Graduate Texts in Mathematics

79

An Introduction to Ergodic Theory

Springer-Verlag

Peter Walters

An Introduction to Ergodic Theory

With 8 Illustrations

Springer-Verlag
New York Heidelberg Berlin

Peter Walters
Mathematics Institute
University of Warwick
Coventry CV4 7AL
England

AMS Subject Classification: 28-01, 28DXX, 47A35, 54H20

Library of Congress Cataloging in Publication Data

Walters, Peter, 1943–
An introduction to ergodic theory.

(Graduate texts in mathematics; 79)
Previously published as: Ergodic theory. 1975.
Bibliography: p.
Includes index.
1. Ergodic theory. I. Title. II. Series.
QA313.W34 1981 515.4'2 81-9319
ISBN 0-387-90599-5 AACR2

Reprinted in China by Beijing World Publishing Corporation, 2001

ISBN 0-387-90599-5 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90599-5 Springer-Verlag Berlin Heidelberg New York

Preface

In 1970 I gave a graduate course in ergodic theory at the University of Maryland in College Park, and these lectures were the basis of the Springer Lecture Notes in Mathematics Volume 458 called "Ergodic Theory—Introductory Lectures" which was published in 1975. This volume is now out of print, so I decided to revise and add to the contents of these notes. I have updated the earlier chapters and have added some new chapters on the ergodic theory of continuous transformations of compact metric spaces. In particular, I have included some material on topological pressure and equilibrium states. In recent years there have been some fascinating interactions of ergodic theory with differentiable dynamics, differential geometry, number theory, von Neumann algebras, probability theory, statistical mechanics, and other topics. In Chapter 10 I have briefly described some of these and given references to some of the others. I hope that this book will give the reader enough foundation to tackle the research papers on ergodic theory and its applications.

I would like to dedicate this volume to the memory of Rufus Bowen who died on July 30, 1978 at the age of 31. He made outstanding contributions to ergodic theory and his friendship enhanced the lives of all who knew him.

April, 1981 PETER WALTERS

Contents

CHAPTER 0
Preliminaries

§0.1 Introduction

In its broadest interpretation ergodic theory is the study of the qualitative properties of actions of groups on spaces. The space has some structure (e.g. the space is a measure space, or a topological space, or a smooth manifold) and each element of the group acts as a transformation on the space and preserves the given structure (e.g. each element acts as a measure-preserving transformation, or a continuous transformation, or a smooth transformation).

To see how this type of study arises consider a system of k particles moving in R^3 under known forces. Suppose that the state of the system at a given time is determined by knowing the positions and the momenta of each of the k particles. Thus at a given time the system is determined by a point in R^{6k}. As time continues the system alters according to the differential equations governing the motion, e.g., Hamilton's equations

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}.$$

If we are given an initial condition and if the equations can be uniquely solved then the corresponding solution gives us the entire history of the system, which is determined by a curve in R^{6k}.

If x is a point in the state space representing the system at a time t_0, let $T_t(x)$ denote the point of the state space representing the system at time $t + t_0$. From this we see that T_t is a transformation of the state space and, moreover, $T_0 = id$ and $T_{t+s} = T_t \circ T_s$. Thus $t \to T_t$ is an action of the group R on the state space. Because the Hamiltonian H is constant along solution curves, each energy surface $H^{-1}(e)$ is invariant for the transformation T_t

so that we get an action of R on each energy surface. One is interested in the asymptotic properties of the action i.e. in T_t for large t. The transformations $T_t|H^{-1}(e)$ are continuous and are smooth if $H^{-1}(e)$ is smooth. Measure theory enters this picture via a theorem of Liouville which tells us that if the forces are of a certain type one can choose coordinates in the state space so that the usual $6k$-dimensional measure in these coordinates is preserved by each transformation T_t.

The word "ergodic" was introduced by Boltzmann to describe a hypothesis about the action of $\{T_t|t \in R\}$ on an energy surface $H^{-1}(e)$ when the Hamiltonian H is of the type that arises in statistical mechanics. Boltzmann had hoped that each orbit $\{T_t(x)|t \in R\}$ would equal the whole energy surface $H^{-1}(e)$ and he called this statement the ergodic hypothesis. The word "ergodic" is an amalgamation of the Greek words ergon (work) and odos (path). Boltzmann made the hypothesis in order to deduce the equality of time means and phase means which is a fundamental algorithm in statistical mechanics. The ergodic hypothesis, as stated above, is false. The property the flow needs to satisfy in order to equate time means and phase means of real-valued functions is what is now called ergodicity.

It is common to use the name ergodic theory to describe only the qualitative study of actions of groups on measure spaces. The actions on topological spaces and smooth manifolds are often called topological dynamics and differentiable dynamics. This measure theoretic study began in the early 1930's and the ergodic theorems of Birkhoff and von Neumann were proved then. The next major advance was the introduction of entropy by Kolmogorov in 1958. The proof, by Ornstein in 1969, that entropy was complete for Bernoulli shifts revitalised the work on the isomorphism problem. During recent years ergodic theory had been used to give important results in other branches of mathematics.

