# 4. Ergodicity and mixing

## §4.1 Introduction

In the previous lecture we defined what is meant by an invariant measure. In this lecture, we define what is meant by an ergodic measure. The primary motivation for ergodicity is that in this setting Birkhoff's Ergodic Theorem has a particularly simple statement: if $T$ is an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$ then, for each $f \in L^1(X, \mathcal{B}, \mu)$ we have that

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

as $n \to \infty$, for $\mu$-a.e. $x \in X$. Checking that a given measure-preserving transformation is ergodic is often a non-trivial task. We will discuss other, stronger, properties that a measure-preserving transformation may enjoy, and that in some cases are easier to check.

## §4.2 Ergodicity

We define what it means to say that a measure-preserving transformation is ergodic.

**Definition.** Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving transformation. We say that $T$ is an *ergodic* transformation (or that $\mu$ is an *ergodic measure*) if whenever $B \in \mathcal{B}$ satisfies $T^{-1}B = B$ then $\mu(B) = 0$ or 1.

**Remark** Ergodicity can be viewed as an indecomposability condition. If ergodicity does not hold then we can find a set $B \in \mathcal{B}$ such that $T^{-1}B = B$ and $0 < \mu(B) < 1$. We can then split $T : X \to X$ into $T : B \to B$ and $T : X \setminus B \to X \setminus B$ with invariant probability measures $\frac{1}{\mu(B)}\mu(\cdot \cap B)$ and $\frac{1}{1-\mu(B)}\mu(\cdot \cap (X \setminus B))$, respectively.

It will sometimes be convenient for us to weaken the condition $T^{-1}B = B$ to $\mu(T^{-1}B \triangle B) = 0$, where $\triangle$ denotes the symmetric difference:

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

We will often write that $A = B$ $\mu$-a.e. or $A = B \bmod 0$ to mean that $\mu(A \triangle B) = 0$.

**Remark** It is easy to see that if $A = B$ $\mu$-a.e. then $\mu(A) = \mu(B)$.

**Lemma 4.1**
*Suppose that $B \in \mathcal{B}$ is such that $\mu(T^{-1}B \triangle B) = 0$. Then there exists $B' \in \mathcal{B}$ with $T^{-1}B' = B'$ and $\mu(B \triangle B') = 0$. (In particular, $\mu(B) = \mu(B')$.)*

**Proof.** For each $n \geq 0$, we have the inclusion

$$\begin{aligned} T^{-n}B \triangle B &\subset \bigcup_{j=0}^{n-1} \left( T^{-(j+1)}B \triangle T^{-j}B \right) \\ &= \bigcup_{j=0}^{n-1} T^{-j}(T^{-1}B \triangle B). \end{aligned}$$

Hence, as $T$ preserves $\mu$,

$$\mu(T^{-n}B \triangle B) \leq n\mu(T^{-1}B \triangle B) = 0.$$

Let

$$B' = \bigcap_{n=0}^{\infty} \bigcup_{j=n}^{\infty} T^{-j}B.$$

We have that

$$\mu\left( B \triangle \bigcup_{j=n}^{\infty} T^{-j}B \right) \leq \sum_{j=n}^{\infty} \mu(B \triangle T^{-n}B) = 0.$$

Since the sets $\bigcup_{j=n}^{\infty} T^{-j}B$ decrease as $n$ increases we hence have $\mu(B \triangle B') = 0$. Also,

$$\begin{aligned} T^{-1}B' &= \bigcap_{n=0}^{\infty} \bigcup_{j=n}^{\infty} T^{-(j+1)}B \\ &= \bigcap_{n=0}^{\infty} \bigcup_{j=n+1}^{\infty} T^{-j}B = B', \end{aligned}$$

as required. □

**Corollary 4.2**
*If $T$ is ergodic and $\mu(T^{-1}B \triangle B) = 0$ then $\mu(B) = 0$ or $1$.*

We have the following convenient characterisation of ergodicity.

**Proposition 4.3**
*Let $T$ be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. The following are equivalent:*

*(i) $T$ is ergodic;*

*(ii) whenever $f \in L^1(X, \mathcal{B}, \mu)$ satisfies $f \circ T = f$ $\mu$-a.e. we have that $f$ is constant $\mu$-a.e.*

**Remark** We can replace $L^1$ in Proposition 4.3(ii) by measurable or by $L^2$.

**Proof.** We prove that (i) implies (ii). Suppose that $T$ is ergodic. Suppose that $f \in L^1(X, \mathcal{B}, \mu)$ is such that $f \circ T = f$ $\mu$-a.e. For $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, define

$$X(k,n) = \left\{ x \in X \mid \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\} = f^{-1}\left( \left[ \frac{k}{2^n}, \frac{k+1}{2^n} \right) \right).$$

Since $f$ is measurable, we have that $X(k,n) \in \mathcal{B}$.

We have that

$$T^{-1}X(k,n) \triangle X(k,n) \subset \{x \in X \mid f(Tx) \neq f(x)\}$$

so that

$$\mu(T^{-1}X(k,n) \triangle X(k,n)) = 0.$$

Hence $\mu(X(k,n)) = 0$ or $\mu(X(k,n)) = 1$.

