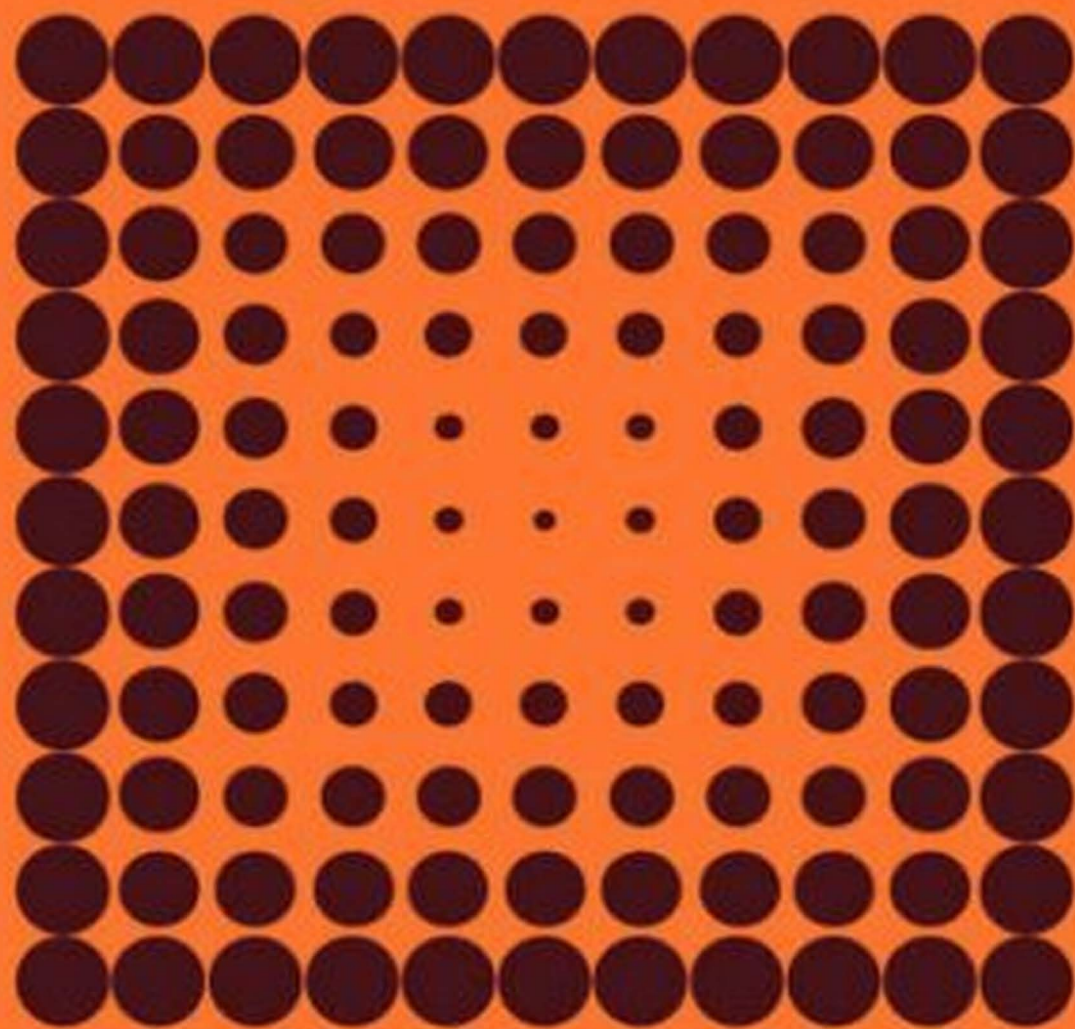


S. J. TAYLOR

Introduction to Measure and Integration



CAMBRIDGE UNIVERSITY PRESS

INTRODUCTION TO MEASURE AND INTEGRATION

BY

S. J. TAYLOR

*Professor of Mathematics at Westfield College,
University of London*



CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521098045

© Cambridge University Press 1966, 1973

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published as Chs. 1–9 of Kingman and Taylor
Introduction to Measure and Probability 1966
Reprinted as *Introduction to Measure and Integration* 1973
Re-issued in this digitally printed version 2008