We shall study actions of the group Z of integers on a space X i.e. we study a transformation $T: X \to X$ and its iterates T^n, $n \in Z$. This is simpler than studying the actions of R. Of course, if $\{T_t|t \in R\}$ is an action of R on X, then by choosing $t_0 \neq 0$ and observing the system at the times $\ldots, -t_0, 0, t_0, 2t_0, 3t_0, \ldots$, we are considering $(T_{t_0})^n$, $n \in Z$.

In the following sections we summarise some of the background ideas and notation we shall be using.

We shall use Z to denote the set of integers, Z^+ to denote the non-negative integers, R to denote the real numbers, R^+ to denote the non-negative reals, and $\mathbb{C}$ to denote the complex numbers. The empty set will be denoted by $\varnothing$.

If A,B are subsets of a set X, then $B\backslash A$ denotes the difference set $\{x \in X | x \in B, x \notin A\}$, and $A \triangle B$ denotes the symmetric difference $(A\backslash B) \cup (B\backslash A)$. We use 2^X to denote the collection of all subsets of X.

We use "iff" to denote "if and only if." We number lemmas and theorems in a single sequence (Theorem 5.6 is the sixth theorem in Chapter 5) but give a corollary the same number as the corresponding theorem (Corollary 5.6.2 is the second corollary of Theorem 5.6).

§0.2 Measure Spaces

We shall generally refer to Kingman and Taylor [1] and Parthasarathy [2]. Let X be a set. A *σ-algebra* of subsets of X is a collection $\mathscr{B}$ of subsets of X satisfying the following three conditions: (i) $X \in \mathscr{B}$; (ii) if $B \in \mathscr{B}$ then $X \backslash B \in \mathscr{B}$; (iii) if $B_n \in \mathscr{B}$ for $n \geq 1$ then $\bigcup_{n=1}^{\infty} B_n \in \mathscr{B}$.

We then call the pair $(X, \mathscr{B})$ a *measurable space*. A *finite measure* on $(X, \mathscr{B})$ is a function $m: \mathscr{B} \to R^+$ satisfying $m(\varnothing) = 0$ and $m(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} m(B_n)$ whenever $\{B_n\}_1^{\infty}$ is a sequence of members of $\mathscr{B}$ which are pairwise disjoint subsets of X. (Actually the latter condition implies $m(\varnothing) = 0$ since m is finite-valued.) A finite measure space is a triple $(X, \mathscr{B}, m)$ where $(X, \mathscr{B})$ is a measurable space and m is a finite measure on $(X, \mathscr{B})$. We say $(X, \mathscr{B}, m)$ is a *probability space*, or a normalised measure space, if $m(X) = 1$. We then say m is a probability measure on $(X, \mathscr{B})$. We shall usually consider only probability spaces.

A *finite signed measure* on a measurable space $(X, \mathscr{B})$ is a function $m: \mathscr{B} \to R$ satisfying $m(\varnothing) = 0$ and $m(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} m(B_n)$ whenever $\{B_n\}_1^{\infty}$ is a sequence of members of $\mathscr{B}$ which are pairwise disjoint subsets of X. The *Jordan decomposition* says that a finite signed measure m on $(X, \mathscr{B})$ can be written as the difference $m = m_1 - m_2$ of two finite measures on $(X, \mathscr{B})$ which are uniquely determined by m (see Kingman and Taylor [1], pp. 62 and 64).

Measurable spaces are usually constructed by having a collection $\mathscr{S}$ of interesting subsets of a set X (such as the collection of all subintervals of $[0, 1]$) and then considering the smallest σ-algebra $\mathscr{B}$ containing all these subsets. This makes sense because 2^X is a σ-algebra and any intersection of σ-algebras of subsets of X is also a σ-algebra of subsets of X. It is then usually difficult to decide which subsets of X are in $\mathscr{B}$. When constructing a measure on a measurable space $(X, \mathscr{B})$ obtained in this way, one usually knows what values the measure should take on members of $\mathscr{S}$ and then one needs to extend it to be defined on $\mathscr{B}$. We now describe the basic extension theorem of this type. This involves discussion of the properties the collection $\mathscr{S}$ should have.

A collection $\mathscr{S}$ of subsets of X is called a *semi-algebra* if the following three conditions hold: (i) $\varnothing \in \mathscr{S}$; (ii) if $A, B \in \mathscr{S}$, then $A \cap B \in \mathscr{S}$; (iii) if $A \in \mathscr{S}$, then $X \backslash A = \bigcup_{i=1}^{n} E_i$ where each $E_i \in \mathscr{S}$ and $E_1, \ldots, E_n$ are pairwise disjoint subsets of X. For example, the collection of all subintervals of $[0, 1]$ is a semi-algebra. Also, the collection of all subintervals of $[0, 1]$ of the forms $[0, b]$ and $(a, b]$, with $0 \leq a < b \leq 1$, forms a semi-algebra.

