frequency with which the natural number k occurs in the continued fraction expansion of x is

$$\frac{1}{\log 2} \log\left(\frac{(k+1)^2}{k(k+2)}\right).$$

Proof. Let λ denote Lebesgue measure and let μ denote Gauss' measure. Then λ-a.e. and μ-a.e. $x \in (0,1)$ is irrational and has an infinite continued fraction expansion

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \cdots}}}}.$$

Let T denote the continued fraction map. Then $x_n = [1/T^n x]$.

Fix $k \in \mathbb{N}$. Then $x_n = k$ precisely when $[1/T^n x] = k$, i.e.

$$k \leq \frac{1}{T^n x} < k+1$$

which is equivalent to requiring

$$\frac{1}{k+1} < T^n x \leq \frac{1}{k}.$$

Hence

$$\begin{aligned}
\frac{1}{n}\mathrm{card}\{0 \leq j \leq n-1 \mid x_j = k\} &= \frac{1}{n}\sum_{j=0}^{n-1} \chi_{(1/(k+1),1/k]}(T^i x) \\
&\to \int \chi_{(1/(k+1),1/k]}\, d\mu \text{ for } \mu\text{-a.e. } x \\
&= \frac{1}{\log 2}\left[\log\left(1+\frac{1}{k}\right) - \log\left(1+\frac{1}{k+1}\right)\right] \\
&= \frac{1}{\log 2}\log\frac{(k+1)^2}{k(k+2)}.
\end{aligned}$$

As μ and λ are equivalent, this holds for Lebesgue almost every point. $\square$

Proposition 5.12

(i) For Lebesgue-almost every $x \in [0,1]$, the arithmetic mean of the digits in the continued fraction expansion of x is infinite.

(ii) For Lebesgue-almost every $x \in [0,1]$, the geometric mean of the digits in the continued fraction expansion of x is

$$\prod_{k=1}^{\infty}\left(1+\frac{1}{k^2+2k}\right)^{\log k/\log 2}.$$

Proof. Writing

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \cdots}}}}.$$

the proposition claims that

$$\lim_{n\to\infty}\frac{1}{n}(x_0 + x_1 + \cdots + x_{n-1}) = \infty \tag{5.3}$$

almost everywhere, and

$$\lim_{n\to\infty}(x_0 x_1 \cdots x_{n-1})^{1/n} = \prod_{k=1}^{\infty}\left(1+\frac{1}{k^2+2k}\right)^{\log k/\log 2} \tag{5.4}$$

almost everywhere.

We leave (5.3) as an exercise.

We prove (5.4). Define $f(x) = \log k$ for $x \in (1/(k+1), 1/k]$. Then

$$\begin{aligned}
\frac{1}{n}(\log a_0 + \log a_1 + \cdots + a_{n-1}) &= \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \\
&\to \frac{1}{\log 2}\int_0^1 \frac{f(x)}{1+x}\,dx \\
&= \frac{1}{\log 2}\sum_{k=1}^{\infty}\int_{1/(k+1)}^{1/k}\frac{\log k}{1+x}\,dx \\
&= \sum_{k=1}^{\infty}\frac{\log k}{\log 2}\log\left(1+\frac{1}{k^2+2k}\right),
\end{aligned}$$

for Gauss-almost every, hence Lebesgue-almost every, point $x \in [0,1]$. □

§5.8 Appendix: The proof of Birkhoff's Ergodic Theorem

The proof is something of a tour de force of hard analysis. It is based on the following inequality.