For each fixed $n$, the union $\bigcup_{k \in \mathbb{Z}} X(k,n)$ is equal to $X$ up to a set of measure zero, i.e.,

$$\mu\left( X \triangle \bigcup_{k \in \mathbb{Z}} X(k,n) \right) = 0;$$

moreover, this union is disjoint. Hence we have

$$\sum_{k \in \mathbb{Z}} \mu(X(k,n)) = \mu(X) = 1$$

and so there is a unique $k_n$ for which $\mu(X(k_n, n)) = 1$. Let

$$Y = \bigcap_{n=1}^{\infty} X(k_n, n).$$

Then $\mu(Y) = 1$ and, by construction, $f$ is constant on $Y$, i.e., $f$ is constant $\mu$-a.e.

Conversely, we prove that (ii) implies (i). Suppose that $B \in \mathcal{B}$ is such that $T^{-1}B = B$. Then $\chi_B \in L^1(X, \mathcal{B}, \mu)$ and $\chi_B \circ T(x) = \chi_B(x)$ for all $x \in X$. Hence $\chi_B$ is constant $\mu$-a.e. Since $\chi_B$ only takes the values 0 and 1, we must have $\chi_B = 0$ $\mu$-a.e. or $\chi_B = 1$ $\mu$-a.e. Therefore

$$\mu(B) = \int_X \chi_B \, d\mu = 0 \text{ or } 1,$$

and $T$ is ergodic. $\square$

### §4.3 Examples: Using Fourier series to prove ergodicity

In the previous lecture we studied a number of examples of dynamical systems defined on the circle or the torus and we proved that Lebesgue measure is invariant. We show how Proposition 4.3 can be used in conjuction with Fourier Series to determine whether Lebesgue measure is ergodic.

#### §4.3.1 Rotations on a circle

Fix $\alpha \in \mathbb{R}$ and define $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by $T(x) = x + \alpha \bmod 1$. We have already seen that $T$ preserves Lebesgue measure. The following result gives a necessary and sufficient condition for $T$ to be ergodic.

**Theorem 4.4**
*Let $T(x) = x + \alpha \bmod 1$.*

*(i) If $\alpha \in \mathbb{Q}$ then $T$ is not ergodic with respect to Lebesgue measure.*

*(ii) If $\alpha \notin \mathbb{Q}$ then $T$ is ergodic with respect to Lebesgue measure.*

**Proof.** Suppose that $\alpha \in \mathbb{Q}$ and write $\alpha = p/q$ for $p, q \in \mathbb{Z}$ with $q \neq 0$. Define

$$f(x) = e^{2\pi iqx} \in L^2(X, \mathcal{B}, \mu).$$

Then $f$ is not constant but

$$f(Tx) = e^{2\pi iq(x+p/q)} = e^{2\pi i(qx+p)} = e^{2\pi iqx} = f(x).$$

Hence $T$ is not ergodic.

Suppose that $\alpha \notin \mathbb{Q}$. Suppose that $f \in L^2(X, \mathcal{B}, \mu)$ is such that $f \circ T = f$ a.e. Suppose that $f$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi inx}.$$

Then $f \circ T$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi in\alpha} e^{2\pi inx}.$$

Comparing Fourier coefficients we see that

$$c_n = c_n e^{2\pi in\alpha},$$

for all $n \in \mathbb{Z}$. As $\alpha \notin \mathbb{Q}$, we see that $e^{2\pi in\alpha} \neq 1$ unless $n = 0$. Hence $c_n = 0$ for $n \neq 0$. Hence $f$ has Fourier series $c_0$, i.e. $f$ is constant a.e. □

### §4.3.2 The doubling map

Let $X = \mathbb{R}/\mathbb{Z}$. Recall that if $f \in L^2(X, \mathcal{B}, \mu)$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}$$

then the Riemann-Lebesgue lemma tells us that $c_n \to 0$ as $n \to \infty$.

**Proposition 4.5**
*The doubling map $T : X \to X$ defined by $T(x) = 2x \bmod 1$ is ergodic with respect to Lebesgue measure $\mu$.*

**Proof.** Let $f \in L^2(X, \mathcal{B}, \mu)$ and suppose that $f \circ T = f$ $\mu$-a.e. Let $f$ have Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x} \text{ (in } L^2).$$

For each $p \geq 0$, $f \circ T^p$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n 2^p x}.$$

Comparing Fourier coefficients we see that

$$c_n = c_{2^p n}$$

for all $n \in \mathbb{Z}$ and each $p = 0, 1, 2, \ldots$. By the Riemann-Lebesgue lemma, for each $n \neq 0$, $c_{2^p n} \to 0$ as $p \to \infty$. Hence $c_n = 0$ for $n \neq 0$. Thus $f$ has Fourier series $c_0$, and so must be equal to a constant a.e. Hence $T$ is ergodic with respect to $\mu$. □

### §4.3.3 Toral endomorphisms

The argument for the doubling map can be generalised using higher-dimensional Fourier series to study toral endomorphisms. Let $X = \mathbb{R}^k/\mathbb{Z}^k$ and let $\mu$ denote Lebesgue measure. Recall that $f \in L^2(X, \mathcal{B}, \mu)$ has Fourier series

$$\sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle n, x \rangle},$$

where $n = (n_1, \ldots, n_k)$, $x = (x_1, \ldots, x_k)$. Define $|n| = \max_{1 \leq j \leq n} |n_j|$. Then the Riemann Lebesgue lemma tells us that $c_n \to 0$ as $|n| \to \infty$.