A collection $\mathscr{A}$ of subsets of X is called an *algebra* if the following three conditions hold: (i) $\varnothing \in \mathscr{A}$; (ii) if $A, B \in \mathscr{A}$, then $A \cap B \in \mathscr{A}$; (iii) if $A \in \mathscr{A}$, then $X \backslash A \in \mathscr{A}$. Clearly every algebra is a semi-algebra and every σ-algebra is an algebra. In the definition of an algebra we can replace (ii) by the condition that whenever $A_1, \ldots, A_n \in \mathscr{A}$ then $\bigcup_{i=1}^{n} A_i \in \mathscr{A}$.

Since the intersection of any family of algebras of a set X is again an algebra of subsets of X it makes sense to speak of the algebra generated by any given collection of subsets of X. There is the following simple theorem (Parthasarathy [2], p. 19)

Theorem 0.1. *Let $\mathscr{S}$ be a semi-algebra of subsets of X. The algebra, $\mathscr{A}(\mathscr{S})$, generated by $\mathscr{S}$ consists precisely of those subsets of X that can be written in the form $E = \bigcup_{i=1}^{n} A_i$ where each $A_i \in \mathscr{S}$ and $A_1, \ldots, A_n$ are disjoint subsets of X.*

Suppose $\mathscr{S}$ is a semi-algebra of subsets of X. A function $\tau:\mathscr{S} \to R^+$ is called *finitely additive* if $\tau(\varnothing) = 0$ and $\tau(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \tau(E_i)$ whenever $E_1, \ldots, E_n$ are members of $\mathscr{S}$ which are pairwise disjoint subsets of X and $\bigcup_{i=1}^{n} E_i \in \mathscr{S}$. Such a map τ is called *countably additive* if the second condition is replaced by the requirement that $\tau(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \tau(E_i)$ whenever $\{E_i\}_1^{\infty}$ are members of $\mathscr{S}$ which are pairwise disjoint subsets of X and $\bigcup_{i=1}^{\infty} E_i \in \mathscr{S}$. If $\mathscr{S}$ is the semi-algebra of all subintervals of $[0, 1]$ of the form $[0, b]$ and $(a, b]$ then the length function is countably additive. A simple application of Theorem 0.1 gives the following (Parthasarathy [2], pp. 20 and 59).

Theorem 0.2. *If $\mathscr{S}$ is a semi-algebra of subsets of X and $\tau:\mathscr{S} \to R^+$ is finitely additive then there is a unique finitely additive function $\tau_1:\mathscr{A}(\mathscr{S}) \to R^+$ which is an extension of τ (i.e. $\tau_1 = \tau$ on $\mathscr{S}$). If τ is countably additive then so is τ_1.*

It could be that $X \notin \mathscr{S}$ but we always have that X is a disjoint union of a finite number of members $E_1, \ldots, E_n$ of $\mathscr{S}$ so $\tau_1(X) = 1$ if $\sum_{i=1}^{n} \tau(E_i) = 1$.

There is the following theorem on extension from an algebra $\mathscr{A}$ to the σ-algebra $\mathscr{B}(A)$ generated by $\mathscr{A}$. ($\mathscr{B}(A)$ is the intersection of all σ-algebras that contain $\mathscr{A}$.) (See Parthasarathy [2], pp. 70 and 71).

Theorem 0.3. *Let $\mathscr{A}$ be an algebra of subsets of X and let $\tau_1:\mathscr{A} \to R^+$ be countably additive and $\tau_1(X) = 1$. Then there is a unique probability measure τ_2 on $(X, \mathscr{B}(\mathscr{A}))$ which extends τ_1.*

By combining Theorems 0.2 and 0.3 we see that a countably additive function τ on a semi-algebra $\mathscr{S}$ can be uniquely extended to a probability measure on $(X, \mathscr{B}(\mathscr{S}))$ if $\sum_{i=1}^{n} \tau(E_i) = 1$ when $X = \bigcup_{i=1}^{n} E_i$ is a disjoint union of members of $\mathscr{S}$. As an example the length function defined on the semi-algebra of all subintervals of $[0, 1]$ of the form $[0, b]$ and $(a, b]$ can be uniquely extended to a probability measure, called the Lebesgue measure, defined on the Borel subsets of $[0, 1]$.

In checking that the extension works for particular examples the most difficult part is usually showing countable additivity. This can sometimes be done for τ on the semi-algebra $\mathscr{S}$ (as for Lebesgue measure) but it is sometimes more convenient to prove that τ is finitely additive and that the finitely additive extension $\tau_1:\mathscr{A}(\mathscr{S}) \to R^+$ is countably additive. The main tool for this is the following theorem (Kingman and Taylor [1], p. 56).