Theorem 5.13 (Maximal Inequality)
Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T : X \to X$ be a measure-preserving transformation and let $f \in L^1(X, \mathcal{B}, \mu)$. Define $f_0 = 0$ and, for $n \geq 1$,

$$f_n = f + f \circ T + \cdots + f \circ T^{n-1}.$$

For $n \geq 1$, set $F_n(x) = \max_{0 \leq j \leq n} f_j(x)$. Then $F_n(x) \geq 0$. Then

$$\int_{\{x \in X \mid F_n(x) > 0\}} f \, d\mu \geq 0.$$

Proof. Clearly $F_n \in L^1(X, \mathcal{B}, \mu)$. For $0 \leq j \leq n$, we have $F_n \geq f_j$, so $F_n \circ T \geq f_j \circ T$. Hence

$$F_n \circ T + f \geq f_j \circ T + f = f_{j+1}$$

and therefore

$$F_n \circ T(x) + f(x) \geq \max_{1 \leq j \leq n} f_j(x).$$

If $F_n(x) > 0$ then

$$\max_{1 \leq j \leq n} f_j(x) = \max_{0 \leq j \leq n} f_j(x) = F_n(x),$$

so we obtain that

$$f \geq F_n - F_n \circ T$$

on the set $A = \{x \mid F_n(x) > 0\}$.

Hence

$$\begin{aligned}
\int_A f \, d\mu &\geq \int_A F_n \, d\mu - \int_A F_n \circ T \, d\mu \\
&= \int_X F_n \, d\mu - \int_A F_n \circ T \, d\mu \text{ as } F_n = 0 \text{ on } X \setminus A \\
&\geq \int_X F_n \, d\mu - \int_X F_n \circ T \, d\mu \text{ as } F_n \circ T \geq 0 \\
&= 0 \text{ as } \mu \text{ is } T\text{-invariant.}
\end{aligned}$$

□

Corollary 5.14
Let $g \in L^1(X, \mathcal{B}, \mu)$ and let

$$M_\alpha = \left\{ x \in X \mid \sup_{n \geq 1} \frac{1}{n} \sum_{j=0}^{n-1} g(T^j x) > \alpha \right\}.$$

Then for all $B \in \mathcal{B}$ with $T^{-1}B = B$ we have that

$$\int_{M_\alpha \cap A} g \, d\mu \geq \alpha \mu(M_\alpha \cap B).$$

Proof. Suppose first that $B = X$. Let $f = g - \alpha$, then

$$M_\alpha = \bigcup_{n=1}^{\infty} \left\{ x \mid \sum_{j=0}^{n-1} g(T^j x) > n\alpha \right\} = \bigcup_{n=1}^{\infty} \{x \mid f_n(x) > 0\} = \bigcup_{n=1}^{\infty} \{x \mid F_n(x) > 0\}$$

(since $f_n(x) > 0 \Rightarrow F_n(x) > 0$ and $F_n(x) > 0 \Rightarrow f_j(x) > 0$ for some $1 \le j \le n$). Write $C_n = \{x \mid F_n(x) > 0\}$ and observe that $C_n \subset C_{n+1}$. Thus χ_{C_n} converges to χ_{B_α} and so $f\chi_{C_n}$ converges to $f\chi_{M_\alpha}$, as $n \to \infty$. Furthermore, $|f\chi_{C_n}| \le |f|$. Hence, by the Dominated Convergence Theorem,

$$\int_{C_n} f\, d\mu = \int_X f\chi_{C_n}\, d\mu \to \int_X f\chi_{M_\alpha}\, d\mu = \int_{M_\alpha} f\, d\mu, \quad \text{as } n \to \infty.$$

Applying the Maximal Inequality, we have, for all $n \ge 1$ we have that $\int_{C_n} f\, d\mu \ge 0$. Therefore $\int_{M_\alpha} f\, d\mu \ge 0$, i.e., $\int_{B_\alpha} g\, d\mu \ge \alpha\mu(B_\alpha)$.

For the general case, we work with the restriction of T to B, $T : B \to B$, and apply the Maximal Inequality on this subset to get

$$\int_{M_\alpha \cap B} g\, d\mu \ge \alpha\mu(M_\alpha \cap B),$$

as required. □

We will also need the following convergence result.