Let $A$ be a $k \times k$ integer matrix with $\det A = \pm 1$ and define $T : X \to X$ by

$$T(x_1, \ldots, x_k) = A(x_1, \ldots, x_k) \bmod 1.$$

**Proposition 4.6**
*A linear toral automorphism $T$ is ergodic with respect to $\mu$ if and only if no eigenvalue of $A$ is a root of unity.*

**Remark** In particular, hyperbolic toral automorphisms (i.e. no eigenvalues of modulus 1) are ergodic with respect to Lebesgue measure.

**Proof.** We prove that (i) implies (ii). Suppose that $T$ is ergodic but, for a contradition, that $A$ has a $p$th root of unity as an eigenvalue. We choose $p > 0$ to be the least such integer. Then $A^p$ has 1 as an eigenvalue, and so $n(A^p - I) = 0$ for some non-zero vector $n = (n_1, \ldots, n_k) \in \mathbb{R}^k$. Since $A$ is an integer matrix, we have that $A^p - I$ is an integer matrix, and so we can in fact take $n \in \mathbb{Z}^k$. Note that

$$e^{2\pi i \langle n, A^p x\rangle} = e^{2\pi i \langle nA^p, x\rangle} = e^{2\pi i \langle n, x\rangle}.$$

Define

$$f(x) = \sum_{j=0}^{p-1} e^{2\pi i \langle n, A^j x\rangle}.$$

Then $f \in L^2(X, \mathcal{B}, \mu)$ and is $T$-invariant. Since $T$ is ergodic, we must have that $f$ is constant. But the only way in which this can happen is if $n = 0$, a contradiction.

We prove that (ii) implies (i). Suppose that $f \in L^2(X, \mathcal{B}, \mu)$ is $T$-invariant $\mu$-a.e. We show that $f$ is constant $\mu$-a.e. Associate to $f$ its Fourier series:

$$\sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle n, x\rangle}.$$

Since $fT^p = f$ $\mu$-a.e., for all $p > 0$, we have that

$$\sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle nA^p, x\rangle} = \sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle n, x\rangle}.$$

Comparing Fourier coefficients we see that, for every $n \in \mathbb{Z}^k$,

$$c_n = c_{nA} = \cdots = c_{nA^p} = \cdots.$$

If $c_n \neq 0$ then there can only be finitely many indices in the above list, for otherwise it would contradict the fact that $c_n \to 0$ as $|n| \to \infty$, by the Riemann-Lebesgue lemma. Hence there exists $q_1 > q_2 > 0$ such that $nA^{q_1} = nA^{q_2}$. Letting $p = q_1 - q_2 > 0$ we see that $nA^p = n$. Thus $n$ is either equal to 0 or is an eigenvector for $A^p$ with eigenvalue 1. In the latter case, $A$ would have a $p$th root of unity as an eigenvalue; hence $n = 0$. Hence $c_n = 0$ unless $n = 0$ and so $f$ is equal to the constant $c_0$ $\mu$-a.e. Thus $T$ is ergodic. □

## §4.4 Using the Kolmogorov Extension Theorem to prove ergodicity

We illustrate a method for proving that a given transformation is ergodic using the Kolmogorov Extension Theorem. The key observation is the following technical lemma.

**Lemma 4.7**
*Let $(X, \mathcal{B}, \mu)$ be a probability space and suppose that $\mathcal{A} \subset \mathcal{B}$ is an algebra that generates $\mathcal{B}$. Suppose there exists $K > 0$ such that*

$$\mu(B)\mu(I) \leq K\mu(B \cap I) \tag{4.1}$$

*for all $I \in \mathcal{A}$. Then $\mu(B) = 0$ or $1$.*

**Proof.** Let $\varepsilon > 0$. As $\mathcal{A}$ generates $\mathcal{B}$ there exists $I \in \mathcal{A}$ such that $\mu(B^c \triangle I) < \varepsilon$. Hence $|\mu(B^c) - \mu(I)| < \varepsilon$. Moreover, note that $B \cap I \subset B^c \triangle I$ so that $\mu(B \cap I) < \varepsilon$. Hence

$$\mu(B)\mu(B^c) \leq \mu(B)(\mu(I)+\varepsilon) \leq \mu(B)\mu(I)+\mu(B)\varepsilon \leq K\mu(B\cap I)+\varepsilon \leq (K+1)\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, it follows that $\mu(B)\mu(B^c) = 0$. Hence $\mu(B) = 0$ or 1. □

**Remark** We will often apply Lemma 4.7 when $\mathcal{A}$ is an algebra of finite unions of intervals or cylinders. In this case, we need only check that there exists a constant $K > 0$ such that (4.1) holds for intervals or cylinders. To see this, let $I = \bigcup_{j=1}^{k} I_j$ be a finite union of disjoint sets in $\mathcal{A}$. Then if (4.1) holds for $I_j$ then

$$\begin{aligned} \mu(B)\mu(I) &= \mu(B)\mu\left(\bigcup_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} \mu(B)\mu(I_j) \\ &\leq K\sum_{j=1}^{k} \mu(B \cap I_j) = K\mu\left(B \cap \bigcup_{j=1}^{k} I_j\right) = K\mu(B \cap I). \end{aligned}$$

We will also use the change of variables formula for integration. Recall that if $I, J \subset \mathbb{R}$ are intervals, $u : I \to J$ is a differentiable bijection, and $f : J \to \mathbb{R}$ is integrable, then

$$\int_J f(x)\,dx = \int_I f(u(x))|u'(x)|\,dx.$$

### §4.4.1 Bernoulli shifts

Let $S = \{1, \ldots, k\}$ be a finite set of symbols and let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in \{1, 2, \ldots, k\}\}$ denote the shift space on $k$ symbols. Let $\sigma : \Sigma \to \Sigma$ denote the left shift map, so that $(\sigma(\mathbf{x}))_j = x_{j+1}$.

