CHAPTER 9

ERGODIC MEASURES

In this chapter we shall consider the stronger property of ergodicity for an invariant probability measure μ. This property is more appropriate (amongst other things) for understanding the "long term" average behaviour of a transformation.

9.1 Definitions and characterization of ergodic measures

DEFINITION. Given a probability space $(X, \mathcal{B}, \mu)$, a transformation $T : X \to X$ is called *ergodic* if for every set $B \in \mathcal{B}$ with $T^{-1}B = B$ we have that either $\mu(B) = 0$ or $\mu(B) = 1$.

Alternatively we say that μ is T-ergodic.

The following lemma gives a simple characterization in terms of functions.

LEMMA 9.1. *T is ergodic with respect to μ iff whenever $f \in L^1(X, \mathcal{B}, \mu)$ satisfies $f = f \circ T$ then f is a constant function.*

PROOF. This is an easy observation using indicator functions.

■

9.2 Poincaré recurrence and Kac's theorem

We begin with one of the most fundamental results in ergodic theory.

THEOREM 9.2 (POINCARÉ RECURRENCE THEOREM). *Let $T : X \to X$ be a measurable transformation on a probability space $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}$ have $\mu(A) > 0$; then for almost points $x \in A$ the orbit $\{T^n x\}_{n \geq 0}$ returns to A infinitely often.*

PROOF. Let $F = \{x \in A : T^n x \notin A, \forall n \geq 1\}$, then it suffices to show that $\mu(F) = 0$.

Towards this end, we first observe that $T^{-m}F \cap T^{-n}F = \emptyset$ when $n > m$, say. If this were not the case and $w \in T^{-m}F \cap T^{-n}F$ then $T^m w \in F$ and $T^{n-m}(T^m w) \in F \subset A$, which contradicts the definition of F.

Thus since the sets $\{T^{-n}F\}_{n\geq 0}$ are disjoint we see that

$$\sum_{n=0}^{\infty}\mu(T^{-n}F) = \mu(\cup_{n=0}^{\infty}T^{-n}F) \leq \mu(X) = 1$$

and then because μ is T-invariant $\mu(F) = \mu(T^{-1}F) = \ldots = \mu(T^{-n}F) = \ldots$ so we can only have that $\mu(F) = 0$.

■

DEFINITION. Let $n_A : A \to \mathbb{Z}^+ \cup \{+\infty\}$ be the first return time i.e. $n_A(x) > 0$ is the smallest value for which $T^{n_A(x)}x \in A$.

By Theorem 9.2 $n_A(x)$ is finite almost everywhere. The next theorem shows that when μ is an ergodic measure then the *average* return time to A can be calculated explicitly.

THEOREM 9.3 (KAC'S THEOREM). *Let $T : X \to X$ be an ergodic transformation on a probability space $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}$ have $\mu(A) > 0$ then we define the return time function $n_A : A \to \mathbb{Z}^+ \cup \{\infty\}$ (which is finite, almost everywhere). The average return time (with respect to the induced probability measure μ_A) is*

$$\int_A n_A(x) d\mu_A(x) = \frac{1}{\mu(A)}.$$

PROOF. By definition of μ_A it is equivalent to show that $\int_A n_A(x)d\mu(x) = 1$. It is useful to define the following sets.

(a) For each $n \geq 1$ we define $A_n = \{x \in A : n_A(x) = n\}$, and write $A = \cup_{n\geq 1}A_n$ (with $A_i \cap A_j = \emptyset$ for $i \neq j$). In particular, $\sum_{n=1}^{\infty}\mu(A_n) = \mu(A)$.

(b) For $n \geq 1$ we define $B_n = \{x \in X : T^jx \notin A \text{ for } 1 \leq j \leq n-1, T^nx \in A\}$. The sets B_n are disjoint (i.e. $B_i \cap B_j = \emptyset$ for $i \neq j$) and by ergodicity $X = \cup_{n\geq 1}B_n$ (since $\cup_{n\geq 1}B_n \supset \cup_{n\geq 1}T^{-n}A \supset A$) so that $\sum_{n=1}^{\infty}\mu(B_n) = 1$.