Theorem 0.4. *Let $\mathscr{A}$ be an algebra of subsets of X and let $\tau_1 : \mathscr{A} \to R^+$ be finitely additive and let $\tau_1(X) = 1$. Then τ_1 will be countably additive if for every decreasing sequence $E_1 \supset E_2 \supset E_3 \supset \cdots$ of members of $\mathscr{A}$ with $\bigcap_{n=1}^{\infty} E_n = \varnothing$ we have $\tau_1(E_n) \to 0$.*

One situation where this theorem is used is in defining product measures on a countable product of probability spaces. For $i \in Z$ let $(X_i, \mathscr{B}_i, m_i)$ be a probability space. Let $X = \prod_{i=-\infty}^{\infty} X_i$. So a point of X is a bisequence $\{x_i\}_{-\infty}^{\infty}$ with $x_i \in X_i$ for each i. We now define a σ-algebra $\mathscr{B}$ of subsets of X called the product of the σ-algebras $\mathscr{B}_i$. Let $n \geq 0$, let $A_j \in \mathscr{B}_j$ for $|j| \leq n$, and consider the set

$$\prod_{i=-\infty}^{-(n+1)} X_i \times \prod_{j=-n}^{n} A_j \times \prod_{i=n+1}^{\infty} X_i = \{(x_i)_{-\infty}^{\infty} \in X \,|\, x_j \in A_j \text{ for } |j| \leq n\}.$$

Such a set is called a *measurable rectangle* and the collection of all such subsets of X forms a semi-algebra $\mathscr{S}$. The σ-algebra $\mathscr{B}$ is the σ-algebra generated by $\mathscr{S}$. We write $(X, \mathscr{B}) = \prod_{i=-\infty}^{\infty} (X_i, \mathscr{B}_i)$. If we define $\tau : \mathscr{S} \to R^+$ by giving the above rectangle the value $\prod_{j=-n}^{n} m_j(A_j)$, then one can use Theorems 0.2 and 0.4 (see Kingman and Tayler [1], p. 140) to extend τ to a probability measure m on $(X, \mathscr{B})$. The probability space $(X, \mathscr{B}, m)$ is called the *direct product* of the spaces $(X_i, \mathscr{B}_i, m_i)$ and is sometimes denoted $\prod_{i=-\infty}^{\infty} (X_i, \mathscr{B}_i, m_i)$. The corresponding construction holds for a product $\prod_{i=0}^{\infty} (X_i, \mathscr{B}_i, m_i)$.

A special type of product space will be important for us. Here each space $(X_i, \mathscr{B}_i, m_i)$ is the same space $(Y, \mathscr{C}, \mu)$ and Y is the finite set $\{0, 1, \ldots, k-1\}$, $\mathscr{C} = 2^Y$, and μ is given by a probability vector $(p_0, p_1, \ldots, p_{k-1})$ where $p_i = \mu(\{i\})$. We can take elementary rectangles where each A_j (in the description above) is taken to be one point of Y. So if $n \geq 0$ and $a_j \in Y$, $|j| \leq n$, such an elementary rectangle has the form $\{(x_i)_{-\infty}^{\infty} | x_j = a_j \text{ for } |j| \leq n\}$. We shall denote this set by ${}_{-n}[a_{-n}, a_{-(n+1)}, \ldots, a_{n-1}, a_n]_n$ and call it a *block* with end points $-n$ and n. The collection of all these sets form a semi-algebra which generates the product σ-algebra $\mathscr{B}$. We have $m({}_{-n}[a_n, \ldots, a_n]_n) = \prod_{j=n}^{n} p_j$. The measure m is called the $(p_0, \ldots, p_{k-1})$-product measure. Sometimes we consider *blocks* with end points h and l where $h \leq l$. Such a set is one of the form ${}_h[a_h, \ldots, a_l]_l = \{(x_i)_{-\infty}^{\infty} | x_i = a_i \text{ for } h \leq i \leq l\}$. It has measure $\prod_{i=h}^{l} p_i$.

Theorem 0.4 can also be used to obtain further measures on the space $(X, \mathscr{B})$ where $X = \prod_{-\infty}^{\infty} Y$, $Y = \{0, 1, \ldots, k-1\}$, and $\mathscr{B}$ is the product σ-algebra described above. The following is a special case of the Daniell–Kolmogorov consistency theorem (Parthasarathy [2], p. 119).

Theorem 0.5. *Fix $k \geq 1$ and let $Y = \{0, 1, \ldots, k-1\}$ and $(X, \mathscr{B}) = \prod_{-\infty}^{\infty} (Y, 2^Y)$. For each natural number n and $a_0, \ldots, a_n \in Y$ suppose a non-negative real number $p_n(a_0, \ldots, a_n)$ is given so that*

(a) $$\sum_{a_0 \in Y} p_0(a_0) = 1$$

and

(b) $$p_n(a_0, \ldots, a_n) = \sum_{a_{n+1} \in Y} p_{n+1}(a_0, \ldots, a_n, a_{n+1}).$$

Then there is a unique probability measure m on $(X, \mathscr{B})$ *with* $m({}_h[a_h, \ldots, a_l]_l) = p_{l-h}(a_h, \ldots, a_l)$ *for all* $h \leq l$ *and all* $a_i \in Y$, $h \leq i \leq l$.