Recall that we defined the cylinder $[i_0, \ldots, i_{n-1}]$ to be the set of all sequences in $\Sigma$ that start with symbols $i_0, \ldots, i_{n-1}$, that is

$$[i_0, \ldots, i_{n-1}] = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \in \Sigma \mid x_j = i_j,\ j = 0, 1, \ldots, n-1\}.$$

Let $p = (p(1), \ldots, p(k))$ be a probability vector (that is, $p(j) > 0$, $\sum_{j=1}^{k} p(j) = 1$). We defined the Bernoulli measure $\mu_p$ on cylinders by setting

$$\mu_p[i_0, \ldots, i_{n-1}] = p(i_0)p(i_1)\cdots p(i_{n-1}).$$

We have already seen that $\mu_p$ is a $\sigma$-invariant measure.

**Proposition 4.8**
*Let $\mu_p$ be a Bernoulli measure. Then $\mu_p$ is ergodic.*

**Proof.** We first make the following observation: let $I = [i_0, \ldots, i_{p-1}]$, $J = [j_0, \ldots, j_{q-1}]$ be cylinders of ranks $p, q$, respectively. Consider $I \cap \sigma^{-n}J$ where $n \geq p$. Then

$$\begin{aligned} & I \cap \sigma^{-n}J \\ = \; & \{\mathbf{x} = (x_j)_{j=0}^{\infty} \in \Sigma \mid x_j = i_j \text{ for } j = 0, 1, \ldots, p-1, \ x_{j+n} = y_j \text{ for } j = 0, 1, \ldots, q-1\} \\ = \; & \bigcup_{x_p, \ldots, x_{n-1}} [i_0, i_1, \ldots, i_{p-1}, x_p, \ldots, x_{n-1}, j_0, \ldots, j_{q-1}], \end{aligned}$$

a disjoint union. Hence

$$\begin{aligned} & \mu_p(I \cap \sigma^{-n}J) && (4.2) \\ = \; & \sum_{x_p, \ldots, x_{n-1}} \mu_p[i_0, i_1, \ldots, i_{p-1}, x_p, \ldots, x_{n-1}, j_0, \ldots, j_{q-1}] \\ = \; & \sum_{x_p, \ldots, x_{n-1}} p(i_0)p(i_1)\cdots p(i_{n-1})p(x_p)\cdots p(x_{n-1})p(j_0)p(j_1)\cdots p(j_{q-1}) \\ = \; & p(i_0)p(i_1)\cdots p(i_{n-1})p(j_0)p(j_1)\cdots p(j_{q-1}) \text{ as } \sum_{x_p} p(x_p) = \cdots = \sum_{x_{n-1}} p(x_{n-1}) = 1 \\ = \; & \mu_p(I)\mu_p(J). && (4.3) \end{aligned}$$

Let $B \in \mathcal{B}$ be $\sigma$-invariant. By Proposition 4.7 it is sufficient to prove that $\mu_p(B \cap I) \leq \mu_p(B)\mu_p(I)$ for each cylinder $I$. Let $\varepsilon > 0$. We first approximate the invariant set $B$ by a finite union of cylinders. As the algebra of finite unions of cylinders generates the Borel $\sigma$-algebra, we can find a finite disjoint union of cylinders $A = \bigcup_{j=1}^{r} J_j$ such that $\mu_p(B \triangle A) < \varepsilon$. Note that $|\mu_p(A) - \mu_p(B)| < \varepsilon$.

Let $n$ be any integer greater than the rank of $I$. Note that $\sigma^{-n}B \triangle \sigma^{-n}A = \sigma^{-n}(B \triangle A)$. Hence

$$\mu_p(\sigma^{-n}B \triangle \sigma^{-n}A) = \mu_p(\sigma^{-n}(B \triangle A)) = \mu_p(B \triangle A) < \varepsilon,$$

where we have used the facts that $\sigma^{-n}B = B$ and that $\mu_p$ is an invariant measure. Hence $|\mu_p(\sigma^{-n}B) - \mu_p(\sigma^{-n}A)| < \varepsilon$.

As $A = \bigcup_{j=1}^{r} J_j$ is a finite union of cylinders and $n$ is greater than the rank of $I$, it follows from (4.3) that

$$\begin{aligned}
\mu_p(\sigma^{-n}A \cap I) &= \mu_p\left(\sigma^{-n}\left(\bigcup_{j=1}^{r} J_j\right) \cap I\right) = \sum_{j=1}^{r} \mu_p(\sigma^{-n}J_j \cap I) \\
&= \sum_{j=1}^{r} \mu_p(J_j)\mu_p(I) = \mu_p\left(\bigcup_{j=1}^{r} J_j\right)\mu_p(I) \\
&= \mu_p(A)\mu_p(I).
\end{aligned}$$