Proposition 5.15 (Fatou's Lemma)
Let $(X, \mathcal{B}, \mu)$ be a probability space and suppose that $f_n : X \to \mathbb{R}$ are measurable functions. Define $f(x) = \liminf_{n\to\infty} f_n(x)$. Then f is measurable and

$$\int f\, d\mu \le \liminf_{n\to\infty} \int f_n\, d\mu$$

(one or both of these expressions may be infinite).

Proof of Birkhoff's Ergodic Theorem. Let

$$f^*(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x), \;\; f_*(x) = \liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x).$$

These exist (but may be $\pm\infty$, respectively) at all points $x \in X$. Clearly $f_*(x) \le f^*(x)$.

Let

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

Observe that

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

As f is finite μ-a.e., we have that $f(x)/n \to 0$ μ-a.e. as $n \to \infty$. Hence, taking the lim sup and lim inf as $n \to \infty$, gives us that $f^* \circ T = f^*$ μ-a.e. and $f_* \circ T = f_*$ μ-a.e.

We have to show

(i) $f^* = f_*$ μ-a.e

(ii) $f^* \in L^1(X, \mathcal{B}, \mu)$

(iii) $\int f^* \, d\mu = \int f \, d\mu$.

We prove (i). For $\alpha, \beta \in \mathbb{R}$, define

$$E_{\alpha,\beta} = \{x \in X \mid f_*(x) < \beta \text{ and } f^*(x) > \alpha\}.$$

Note that

$$\{x \in X \mid f_*(x) < f^*(x)\} = \bigcup_{\beta<\alpha,\ \alpha,\beta\in\mathbb{Q}} E_{\alpha,\beta}$$

(a countable union). Thus, to show that $f^* = f_*$ μ-a.e., it suffices to show that $\mu(E_{\alpha,\beta}) = 0$ whenever $\beta < \alpha$. Since $f_* \circ T = f_*$ and $f^* \circ T = f^*$, we see that $T^{-1}E_{\alpha,\beta} = E_{\alpha,\beta}$. If we write

$$M_\alpha = \left\{ x \in X \mid \sup_{n\geq 1} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) > \alpha \right\}$$

then $E_{\alpha,\beta} \cap M_\alpha = E_{\alpha,\beta}$.

Applying Corollary 5.14 we have that

$$\begin{aligned} \int_{E_{\alpha,\beta}} f \, d\mu &= \int_{E_{\alpha,\beta}\cap M_\alpha} f \, d\mu \\ &\geq \alpha\mu(E_{\alpha,\beta} \cap M_\alpha) = \alpha\mu(E_{\alpha,\beta}). \end{aligned}$$

Replacing f, α and β by $-f$, $-\beta$ and $-\alpha$ and using the fact that $(-f)^* = -f_*$ and $(-f)_* = -f^*$, we also get

$$\int_{E_{\alpha,\beta}} f \, d\mu \leq \beta\mu(E_{\alpha,\beta}).$$

Therefore

$$\alpha\mu(E_{\alpha,\beta}) \leq \beta\mu(E_{\alpha,\beta})$$

and since $\beta < \alpha$ this shows that $\mu(E_{\alpha,\beta}) = 0$. Thus $f^* = f_*$ μ-a.e. and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = f^*(x) \ \mu\text{-a.e.}$$

We prove (ii). Let

$$g_n(x) = \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \right|.$$

Then $g_n \geq 0$ and

$$\int g_n \, d\mu \leq \int |f| \, d\mu$$

so we can apply Fatou's Lemma (Proposition 5.15) to conclude that $\lim_{n\to\infty} g_n = |f^*|$ is integrable, i.e., that $f^* \in L^1(X, \mathcal{B}, \mu)$.