The proof boils down to showing that the function naturally defined on the algebra of all finite unions of elementary rectangles is countable additive by using Theorem 0.4.

There is another way, which is useful in some proofs, of describing the σ-algebra $\mathscr{B}(\mathscr{A})$ generated by an algebra $\mathscr{A}$. A collection M of subsets of X is called a *monotone class* if whenever $E_1 \subset E_2 \subset E_3 \subset \cdots$ all belong to M then so does $\bigcup_{n=1}^{\infty} E_n$ and whenever $F_1 \supset F_2 \supset F_3 \supset \cdots$ all belong to M then so does $\bigcap_{n=1}^{\infty} F_n$. Since the intersection of any family of monotone classes is a monotone class, we can speak of the monotone class generated by any given collection of subsets of X.

Theorem 0.6. *Let* $\mathscr{A}$ *be an algebra of subsets of* X. *Then* $\mathscr{B}(\mathscr{A})$ *equals the monotone class generated by* $\mathscr{A}$.

As we have seen we usually know the elements of an algebra $\mathscr{A}$ but we do not know which subsets of X belong to $\mathscr{B}(\mathscr{A})$. This problem can sometimes be overcome by using the following approximation theorem (Kingman and Taylor [1], p. 84).

Theorem 0.7. *Let* $(X, \mathscr{B}, m)$ *be a probability space and let* $\mathscr{A}$ *be an algebra of subsets of* X *with* $\mathscr{B}(\mathscr{A}) = \mathscr{B}$. *Then for each* $\varepsilon > 0$ *and each* $B \in \mathscr{B}$ *there is some* $A \in \mathscr{A}$ *with* $m(A \triangle B) < \varepsilon$.

Note that when $m(A \triangle B) < \varepsilon$ then $|m(A) - m(B)| < \varepsilon$ because $m(A) = m(A \backslash B) + m(A \cap B)$ and $m(B) = m(B \backslash A) + m(A \cap B)$, so that $|m(A) - m(B)| \leq m(A \triangle B)$.

§0.3 Integration

Let $\mathscr{B}(R)$ denote the σ-algebra of Borel subsets of R. This is the σ-algebra generated by all open subsets of R and is also generated by the collection of all intervals, or by the collection of all intervals of the form (c, ∞).

Let $(X, \mathscr{B}, m)$ be a measure space. A function $f: X \to R$ is *measurable* if $f^{-1}(D) \in \mathscr{B}$ whenever $D \in \mathscr{B}(R)$ or equivalently if $f^{-1}(c, \infty) \in \mathscr{B}$ for all $c \in R$. A function $f: X \to \mathbb{C}$ is measurable if both its real and imaginary parts are measurable. If X is a topological space and $\mathscr{B}$ the σ-algebra generated by the open subsets of X, then any continuous function $f: X \to \mathbb{C}$ is measurable. We say $f = g$ a.e. if $m(\{x: f(x) \neq g(x)\}) = 0$. Suppose X is a topological space, $\mathscr{B}(X)$ its σ-algebra of Borel sets and m a measure on $(X, \mathscr{B}(X))$ with the property that each non-empty open set has non-zero measure. Then for two continuous functions $f, g: X \to R$, $f = g$ a.e. implies $f = g$ because $\{x: f(x) - g(x) \neq 0\}$ is an open set of zero measure.

Let $(X, \mathscr{B}, m)$ be a probability space. A function $f: X \to R$ is a *simple function* if it can be written in the form $\sum_{i=1}^{n} a_i \chi_{A_i}$, where $a_i \in R$, $A_i \in \mathscr{B}$, the sets A_i are disjoint subsets of X, and χ_{A_i} denotes the characteristic function of A_i. Simple functions are measurable. We define the integral for simple functions by:

$$\int f\, dm = \sum_{i=1}^{n} a_i m(A_i).$$

This value is independent of the representation $\sum_i a_i \chi_{A_i}$.

Suppose $f: X \to R$ is measurable and $f \geq 0$. Then there exists an increasing sequence of simple functions $f_n \nearrow f$. For example, we could take

$$f_n(x) = \begin{cases} \dfrac{i-1}{2^n}, & \text{if } \dfrac{i-1}{2^n} \leq f(x) < \dfrac{i}{2^n} \quad i = 1, \dots, n2^n \\ n, & \text{if } f(x) \geq n. \end{cases}$$

We define $\int f\, dm = \lim_{n\to\infty} \int f_n\, dm$ and note that this definition is independent of the chosen sequence $\{f_n\}$. We say f is integrable if $\int f\, dm < \infty$.