Finally, note that $(\sigma^{-n}A \cap I)\triangle(\sigma^{-n}B \cap I) \subset (\sigma^{-n}A)\triangle(\sigma^{-n}B)$. Hence $\mu_p((\sigma^{-n}A \cap I)\triangle(\sigma^{-n}B \cap I)) < \varepsilon$ so that $\mu_p(\sigma^{-n}A \cap I) < \mu_p(\sigma^{-n}B \cap I) + \varepsilon$. Hence

$$\begin{aligned}
\mu_p(B)\mu_p(I) &= \mu_p(\sigma^{-n}B)\mu_p(I) \leq \mu_p(\sigma^{-n}A)\mu_p(I) + \epsilon \\
&= \mu_p(\sigma^{-n}A \cap I) + \epsilon \leq \mu_p(\sigma^{-n}B \cap I) + 2\epsilon \\
&= \mu(B \cap I) + 2\epsilon.
\end{aligned}$$

As $\varepsilon > 0$ is arbitrary, we have that $\mu_p(B)\mu_p(I) \leq \mu_p(B \cap I)$ for any cylinder $I$. By Proposition 4.7, it follows that $\mu_p(B) = 0$ or 1. Hence $\mu_p$ is ergodic. □

**Remark** One can use a similar argument to show that Markov measures corresponding to irreducible or aperiodic shifts of finite type are ergodic. One can also show that Bernoulli and irreducible Markov measures for two-sided shifts of finite type are ergodic.

### §4.4.2 The continued fraction map

Let $x \in [0,1]$. If $x$ has continued fraction expansion

$$x = \cfrac{1}{x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cdots}}}$$

then for brevity we write $x = [x_0, x_1, x_2, \ldots]$.

Let $X = [0,1]$ and recall that the Gauss map is defined by $T(x) = 1/x \bmod 1$ (with $T$ defined at 0 by setting $T(0) = 0$). If $x$ has continued fraction expansion $[x_0, x_1, x_2, \ldots]$ then $T(x)$ has continued fraction expansion $[x_1, x_2, \ldots]$. We have already seen that $T$ leaves Gauss' measure $\mu$ invariant, where Gauss' measure is defined by

$$\mu(B) = \frac{1}{\log 2}\int_B \frac{1}{1+x}\,dx.$$

We shall also need some basic facts about continued fractions. Let $x \in (0,1)$ be irrational and have continued fraction expansion $[x_0, x_1, \ldots]$. For any $t \in [0,1]$, write

$$[x_0, x_1, \ldots, x_{n-1} + t] = \frac{P_n(x_0, x_1, \ldots, x_{n-1}; t)}{Q_n(x_0, x_1, \ldots, x_{n-1}; t)}$$

where $P_n(x_0, x_1, \ldots, x_{n-1}; t)$ and $Q_n(x_0, x_1, \ldots, x_{n-1}; t)$ are polynomials in $x_0, x_1, \ldots, x_{n-1}$ and $t$. Let $P_n = P_n(x_0, x_1, \ldots, x_{n-1})$, $Q_n = Q_n(x_0, x_1, \ldots, x_{n-1})$ (we suppress the dependence of $P_n$ and $Q_n$ on $x_0, \ldots, x_{n-1}$ for brevity). The following lemma is easily proved using induction.

**Lemma 4.9**
*(i) We have*

$$P_n(x_0, x_1, \ldots, x_{n-1}; t) = P_n + tP_{n-1},\ Q_n(x_0, x_1, \ldots, x_{n-1}; t) = Q_n + tQ_{n-1}.$$

*and the following recurrence relations hold:*

$$P_{n+1} = x_n P_n + P_{n-1},\ Q_{n+1} = x_n Q_n + Q_{n-1}$$

*with initial conditions* $P_0 = 0, P_1 = 1, Q_0 = 1, Q_1 = x_0$.

*(ii) The following identity holds:*

$$Q_n P_{n-1} - Q_{n-1} P_n = (-1)^n.$$

Let $i_0, i_1, \ldots, i_{n-1} \in \mathbb{N}$. Define the cylinder $I(i_0, i_1, \ldots, i_{n-1})$ to be the set of all points $x \in (0,1)$ whose continued fraction expansion starts with $i_0, \ldots, i_{n-1}$. This is easily seen to be an interval; indeed

$$I(i_0, i_1, \ldots, i_{n-1}) = \{[i_0, i_1, \ldots, i_{n-1} + t] \mid t \in [0,1)\}.$$

Let $\mathcal{A}$ denote the algebra of finite unions of cylinders. Then $\mathcal{A}$ generates the Borel $\sigma$-algebra. (This follows as cylinders are clearly Borel sets and they separate points. To see this, note that if $x \neq y$ then they have different continued fraction expansions. Hence there exists $n$ such that $x_n \neq y_n$. Hence $x, y$ are in different cylinders of rank $n$, and these cylinders are disjoint.)

For each $i \in \mathbb{N}$ define the map

$$\phi_i(x) = \frac{1}{i + x} : [0,1) \to I(i).$$

Thus if $x$ has continued fraction expansion $[x_0, x_1, \ldots]$ then $\phi_i(x)$ has continued fraction expansion $[i, x_0, x_1, \ldots]$. Clearly $T(\phi_i(x)) = x$ for all $x \in [0,1)$.