A catalogue record for this publication is available from the British Library

Library of Congress Catalogue Card Number: 73–84325

ISBN 978-0-521-09804-5 paperback

CONTENTS

<i>Preface</i>	<i>page v</i>
1 Theory of sets	
1.1 Sets	1
1.2 Mappings	3
1.3 Cardinal numbers	5
1.4 Operations on subsets	9
1.5 Classes of subsets	14
1.6 Axiom of choice	19
2 Point set topology	
2.1 Metric space	23
2.2 Completeness and compactness	29
2.3 Functions	35
2.4 Cartesian products	38
2.5 Further types of subset	41
2.6 Normed linear space	44
2.7 Cantor set	49
3 Set functions	
3.1 Types of set function	51
3.2 Hahn–Jordan decompositions	61
3.3 Additive set functions on a ring	65
3.4 Length, area and volume of elementary figures	69
4 Construction and properties of measures	
4.1 Extension theorem; Lebesgue measure	74
4.2 Complete measures	81
4.3 Approximation theorems	84
4.4* Geometrical properties of Lebesgue measure	88
4.5 Lebesgue–Stieltjes measure	95
5 Definitions and properties of the integral	
5.1 What is an integral?	100
5.2 Simple functions; measurable functions	101
5.3 Definition of the integral	110
5.4 Properties of the integral	115
5.5 Lebesgue integral; Lebesgue–Stieltjes integral	124
5.6* Conditions for integrability	127

6 Related spaces and measures

6.1	Classes of subsets in a product space	<i>page</i> 134
6.2	Product measures	138
6.3	Fubini's theorem	143
6.4	Radon–Nikodym theorem	148
6.5	Mappings of measure spaces	153
6.6*	Measure in function space	157
6.7	Applications	162

7 The space of measurable functions

7.1	Point-wise convergence	166
7.2	Convergence in measure	171
7.3	Convergence in p th mean	174
7.4	Inequalities	183
7.5*	Measure preserving transformations from a space to itself	187

8 Linear functionals

8.1	Dependence of \mathcal{L}_2 on the underlying $(\Omega, \mathcal{F}, \mu)$	194
8.2	Orthogonal systems of functions	199
8.3	Riesz–Fischer theorem	202
8.4*	Space of linear functionals	209
8.5*	The space conjugate to \mathcal{L}_p	215
8.6*	Mean ergodic theorem	219

9 Structure of measures in special spaces

9.1	Differentiating a monotone function	224
9.2	Differentiating the indefinite integral	230
9.3	Point-wise differentiation of measures	236
9.4*	The Daniell integral	241
9.5*	Representation of linear functionals	250
9.6*	Haar measure	254

<i>Index of notation</i>	261
--------------------------	-----

<i>General Index</i>	263
----------------------	-----

PREFACE

There are many ways of developing the theory of measure and integration. In the present book measure is studied first as the primary concept and the integral is obtained later by extending its definition from the special case of 'simple' functions using monotone limits. The theory is presented for general measure spaces though at each stage Lebesgue measure and the Lebesgue integral in R^n are considered as the most important example, and the detailed properties are established for the Lebesgue case.

The book is designed for use either in the final undergraduate year at British universities or as a basic text in measure theory at the post-graduate level. Though the subject is developed as a branch of pure mathematics, it is presented in such a way that it has immediate application to any branch of applied mathematics which requires the basic theory of measure and integration as a foundation for its mathematical apparatus. In particular, our development of the subject is a suitable basis for modern probability theory – in fact this book first appeared as the initial section of the book *Introduction to measure and probability* (Cambridge University Press, 1966) written jointly with J. F. C. Kingman.

The book is largely self-contained. The first two chapters contain the essential parts of set theory and point set topology; these could well be omitted by a reader already familiar with these subjects. Chapters 3 and 4 develop the theory of measure by the usual process of extension from 'simple sets' to those of a larger class, and the properties of Lebesgue measure are obtained. The integral is defined in Chapter 5, again by extending its definition stage by stage, using monotone sequences. Chapter 6 includes a discussion of product measures and a definition of measure in function space. Convergence in function space is considered in Chapter 7, and Chapter 8 includes a treatment of complete orthonormal sets in Hilbert space. Chapter 9 deals with special spaces; differentiation theory for real functions of a real variable is developed and related to Lebesgue measure theory, and the Haar measure on a locally compact group is defined.

Starred sections contain more advanced material and can be omitted at a first reading.