Suppose $f: X \to R$ is measurable. Then $f = f^+ - f^-$ where $f^+(x) = \max\{f(x), 0\} \geq 0$ and $f^-(x) = \max\{-f(x), 0\} \geq 0$. We say that f is *integrable* if $\int f^+\, dm$, $\int f^-\, dm < \infty$ and we then define

$$\int f\, dm = \int f^+\, dm - \int f^-\, dm.$$

We say $f: X \to \mathbb{C}$ is integrable ($f = f_1 + i f_2$) if f_1 and f_2 are integrable and we define

$$\int f\, dm = \int f_1\, dm + i \int f_2\, dm.$$

Observe that f is integrable if and only if $|f|$ is integrable. If $f = g$ a.e. then one is integrable if the other is and $\int f\, dm = \int g\, dm$.

The two basic theorems on integrating sequences as functions are the following.

Theorem 0.8 (Monotone Convergence Theorem). *Suppose $f_1 \leq f_2 \leq f_3 \leq \cdots$ is an increasing sequence of integrable real-valued functions on $(X, \mathscr{B}, m)$. If $\{\int f_n\, dm\}$ is a bounded sequence of real numbers then* $\lim_{n\to\infty} f_n$ *exists* a.e. *and is integrable and $\int (\lim f_n)\, dm = \lim \int f_n\, dm$. If $\{\int f_n\, dm\}$ is an unbounded sequence then either $\lim_{n\to\infty} f_n$ is infinite on a set of positive measure or $\lim_{n\to\infty} f_n$ is not integrable.*

Theorem 0.9 (Fatou's Lemma). *Let $\{f_n\}$ be a sequence of measurable real-valued functions on $(X, \mathscr{B}, m)$ which is bounded below by an integrable function. If $\liminf_{n\to\infty} \int f_n\, dm < \infty$ then $\liminf_{n\to\infty} f_n$ is integrable and $\int \liminf f_n\, dm \leq \liminf \int f_n\, dm$.*

Corollary 0.9.1 (Dominated Convergence Theorem). *If $g: X \to R$ is integrable and $\{f_n\}$ is a sequence of measurable real-valued functions with $|f_n| \leq g$ a.e. $(n \geq 1)$ and $\lim_{n\to\infty} f_n = f$ a.e. then f is integrable and $\lim \int f_n\, dm = \int f\, dm$.*

We denote by $L^1(X, \mathscr{B}, m)$ (or $L^1(m)$) the space of all integrable functions $f: X \to \mathbb{C}$ where two such functions are identified if they are equal a.e. However we write $f \in L^1(X, \mathscr{B}, m)$ to denote that $f: X \to \mathbb{C}$ is integrable. The space $L^1(X, \mathscr{B}, m)$ is a Banach space with norm $\|f\|_1 = \int |f|\, dm$.

If $f \in L^1(X, \mathscr{B}, m)$, then $\int_A f\, dm$ denotes $\int f \cdot \chi_A\, dm$.

If m is a finite signed measure on $(X, \mathscr{B})$ and $m = m_1 - m_2$ is its unique Jordan decomposition into the difference of two finite measures, then we can define $\int f\, dm = \int f\, dm_1 - \int f\, dm_2$ for $f \in L^1(m_1) \cap L^1(m_2)$.

§0.4 Absolutely Continuous Measures and Conditional Expectations

Let $(X, \mathscr{B})$ be a measurable space and suppose μ, m are two probability measures on $(X, \mathscr{B})$. We say μ is *absolutely continuous* with respect to $m (\mu \ll m)$ if $\mu(B) = 0$ whenever $m(B) = 0$. The measures are *equivalent* if $\mu \ll m$ and $m \ll \mu$. The following theorem characterises absolute continuity.

Theorem 0.10 (Radon–Nikodym Theorem). *Let μ, m be two probability measure on the measurable space $(X, \mathscr{B})$. Then $\mu \ll m$ iff there exists $f \in L^1(m)$, with $f \geq 0$ and $\int f\, dm = 1$, such that $\mu(B) = \int_B f\, dm\ \forall B \in \mathscr{B}$. The function f is unique* a.e. (*in the sense that any other function with these properties is equal to f* a.e.).

The function f is called the Radon–Nikodym derivative of μ with respect to m and denoted by $d\mu/dm$.

The "opposite" notion to absolute continuity is as follows. Two probability measures μ, m on $(X, \mathscr{B})$ are said to be *mutually singular* $(\mu \perp m)$ if there is some $B \in \mathscr{B}$ with $\mu(B) = 0$ and $m(X \backslash B) = 0$. There is the following decomposition theorem.