We can rewrite

$$\int_A n_A(x)d\mu(x) = \sum_{k=1}^{\infty}k\mu(A_k) = \sum_{k=1}^{\infty}\left(\sum_{n=k}^{\infty}\mu(A_n)\right);$$

then if we can show that $\sum_{n=k}^{\infty}\mu(A_n) = \mu(B_k)$ this will complete the proof.

When $k = 1$ we have from the definitions that $B_1 = T^{-1}A$ and so $\sum_{n=1}^{\infty}\mu(A_n) = \mu(A) = \mu(B_1)$, as required. For $k > 1$ we can proceed by induction. We can partition $T^{-1}B_k = B_{k+1} \cup T^{-1}A_k$ (where $T^{-1}A_k = T^{-1}B_k \cap T^{-1}A$ and $B_{k+1} = T^{-1}B_k \cap (X - T^{-1}A)$). Thus $\mu(T^{-1}B_k) = \mu(B_k) = \mu(B_{k+1}) + \mu(T^{-1}A_k) = \mu(B_{k+1}) + \mu(A_k)$ and using the inductive hypothesis we see that $\mu(B_{k+1}) = \mu(B_k) - \mu(A_k) = \sum_{n=k+1}^{\infty}\mu(A_n)$. This completes the inductive step and the proof.

■

9.3 Existence of ergodic measures

When $T : X \to X$ is a continuous map on a compact metric space there is a very simple relationship between ergodic measures and invariant measures which we can now describe.

Let $\mathcal{M}$ denote the set of invariant probability measures on X. There is a natural topology on this space called the *weak-star* topology, i.e. the weakest topology such that a sequence $\mu_n \in \mathcal{M}$ converges to $\mu \in \mathcal{M}$ iff $\forall f \in C^0(X)$, $\int f d\mu_n \to \int f d\mu$.

The following properties of $\mathcal{M}$ are well-known (and easily checked):

(i) $\mathcal{M}$ is convex (i.e. if $\mu_1, \mu_2 \in \mathcal{M}$ and $0 < \alpha < 1$, then $\alpha\mu_1 + (1-\alpha)\mu_2 \in \mathcal{M}$);

(ii) the set $\mathcal{M}$ is compact (in the weak-star topology) [5, Theorem 6.10].

LEMMA 9.4. *The extremal points in the convex set $\mathcal{M}$ are ergodic measures (i.e. $\mu \in \mathcal{M}$ is ergodic if whenever $\exists \mu_1, \mu_2 \in \mathcal{M}$ and $0 < \alpha < 1$ with $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ then $\mu_1 = \mu_2$).*

The converse is also true, but we shall not require it.

PROOF. If μ is not ergodic then we can find $B \in \mathcal{B}$ with $T^{-1}B = B$ and $0 < \mu(B) < 1$. But for any set $A \in \mathcal{B}$ we can write $A = (A \cap B) \cup (A \cap (X - B))$ and thus

$$\begin{aligned} \mu(A) &= \mu\left((A \cap B) \cup (A \cap (X - B))\right) \\ &= \mu(B)\left(\frac{\mu(A \cap B)}{\mu(B)}\right) + \mu(X - B)\left(\frac{\mu(A \cap (X - B))}{\mu(X - B)}\right) \\ &= \alpha\mu_1(A) + (1 - \alpha)\mu_2(A) \end{aligned}$$

where $\alpha = \mu(B)$ and $\mu_1(A) = \frac{\mu(A \cap B)}{\mu(B)}$, $\mu_2(A) = \frac{\mu(A \cap (X - B))}{\mu(X - B)}$. This shows that $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$.