We prove (iii). For $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, define

$$D_k^n = \left\{ x \in X \mid \frac{k}{n} \leq f^*(x) < \frac{k+1}{n} \right\}.$$

For every $\varepsilon > 0$, we have that

$$D_k^n \cap M_{\frac{k}{n}-\varepsilon} = D_k^n.$$

Since $T^{-1}D_k^n = D_k^n$, we can apply Corollary 5.14 again to obtain

$$\int_{D_k^n} f \, d\mu \geq \left(\frac{k}{n} - \varepsilon \right) \mu(D_k^n).$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\int_{D_k^n} f \, d\mu \geq \frac{k}{n} \, \mu(D_k^n).$$

Thus

$$\int_{D_k^n} f^* \, d\mu \leq \frac{k+1}{n} \, \mu(D_k^n) \leq \frac{1}{n} \, \mu(D_k^n) + \int_{D_k^n} f \, d\mu$$

(where the first inequality follows from the definition of D_k^n). Since

$$X = \bigcup_{k \in \mathbb{Z}} D_k^n$$

(a disjoint union), summing over $k \in \mathbb{Z}$ gives

$$\begin{aligned} \int_X f^* \, d\mu &\leq \frac{1}{n} \, \mu(X) + \int_X f \, d\mu \\ &= \frac{1}{n} + \int_X f \, d\mu. \end{aligned}$$

Since this holds for all $n \geq 1$, we obtain

$$\int_X f^* \, d\mu \leq \int_X f \, d\mu.$$

Applying the same argument to $-f$ gives

$$\int (-f)^* \, d\mu \leq \int -f \, d\mu$$

so that

$$\int f^* \, d\mu = \int f_* \, d\mu \geq \int f \, d\mu.$$

Therefore

$$\int f^* \, d\mu = \int f \, d\mu,$$

as required.

Finally, we prove that $f^* = E(f \mid \mathcal{I})$. First note that as f^* is T-invariant, it is measurable with respect to $\mathcal{I}$. Moreover, if I is any T-invariant set then

$$\int_I f \, d\mu = \int_I f^* \, d\mu.$$

Hence $f^* = E(f \mid \mathcal{I})$. □

§5.9 References

Most of the material in this lecture is standard in ergodic theory. The presentation of Ehrenfests' example (originally an example in statistical mechanics) is taken from

K. Petersen, *Ergodic Theory*, C.U.P., Cambridge, 1983.

Additional applications of the Ergodic Theorem to continued fractions can be found in

I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.

A.M. Rockett and P. Szusz, *Continued Fractions*, World Scientific, 1992.

§5.10 Exercises

Exercise 5.1
Construct an example to show that Poincaré's recurrence theorem does not hold on infinite measure spaces. (Recall that a measure space $(X, \mathcal{B}, \mu)$ is infinite if $\mu(X) = \infty$.)

Exercise 5.2
Prove Proposition 5.10: For Lebesgue-almost every point $x \in [0, 1]$, the arithmetic mean of the digits occurring in the base r expansion of x is $(r - 1)/2$.

Exercise 5.3

(i) *Let $T : X \to X$ be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Let $f : X \to \mathbb{R}$ be measurable, and suppose that $\int f \, d\mu = \infty$. Prove that*

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

for μ-almost every x.

(ii) *Prove Proposition 5.12(i).*

Exercise 5.4

Let B be a Banach space and U a bounded linear operator of B such that $\sup_k \|U^k\| < \infty$. Prove that the following are equivalent:

(i) $\frac{1}{n} \sum_{j=0}^{n-1} U^j v$ *converges in norm;*

(ii) $\frac{1}{n} \sum_{j=0}^{n-1} U^j v$ *has a limit point in the weak topology,*

(iii) *U has a fixed point in the weakly closed convex hull of $\{U^n v\}$ (i.e. the smallest weakly closed convex set that contains all the $U^n v$).*

Hence prove the L^p-Ergodic Theorem when T is an ergodic measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$, namely that if $f \in L^p(X, \mathcal{B}, \mu)$ then

$$\frac{1}{n} \sum_{j=0}^{n-1} f T^j \to \int f \, d\mu$$

in L^p.