Theorem 0.11 (Lebesgue Decomposition Theorem). *Let μ, m be two probability measures on $(X, \mathscr{B})$. There exists $p \in [0, 1]$ and probability measures μ_1, μ_2 on $(X, \mathscr{B})$ such that $\mu = p\mu_1 + (1 - p)\mu_2$ and $\mu_1 \ll m$, $\mu_2 \perp m$. ($\mu = p\mu_1 + (1 - p)\mu_2$ means $\mu(B) = p\mu_1(B) + (1 - p)\mu_2(B)\ \forall B \in \mathscr{B}$). The number p and probabilities μ_1, μ_2 are uniquely determined.*

The Radon–Nikodym theorem allows us to define conditional expectations. Let $(X, \mathscr{B}, m)$ be a measure space and let $\mathscr{C}$ be a sub-σ-algebra of $\mathscr{B}$. We now define the conditional expectation operator $E(\cdot/\mathscr{C})\colon L^1(X, \mathscr{B}, m) \to L^1(X, \mathscr{C}, m)$. If $f \in L^1(X, \mathscr{B}, m)$ takes non-negative real values then $\mu_f(C) = a^{-1} \int_C f\, dm$ (where $a = \int_X f\, dm$) defines a probability measue, μ_f, on $(X, \mathscr{C}, m)$ and $\mu_f \ll m$. By Theorem 0.10 there is a function $E(f/\mathscr{C}) \in L^1(X, \mathscr{C}, m)$ such that $E(f/\mathscr{C}) \geq 0$ and $\int_C E(f/\mathscr{C})\, dm = \int_C f\, dm\ \forall C \in \mathscr{C}$. Moreover $E(f/\mathscr{C})$ is unique a.e. If f is real-valued we can consider the positive and negative parts of f and define $E(f/\mathscr{C})$ linearly. Similarly when f is complex-valued we can use the real and imaginary parts to define $E(f/\mathscr{C})$ linearly. Therefore if $f \in L^1(X, \mathscr{B}, m)$ then $E(f/\mathscr{C})$ is the only $\mathscr{C}$-measurable function with $\int_C E(f/\mathscr{C})\, dm = \int_C f\, dm\ \forall C \in \mathscr{C}$. The following properties of the map $E(\cdot/\mathscr{C})\colon L^1(X, \mathscr{B}, m) \to L^1(X, \mathscr{C}, m)$ hold (Parthasarathy [2], p. 225):

(i) $E(\cdot/\mathscr{C})$ is linear.
(ii) If $f \geq 0$, then $E(f/\mathscr{C}) \geq 0$.
(iii) If $f \in L^1(X, \mathscr{B}, m)$ and g is $\mathscr{C}$-measurable and bounded,

$$E(fg/\mathscr{C}) = gE(f/\mathscr{C}).$$

(iv) $|E(f/\mathscr{C})| \leq E(|f|/\mathscr{C}), \qquad f \in L^1(X, \mathscr{B}, m)$
(v) If $\mathscr{C}_2 \subset \mathscr{C}_1$, then $E(E(f/\mathscr{C}_1)/\mathscr{C}_2) = E(f/\mathscr{C}_2)$, $f \in L^1(X, \mathscr{B}, m)$.

§0.5 Function Spaces

One way to deal with some problems on a measure space is to use certain natural Banach spaces of functions associated with the measure space.

Let $(X, \mathscr{B}, m)$ be a measure space and let $p \in R$ with $p \geq 1$. Consider the set of all measurable functions $f\colon X \to \mathbb{C}$ with $|f|^p$ integrable. This space is a vector space under the usual addition and scalar multiplication of functions. If we define an equivalence relation on this set by $f \sim g$ iff $f = g$ a.e. then the space of equivalence classes is also a vector space. Let $L^p(X, \mathscr{B}, m)$ denote the space of equivalence classes, although we write $f \in L^p(X, \mathscr{B}, m)$ to denote that the function $f\colon X \to \mathbb{C}$ has $|f|^p$ integrable. The formula $\|f\|_p = [\int |f|^p\, dm]^{1/p}$ defines a norm on $L^p(X, \mathscr{B}, m)$ and this norm is complete. Therefore $L^p(X, \mathscr{B}, m)$ is a Banach space. If $L^p_R(X, \mathscr{B}, m)$ denotes those equivalence classes containing real-valued functions then $L^p_R(X, \mathscr{B}, m)$ is a real Banach space. The bounded measurable functions are dense in $L^p(X, \mathscr{B}, m)$. If $m(X) < \infty$ and $1 \leq p < q$ then $L^q(X, \mathscr{B}, m) \subset L^p(X, \mathscr{B}, m)$. We sometimes write $L^p(m)$ or $L^p(\mathscr{B})$ instead of $L^p(X, \mathscr{B}, m)$ when no confusion can arise.