■

PROPOSITION 9.5 (EXISTENCE OF ERGODIC MEASURES). *Let X be a compact metric space and $\mathcal{B}$ be the Borel sigma-algebra. Given any continuous map $T : X \to X$ there exists at least one T-ergodic probability measure μ.*

PROOF. Choose a dense set of functions $f_k \in C^0(X)$, $k \geq 0$. Since the map $\mu \to \int f_0 d\mu$ is continuous on $\mathcal{M}$ there exists by (weak-star) compactness at least one $\nu \in \mathcal{M}$ such that $\int f_0 d\nu = \sup_{\mu \in \mathcal{M}}\{\int f_0 d\mu\}$. We let

$$\mathcal{M}_0 = \left\{\nu \in \mathcal{M} : \int f_0 d\nu = \sup_{\mu \in \mathcal{M}}\{\int f_0 d\mu\}\right\};$$

then clearly $\mathcal{M}_0$ is non-empty and closed. Similarly, define

$$\mathcal{M}_1 = \left\{\nu \in \mathcal{M}_0 : \int f_1 d\nu = \sup_{\mu \in \mathcal{M}_0}\{\int f_1 d\mu\}\right\}$$

and the same reasoning shows that $\mathcal{M}_1 \subset \mathcal{M}_0 \subset \mathcal{M}$ is non-empty and closed.

Proceeding inductively we define

$$\mathcal{M}_k = \left\{ \nu \in \mathcal{M}_{k-1} : \int f_k d\nu = \sup_{\mu \in \mathcal{M}_{k-1}} \{ \int f_k d\mu \} \right\}$$

and arrive at a nested sequence $\mathcal{M} \supset \mathcal{M}_0 \supset \mathcal{M}_1 \supset \mathcal{M}_2 \supset \ldots \supset \mathcal{M}_k \supset \ldots$. Since the sets are all closed in $\mathcal{M}$ (and hence compact) we have that the intersection is non-empty. Assume $\mu \in \cap_{k \in \mathbb{Z}^+} \mathcal{M}_k$. We want to show that μ is ergodic by showing that it is an extreme point in $\mathcal{M}$.

Assume that μ can be written as an affine combination $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ (with $0 < \alpha < 1$); then to show that μ is ergodic we need to show that $\mu_1 = \mu_2$. Thus it suffices to show that for every $f_k \in C^0(X)$ we have that $\int f_k f d\mu_1 = \int f_k f d\mu_2$ (since the set f_k is dense).

We begin with $k = 0$ and observe that by assumption $\int f_0 d\mu = \alpha \int f_0 d\mu_1 + (1-\alpha) \int f_0 d\mu_2$. Since $\mu \in \mathcal{M}_0$ we see that $\sup_{m \in \mathcal{M}} \{\int f_0 dm\} = \int f_0 d\mu$ implies that $\int f_0 d\mu_1 = \int f_0 d\mu_2 = \sup_{m \in \mathcal{M}} \{\int f_0 dm\}$. We thus conclude

(1) the first identity $\int f_0 d\mu_1 = \int f_0 d\mu_2$ is proved.
(2) $\mu_1, \mu_2 \in \mathcal{M}_0$.

Continuing inductively, we establish that for arbitrary $k \geq 0$ we have $\int f_k d\mu_1 = \int f_k d\mu_2$ and $\mu_1, \mu_2 \in \mathcal{M}_k$. This completes the proof (i.e. $\mu_1 = \mu_2$ and μ is an extremal measure).

■

REMARK. The following facts are easy to check.

(3) If ν, μ are distinct T-ergodic measures then $\nu \perp \mu$.
(4) If μ is ergodic then it is an extremal measure in $\mathcal{M}$. (The converse to Lemma 9.4.)

Since $\mathcal{M}$ is a compact convex metric space there is a general theorem of Choquet that says *every invariant measure $\mu \in \mathcal{M}$ can be written as a convex combination of extremal measures in $\mathcal{M}$.* More precisely, we can find a measure $\rho = \rho_\mu$ on the space $\mathcal{M}$ (with respect to the Borel sigma-algebra associated to the weak-star topology) such that

(1) for any function $f \in C^0(X)$ we have

$$\int f d\mu = \int_{\mathcal{M}} \left(\int f d\nu \right) d\rho(\nu).$$

(2) $\rho(\{\nu : \nu \text{ is extremal}\}) = 1$.

9.4 Some basic constructions in ergodic theory

In this final section of chapter 9 we shall describe two basic constructions in ergodic theory.

9.4.1. Skew products. Let $T : X \to X$ be a measure preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Let $(G, \mathcal{B})$ be a compact Lie group with the Borel sigma-algebra $\mathcal{B}$. We can consider the product space $X \times G$ with the product sigma-algebra $\mathcal{A}$.