For $i_0, i_1, \ldots, i_{n-1} \in \mathbb{N}$, define

$$\phi_{i_0, i_1, \ldots, i_{n-1}} = \phi_{i_0} \phi_{i_1} \cdots \phi_{i_{n-1}} : [0,1) \to I(i_0, i_1, \ldots, i_{n-1}).$$

Then $\phi_{i_0,i_1,\ldots,i_{n-1}}(x)$ takes the continued fraction expansion of $x$, shifts every digit $n$ places to the right, and inserts the digit $i_0, i_1, \ldots, i_{n-1}$ in the first $n$ places. Clearly

$$T^n(\phi_{i_0,i_1,\ldots,i_{n-1}}(x)) = x$$

for all $x \in [0, 1)$.

We first need an estimate on the length of (i.e. the Lebesgue measure of) the cylinder $I(i_0, i_1, \ldots, i_{n-1})$. Note that

$$\phi_{i_0,i_1,\ldots,i_{n-1}}(t) = \frac{P_n(i_0,\ldots,i_{n-1};t)}{Q_n(i_0,\ldots,i_{n-1};t)} = \frac{P_n + tP_{n-1}}{Q_n + tQ_{n-1}}.$$

Differentiating this expression with respect to $t$ and using Lemma 4.9(ii), we see that

$$|\phi'_{i_0,i_1,\ldots,i_{n-1}}(t)| = \left|\frac{Q_nP_{n-1} - P_nQ_{n-1}}{(Q_n + tQ_{n-1})^2}\right| = \frac{1}{(Q_n + tQ_{n-1})^2}.$$

It follows from Lemma 4.9(i) that $Q_n + Q_{n-1} \leq 2Q_n$. Hence

$$\frac{1}{4}\frac{1}{Q_n^2} \leq \frac{1}{(Q_n + Q_{n-1})^2} \leq |\phi'_{i_0,i_1,\ldots,i_{n-1}}(t)| \leq \frac{1}{Q_n^2}. \tag{4.4}$$

Hence

$$\lambda(I(i_0, i_1, \ldots, i_{n-1})) = \int \chi_{I(i_0,i_1,\ldots,i_{n-1})}(t)\,dt = \int_{I(i_0,i_1,\ldots,i_{n-1})} dt = \int_0^1 |\phi'_{i_0,i_1,\ldots,i_{n-1}}(t)|\,dt \tag{4.5}$$

where we have used the change of variables formula. Combining (4.5) with (4.4) we see that

$$\frac{1}{4}\frac{1}{Q_n^2} \leq \lambda(I(i_0, i_1, \ldots, i_{n-1})) \leq \frac{1}{Q_n^2}. \tag{4.6}$$

We can now prove that the Gauss map is ergodic with respect to Gauss' measure $\mu$. Suppose that $T^{-1}B = B$ where $B \in \mathcal{B}$. Let $I(i_0, i_1, \ldots, i_{n-1})$ be a cylinder. Then

$$\begin{aligned}
&\lambda(B \cap I(i_0, i_1, \ldots, i_{n-1})) \\
&= \int_{I(i_0,i_1,\ldots,i_{n-1})} \chi_B(x)\,dx \\
&= \int_0^1 \chi_B(\phi_{i_0,i_1,\ldots,i_{n-1}}(x))|\phi'_{i_0,i_1,\ldots,i_{n-1}}(x)|\,dx \text{ by the change of variables formula.} \\
&= \int_0^1 \chi_{T^{-n}B}(\phi_{i_0,i_1,\ldots,i_{n-1}}(x))|\phi'_{i_0,i_1,\ldots,i_{n-1}}(x)|\,dx \text{ as } T^{-n}B = B \\
&= \int_0^1 \chi_B(T^n(\phi_{i_0,i_1,\ldots,i_{n-1}}(x)))|\phi'_{i_0,i_1,\ldots,i_{n-1}}(x)|\,dx \text{ as } \chi_{T^{-n}B} = \chi_B \circ T^n \\
&= \int_0^1 \chi_B(x)|\phi'_{i_0,i_1,\ldots,i_{n-1}}(x)|\,dx \text{ as } T^n\phi_{i_0,i_1,\ldots,i_{n-1}}(x) = x.
\end{aligned}$$

By (4.4) and (4.6) it follows that

$$\lambda(B \cap I(i_0, i_1, \ldots, i_{n-1})) \geq \frac{1}{4Q_n^2}\lambda(B) \geq \frac{1}{4}\lambda(B)\lambda(I(i_0, i_1, \ldots, i_{n-1}))$$

so that

$$\lambda(B)\lambda(I(i_0, i_1, \ldots, i_{n-1})) \leq 4\lambda(B \cap I(i_0, i_1, \ldots, i_{n-1})).$$

By Lemma 4.7 it follows that $\lambda(B) = 0$ or 1. Hence, as Lebesgue measure and Gauss' measure have the same sets of measure zero, it follows that either $\mu(B) = 0$ or $\mu(B) = 1$. Hence $T$ is ergodic with respect to Gauss' measure.

## §4.5 Mixing

There are many other properties that a given measure-preserving transformation $T$ may enjoy that imply ergodicity. We briefly describe some of the more important properties below.

Let $T$ be a measure-presering transformation of a probability space $(X, \mathcal{B}, \mu)$. We will see in the next lecture that if $T$ is ergodic with respect to $\mu$, then the following result holds:

**Theorem 4.10 (Birkhoff's Ergodic Theorem)**
*Let $T$ be an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) = \int f\, d\mu \tag{4.7}$$

*for $\mu$-a.e. $x \in X$.*

The following lemma follows easily from Theorem 4.10.