A Hilbert space $\mathscr{H}$ is a Banach space in which the norm is given by an inner product, i.e., $\mathscr{H}$ is a Banach space and there is a map $(\cdot, \cdot)\colon \mathscr{H} \times \mathscr{H} \to \mathbb{C}$ such that $(\cdot, \cdot)$ is bilinear, $(f, g) = \overline{(g, f)}\ \forall g, f \in \mathscr{H}$, $(f, f) \geq 0\ \forall f \in \mathscr{H}$, and $f = (f, f)^{1/2}$ is the norm on $\mathscr{H}$.

The Banach space $L^p(X, \mathscr{B}, m)$ is a Hilbert space iff $p = 2$. The inner product in $L^2(X, \mathscr{B}, m)$ is given by $(f, g) = \int f\bar{g}\, dm$.

In every Hilbert space $\mathscr{H}$ we have the Schwarz inequality:

$$|(f,g)| \leq \|f\| \cdot \|g\| \qquad \forall f, g \in \mathscr{H}.$$

Separable Hilbert spaces (i.e. those having a countable dense set) are the simplest. The space $L^2(X, \mathscr{B}, m)$ is separable iff $(X, \mathscr{B}, m)$ has a *countable basis*, in the sense that there is a sequence of elements $\{E_n\}_1^\infty$ of $\mathscr{B}$ such that for every $\varepsilon > 0$ and every $B \in \mathscr{B}$ with $m(B) < \infty$ there is some n with $m(B \triangle E_n) < \varepsilon$. If X is a metric space and $\mathscr{B}$ is the σ-algebra of Borel subsets of X (the σ-algebra generated by the open sets) and m is any probability measure on $(X, \mathscr{B})$ then $(X, \mathscr{B}, m)$ has a countable basis. (This follows from Theorem 6.1.) Therefore most of the spaces we shall deal with have $L^2(X, \mathscr{B}, m)$ separable.

Any separate Hilbert space $\mathscr{H}$ contains a basis $\{e_n\}_1^\infty$, i.e. $(e_n, e_k) = 0$ if $n \neq k$ and only the zero element is orthogonal to all the e_n. If $\{e_n\}_1^\infty$ is a basis then each $v \in \mathscr{H}$ is uniquely expressible as $v = \sum_{n=1}^\infty a_n e_n$ where $a_n \in \mathbb{C}$. We have

$$\|v\|^2 = \sum_{n=1}^\infty |a_n|^2 \quad \text{so that } \sum_{n=1}^\infty |a_n|^2 < \infty.$$

An isomorphism between two Hilbert spaces $\mathscr{H}_1$, $\mathscr{H}_2$ is a linear bijection $W: \mathscr{H}_1 \to \mathscr{H}_2$ that preserves norms ($\|Wv\| = \|v\|\ \forall v \in \mathscr{H}_1$). The norm-preserving condition can also be written as $(Wu, Wv) = (u, v)\ \forall u, v \in \mathscr{H}_1$. Any two separable Hilbert spaces are isomorphic if they both have a basis with an infinite number of elements. A Hilbert space with a basis of k elements is isomorphic to $\mathbb{C}^k$. An isomorphism of a Hilbert space $\mathscr{H}$ to itself is called a unitary operator.

If V is a closed subspace of a Hilbert space $\mathscr{H}$ then $V^\perp = \{h \in \mathscr{H} \mid (v, h) = 0\ \forall v \in V\}$ is a closed subspace of $\mathscr{H}$ and $V \oplus V^\perp = \mathscr{H}$ (i.e. each $f \in \mathscr{H}$ has a unique representation $f = f_1 + f_2$ where $f_1 \in V$ and $f_2 \in V^\perp$.) The linear operator $P: \mathscr{H} \to V$ given by $P(f) = f_1$ is called the orthogonal projection of $\mathscr{H}$ onto V. In fact $P(f)$ is the unique member of V that satisfies $\|f - P(f)\| = \inf\{\|f - v\| \mid v \in V\}$. We have $P|V = id$ and $(Pf, g) = (f, Pg)\ \forall f, g \in \mathscr{H}$.

Let $(X, \mathscr{B}, m)$ be a probability space and recall from §0.4 that if $\mathscr{C}$ is a sub-σ-algebra of $\mathscr{B}$ then the conditional expectation operator

$$E(\cdot/\mathscr{C}): L^1(X, \mathscr{B}, m) \to L^1(X, \mathscr{C}, m)$$

is defined. Since $L^2(X, \mathscr{B}, m) \subset L^1(X, \mathscr{B}, m)$ the conditional expectation operator acts on $L^2(X, \mathscr{B}, m)$ and the following result describes what this restriction is.

Theorem 0.12. *Let $(X, \mathscr{B}, m)$ be a probability space and let $\mathscr{C}$ be a sub-σ-algebra of $\mathscr{B}$. The restriction of the conditional expectation operator $E(\cdot/\mathscr{C})$ to $L^2(X, \mathscr{B}, m)$ is the orthogonal projection of $L^2(X, \mathscr{B}, m)$ onto $L^2(X, \mathscr{C}, m)$.*