DEFINITION. Given a measure preserving transformation of $T : X \to X$ and a measurable map $\phi : X \to G$ we define a *skew product* to be the transformation $S : X \times G \to X \times G$ defined by $S(x, g) = (Tx, \phi(x)g)$. Given any T-invariant probability measure μ we can associate the S-invariant measure ν defined by $d\nu = d\mu \times dt$.

A simple example is the following.

EXAMPLE. Let $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be given by $T(x) = x + \alpha \pmod 1$ for some $\alpha \in \mathbb{R}$. Let $G = \mathbb{R}/\mathbb{Z}$ and we define $\phi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by $\phi(x) = x \pmod 1$ (i.e. the identity map). The associated skew product is then the map $S : \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ given by $S(x, y) = (x + \alpha, x + y) \pmod 1$.

9.4.2. Induced transformations and Rohlin towers. Assume that $T : X \to X$ is a measurable transformation on a measurable space $(X, \mathcal{B})$. Assume that $A \subset X$ with $A \in \mathcal{B}$.

DEFINITION. The transformation $T_A : A \to A$ defined by $T_A(x) = T^{n_A(x)}x$ is called the *induced transformation* on A. We denote by $\mathcal{B}_A = \{B \cap A : B \in \mathcal{B}\}$ the restriction of the sigma-algebra $\mathcal{B}$ to A.

If μ is a T-invariant sigma-finite measure on $(X, \mathcal{B})$ and $0 < \mu(A) < \infty$ then we can define a T_A-invariant measure μ_A on $(A, \mathcal{B}_A)$ by $\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}$.

EXAMPLE (CONTINUED FRACTION TRANSFORMATION). Consider the case where $(X, \mathcal{B})$ is the positive half-line $\mathbb{R}^+ = (0, +\infty)$ with the Borel sigma-algebra. We define a transformation $T : \mathbb{R}^+ \to \mathbb{R}^+$ by

(1) $Tx = x - 1$ if $x \in [1, +\infty)$, and
(2) $Tx = \frac{1}{x}$ if $x \in (0, 1)$.

We can consider the induced transformation $T_A : A \to A$ on the interval $A = (0, 1]$ defined by $T_A x = \frac{1}{x} - \left[\frac{1}{x}\right]$.

The measure μ_A defined by $\mu_A(B) = \frac{1}{\log 2} \int_B \frac{1}{1+x} dx$ is T_A-invariant.

REMARK. We need not be too careful about the definition of T and T_A on a countable set of points since they have zero measure.

Consider an (ergodic) transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ and let $A \in \mathcal{B}$ have $\mu(A) > 0$.

DEFINITION. We can define a space

$$A^{n_A} = \{(x,k) \in A \times \mathbb{Z}^+ : 0 \le k \le n_A(x)\},$$

where we identify $(x, n_A(x)) \sim (T^{n_A(x)}x, 0)$, and introduce the product sigma-algebra $\bar{\mathcal{B}}$ (i.e. the smallest sigma-algebra containing the products of sets in $\mathcal{B}_A$ and $\mathcal{B}_{\mathbb{Z}^+}$).

We define a probability measure on the space A^{n_A} by $\nu = \frac{\mu \times dn}{\int n_A d\mu}$ (where dn corresponds to the usual counting measure on $\mathbb{Z}^+$).

Finally, we define a transformation $T_A^{n_A} : A^{n_A} \to A^{n_A}$ by

(1) $T_A^{n_A}(x,k) = (x, k+1)$ if $0 \le k < n_A(x)$, and
(2) $T_A^{n_A}(x, n_A(x)) = T_A^{n_A}(T_A x, 0)) = (T_A x, 1)$.

This construction is called the *Rohlin tower* over A.

(N.B. A Rohlin tower is the converse process to induced transformations. We reproduce the original transformation on X from the induced transformation on A.) The following lemma tells us the Rohlin tower is a good model of the original transformation.