**Lemma 4.11**
*Let $T$ be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then $T$ is ergodic if and only if for all $A, B \in \mathcal{B}$ we have*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A)\mu(B). \tag{4.8}$$

**Proof.** Take $f = \chi_A$ in (4.7). Multiply (4.7) by $\chi_B$, and recall that $\chi_A \circ T^j \cdot \chi_B = \chi_{T^{-j}A\cap B}$. By the Lebesgue Dominated Convergence Theorem, we can integrate the resulting expression to obtain (4.8).

Conversely, suppose that $T^{-1}B = B$, $B \in \mathcal{B}$. Taking $A = B$ in (4.8) we have that $\mu(B)^2 = \mu(B)$. Hence $\mu(B) = 0$ or 1. □

Recall from abstract probability theory that two events $A, B$ are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$. Also recall that a sequence $c_n$ is said to *Cesàro converge* to $a$ if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} c_n = a.$$

Thus $T$ is ergodic if and only if the Cesàro averages of the sequence $\mu(T^{-n}A\cap B)$ converge to $\mu(A)\mu(B)$. That is, given two sets $A, B \in \mathcal{B}$, the sets $T^{-n}A$, $B$ become independent as $n \to \infty$ in some appropriate sense. We can change the sense in which $T^{-n}A$, $B$ become independent to obtain the following definitions.

**Definition.** Let $T$ be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$.

(i) We say that $T$ is *weak-mixing* if, for all $A, B \in \mathcal{B}$, we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}A \cap B) - \mu(A)\mu(B)| = 0. \tag{4.9}$$

(ii) We say that $T$ is *strong-mixing* if, for all $A, B \in \mathcal{B}$, we have

$$\lim_{n\to\infty} \mu(T^{-n}A) \cap B) = \mu(A)\mu(B). \tag{4.10}$$

**Remark** It is easy to see that strong-mixing implies weak-mixing, and that weak-mixing implies ergodicity.

**Proposition 4.12**

*(i) Let $T$ be an exact transformation of the probability space $(X, \mathcal{B}, \mu)$. Then $T$ is strong-mixing.*

*(ii) Let $T$ be a $K$-automorphism of the probability space $(X, \mathcal{B}, \mu)$. Then $T$ is strong-mixing.*

## §4.6 Spectral theory

Let $(X, \mathcal{B}, \mu)$ be a probability space. Recall that the space $L^2(X, \mathcal{B}, \mu)$ of complex-valued square-integrable functions is a Hilbert space with respect to the inner-product

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

Let $T$ be a measurable transformation of a probability space $(X, \mathcal{B}, \mu)$. Then $T$ induces a linear operator $U$ on $L^2(X, \mathcal{B}\, mu)$ by defining $Uf = f \circ T$. One can often relate ergodic-theoretic properties of $T$ with spectral properties of $U$.

**Proposition 4.13**
*The transformation $T$ is measure-preserving if and only if the operator $U$ is an isometry: i.e. $\langle Uf, Ug\rangle = \langle f, g\rangle$ for all $f, g \in L^2(X, \mathcal{B}, \mu)$. Moreover, if $T$ is invertible and measure-preserving then $U$ is unitary.*

**Proof.** If $T$ is measure-presering then clearly

$$\langle Uf, Ug\rangle = \int (f \circ T)(g \bar{\circ} T)\, d\mu = \int f\bar{g} \circ T\, d\mu = \int f\bar{g}\, d\mu = \langle f, g\rangle.$$

Hence $U$ is an isometry. if $T$ is invertible that $U^{-1}f = f \circ T^{-1}$, so that $U$ is unitary.

Conversely, if $U$ is an isometry then, noting that $U1 = 1$ where 1 denotes the function that is constantly equal to 1, we have that

$$\int f \circ T\, d\mu = \langle Uf, U1\rangle = \langle f, 1\rangle = \int f\, d\mu.$$

Hence $T$ is measure-preserving. □

Recall that the eigenvalues of a unitary operator acting on a Hilbert space all have modulus 1. Recall that an eigenvalue is said to be *simple* if the corresponding eigenspace is 1-dimensional. The following is a restatement of Proposition 4.3.

**Proposition 4.14**
*Let $T : X \to X$ be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then $T$ is ergodic with respect to $\mu$ if and only if 1 is a simple eigenvalue for $U$.*

**Proof.** Notice that 1 is always an eigenvalue for $U$ as $U$ fixes the constant functions. Recalling Proposition 4.3 we know that $T$ is ergodic if and only if the only $L^2$ functions $f$ for which $f \circ T = f$ are the constants, the result follows immediately. □

We have the following characterisation of weak-mixing.

**Proposition 4.15**
*Let $T$ be an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. The following are equivalent:*

*(i) $T$ is weak-mixing;*

*(ii) whenever $f \in L^2(X, \mathcal{B}, \mu)$ satisfies $f \circ T = \alpha f$ $\mu$-a.e. where $\alpha \in \mathbb{C}$, $|\alpha| = 1$, we have that $f$ is constant $\mu$-a.e.*

**Proof.** See any standard text on ergodic theory. □

## §4.7 Bernoulli transformations

**Definition.** We say that two measure-preserving transformations $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, m, S)$ are *(measure theoretically) isomorphic* if there exist $M \in \mathcal{B}$ and $N \in \mathcal{C}$ such that

(i) $TM \subset M$, $SN \subset N$,

(ii) $\mu(M) = 1$, $m(N) = 1$,

and there exists a bijection $\phi : M \to N$ such that

(i) $\phi$, $\phi^{-1}$ are measurable and measure-preserving (i.e. $\mu(\phi^{-1}A) = m(A)$ for all $A \in \mathcal{C}$),

(ii) $\phi \circ T = S \circ \phi$.

We say that a measure-preserving transformation is *Bernoulli* if it isomorphic to a Bernoulli $(p_1, \ldots, p_k)$-shift, for some $(p_1, \ldots, p_k)$. If $T$ is non-invertible then the shift is understood to be one-sided, and if $T$ is invertible then the shift if understood to be two-sided.

**Example.** The construction of symbolic dynamics for the doubling map in Lecture 1 shows that the doubling map is isomorphic to the Bernoulli $(1/2, 1/2)$-shift.

For non-invertible transformations we have the following hierachy:

$$\text{Bernoulli} \Rightarrow \text{strong-mixing} \Rightarrow \text{weak-mixing} \Rightarrow \text{ergodic.}$$

For invertible transformations this hierachy has the form:

$$\text{Bernoulli} \Rightarrow \text{strong-mixing} \Rightarrow \text{weak-mixing} \Rightarrow \text{ergodic.}$$

There are examples to show that none of the implications can be reversed.

## §4.8 References

All of the material in this lecture is standard in ergodic theory, and can be found in, for example,

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

W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.

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

P. Walters, *An introduction to ergodic theory*, Springer, Berlin, 1982.

It is straightforward to find examples of transformations that are ergodic but not weak-mixing. Examples of transformations that are weak-mixing but not strong-mixing have been constructed by Kakutani, Chacon, Katok & Stepin, and others. See the references in Walters' book above.

## §4.9 Exercises

**Exercise 4.1**
*Let $T$ be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Prove that each of the following two conditions is equivalent to the ergodicity of $T$.*

*(i) For every $B \in \mathcal{B}$ with $\mu(B) > 0$ we have*

$$\mu\left(\bigcup_{j=1}^{\infty} T^{-j}B\right) = 1.$$

*(ii) For every $A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0$ there exists $n > 0$ with $\mu(T^{-n}A \cap B) > 0$.*

*In particular, (ii) says that ergodicity is equivalent to the following notion of recurrence: given any two sets of positive measure, the orbit of almost every point in the first set will hit the second set (and the time at which this happens depends only on the sets).*

**Exercise 4.2**
*It is straightforward to construct hyperbolic toral automorphisms (i.e. no eigenvalues of modulus 1—the cat map is such an example), which must necessarily be ergodic with respect to Lebesgue measure. It is harder to show that there are ergodic toral automorphisms with some eigenvalues of modulus 1.*

*(i) Show that to have ergodic toral automorphism of $\mathbb{R}^k/\mathbb{Z}^k$ with an eigenvalue of modulus 1, we must have $k \geq 4$.*

*(ii) Show that the linear toral automorphism of the 4-dimensional torus induced by the matrix*

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 8 & -6 & 8 \end{pmatrix}.$$

*is ergodic.*

**Exercise 4.3**
*Let $T(x, y) = (x + \alpha \bmod 1, x + y)$ be defined on $\mathbb{R}^2/\mathbb{Z}^2$. Describe when Lebesgue measure $\mu$ is invariant, ergodic, or weak-mixing. What are the eigenvalues of the induced operator $U : L^2(\mathbb{R}^2/\mathbb{Z}^2, \mathcal{B}, \mu) \to L^2(\mathbb{R}^2/\mathbb{Z}^2, \mathcal{B}, \mu)$?*

**Exercise 4.4**
*Prove, using induction, Lemma 4.9.*

**Exercise 4.5**
*Use the method in §4.4 to show that Lebesgue measure is an ergodic measure for the doubling map.*

**Exercise 4.6**
*The Perron-Frobenius theorem states the following:*

> *Suppose that* $P$ *is a* $k \times k$ *aperiodic matrix stochastic matrix. (Aperiodic means: there exists* $n \geq 1$ *such that* $P^n(i,j) > 0$ *for all* $i, j$*. Stochastic means that* $0 \leq P(i,j) \leq 1$ *for all* $i, j$ *and each row of* $P$ *sums to 1.) Then 1 is a simple eigenvalue for* $P$ *and all other eigenvalues of* $P$ *have modulus less than 1. There exists a unique left-eigenvector* $p = (p(1), \ldots, p(k))$ *such that* $pP = p$*. Moreover, for all* $i, j$*,* $P^n(i,j) \to p(j)$ *as* $n \to \infty$*.*

*Let* $\Sigma_A^+$ *be a one-sided aperiodic shift of finite type with transition matrix* $A$*. Suppose that* $P$ *is a stochastic matrix that is compatible with* $A$ *(so that* $A(i,j) = 1 \Leftrightarrow P(i,j) > 0$*). Recall that the Markov measure* $\mu_P$ *is defined on ccylinders by*

$$\mu_P([i_0, i_1, \ldots, i_n]) = p(i_0)P(i_0, i_1) \cdots P(i_{n-1}, i_n).$$

*Suppose that* $I, J$ *are cylinders. Prove that* $\mu_P(I \cap \sigma^{-n} J) \to \mu_P(I)\mu_P(J)$ *as* $n \to \infty$*.*

*Hence use the method in §4.4 to prove that* $\mu_P$ *is ergodic.*