LEMMA 9.6. *The map $\phi : (A^{n_A}, \bar{\mathcal{B}}, \nu) \to (X, \mathcal{B}, \mu)$ defined by $\phi(x,k) = T^k(x)$ is measurable and satisfies the following:*

(1) *ϕ is a bijection (almost everywhere);*
(2) *$\forall B \in \mathcal{B}$ we have that $\nu(\phi^{-1}B) = \mu(B)$; and*
(3) *$\phi T_A^{n_A} = T\phi$ (almost everywhere).*

PROOF. The result follows almost immediately from the definitions. ∎

REMARK. The map ϕ is an *isomorphism* which implies that from the point of view of ergodic theory the transformations T and $T_A^{n_A}$ are the same.

9.4.3 Natural extensions. Given a non-invertible map $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ there is a natural way of associating to it an invertible transformation $\hat{T} : (\hat{X}, \hat{\mathcal{B}}, \hat{\mu}) \to (\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ with similar dynamical properties.

We define

$$\hat{X} = \{(x_n)_{n\in\mathbb{Z}^+} \in \prod_{n\in\mathbb{Z}^+} X : T(x_n) = x_{n+1}, n \ge 0\}$$

and associate the sigma-algebra generated by the sets

$$B_m := \left\{(x_n)_{n\in\mathbb{Z}^+} \in \hat{X} : x_m \in B\right\} \text{ for } B \in \mathcal{B} \text{ and } m \in \mathbb{Z}^+.$$

We next define a probability measure $\hat{\mu}$ on $\hat{\mathcal{B}}$ by $\hat{\mu}(B_m) = \mu(B)$. Finally, we define the (invertible) transformation $\hat{T} : \hat{X} \to \hat{X}$ by

$$\hat{T}(x_0, x_1, x_2, \dots) = (Tx_0, x_0, x_1, x_2, \dots).$$

It is easy to see from the construction that $\hat{T}$ is measurable and preserves the probability measure $\hat{\mu}$.

DEFINITION. We call $\hat{T} : \hat{X} \to \hat{X}$ the *natural extension* of T.

There is a canonical map $\pi : \hat{X} \to X$ defined by $\pi\left((x_n)_{n \in \mathbb{Z}^+}\right) = x_0$. The natural extension $\hat{T}$ has the following properties:

(i) $\hat{T}$ is an extension of T in the sense that $\pi \circ \hat{T} = T \circ \pi$; and

(ii) if we denote by $\hat{\mathcal{B}}^+ \subset \hat{\mathcal{B}}$ the sub-sigma-algebra generated by sets $\{\pi^{-1}(B) : B \in \mathcal{B}\}$ then

$$\ldots \subset \hat{T}^{-1}\hat{\mathcal{B}}^+ \subset \hat{\mathcal{B}}^+ \subset T\hat{\mathcal{B}}^+ \subset \ldots \subset \cup_{n \in \mathbb{Z}^+} T^n \hat{\mathcal{B}}^+ = \mathcal{B}.$$

REMARK. In fact, any transformation satisfying (i) and (ii) will be isomorphic to the natural extension as we have defined it above [3].

EXAMPLE (SUBSHIFTS OF FINITE TYPE). Let $\sigma : X_A^+ \to X_A^+$ be a (one-sided) subshift of finite type, defined by the $k \times k$ matrix A. Relative to a Markov measure, say, its natural extension is the shift $\sigma : X \to X$.

9.5 Comments and references

More can be found on ergodic measures in [1], [2] and [5].

Important applications of ergodic theory beyond the scope of these notes are Mostow's rigidity theorem [4] and the Margulis super-rigidity theorem [6, §5.1].

The skew product example in subsection 9.4.1 was used by Furstenburg to give a simple proof of a result on diophantine approximation due to Hardy and Littlewood [2].

References

1. P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965.
2. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.
3. V. Rohlin, *Exact endomorphisms of lebesgue space*, Amer. Math. Soc. Transl. (2) **39** (1964), 1-36.
4. W. Thurston, *Topology and geometry of three manifolds*, unpublished notes, Princeton University N.J., Princeton, 1978.
5. P. Walters, *An Introduction to Ergodic Theory*, Springer, New York, 1989.
6. R. Zimmer, *Ergodic Theory and Semi-simple Groups*, Birkhäuser, Basel, 1984.