It will be clear to any reader familiar with the standard treatises that this book owes much to what has gone before. I do not claim any particular originality for the treatment, but the form of presentation owes much to my experience of teaching this subject – at Birmingham

University, Cornell University and the University of London – and I readily acknowledge the stimulus received from this source. I am grateful to Dr B. Fishel and Professor G. E. H. Reuter who made helpful criticisms of an early draft, and to a great number of students and colleagues who pointed out misprints and errors in the first edition. However my main debt of gratitude is to Professor J. F. C. Kingman who was co-author of the first edition of this book, and who was much involved in every detail of it.

S.J.T.

London
December 1972

THEORY OF SETS

1.1 Sets

We do not want to become involved in the logical foundations of mathematics. In order to avoid these we will adopt a rather naïve attitude to set theory. This will not lead us into difficulties because in any given situation we will be considering sets which are all contained in (are subsets of) a fixed set or space or suitable collections of such sets. The logical difficulties which can arise in set theory only appear when one considers sets which are 'too big'—like the set of all sets, for instance. We assume the basic algebraic properties of the positive integers, the real numbers, and Euclidean spaces and make no attempt to obtain these from more primitive set theoretic notions. However, we will give an outline development (in Chapter 2) of the topological properties of these sets.

In a space X a set E is well defined if there is a rule which determines, for each *element* (or point) x in X , whether or not it is in E . We write $x \in E$ (read ' x belongs to E ') whenever x is an element of E , and the negation of this statement is written $x \notin E$. Given two sets E, F we say that E is contained in F , or E is a subset of F , or F contains E and write $E \subset F$ if every element x in E also belongs to F . If $E \subset F$ and there is at least one element in F but not in E , we say that E is a proper subset of F .

Two sets E, F are *equal* if and only if they contain the same elements; i.e. if and only if $E \subset F$ and $F \subset E$. In this case we write $E = F$. This means that if we want to prove that $E = F$ we must prove both $x \in E \Rightarrow x \in F$ and $x \in F \Rightarrow x \in E$ (the symbol \Rightarrow should be read 'implies').

Since a set is determined by its elements, one of the commonest methods of describing a set is by means of a defining sentence: thus E is the set of all elements (of X) which have the property P (usually delineated). The notation of 'braces' is often used in this situation

$$E = \{x: x \text{ has property } P\}$$

but when we use this notation we will always assume that only elements x in some fixed set X are being considered—as otherwise logical paradoxes can arise. When a set has only a finite number of

elements we can write them down between braces $E = \{x, y, z, a, b\}$. In particular $\{x\}$ stands for the set containing the single element x . One must always distinguish between the element x and the set $\{x\}$, for example, the empty set \emptyset defined below is not the same as the class $\{\emptyset\}$ containing the empty set.

Empty set (or null set)

The set which contains no elements is called the empty set and will be denoted by \emptyset . Clearly

$$\emptyset = \{x: x \neq x\}, \quad \text{and} \quad \emptyset \subset E \text{ for all sets } E.$$

In fact since \emptyset contains no element, any statement made about the elements of \emptyset is true (as well as its negative).

There are some sets which will be considered very frequently, and we consistently use the following notation:

\mathbf{Z} , for the set of positive integers,

\mathbf{Q} , for the set of rationals,

$\mathbf{R} = \mathbf{R}^1$, for the set of all real numbers,

\mathbf{C} , for the set of complex numbers,

\mathbf{R}^n , for Euclidean n -dimensional space, i.e. the set of ordered n -tuples (x_1, x_2, \dots, x_n) where all the x_i are in \mathbf{R} .

We assume that the reader is familiar with the algebraic and order properties of these sets. In particular we will use the fact that \mathbf{Z} is *well ordered*, that is, that every non-empty set of positive integers has a least member: this is equivalent to the principle of mathematical induction.

We frequently have to consider sets of sets, and occasionally sets of sets of sets. It is convenient to talk of *classes* of sets and *collections* of classes to distinguish these types of set, and we will use italic capitals A, B, \dots for sets, script capitals $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ for classes and Greek capitals Δ, Γ, \dots for collections. Thus $C \in \mathcal{C}$ is read 'the set C belongs to the class \mathcal{C} '; and $\mathcal{A} \subset \mathcal{B}$ means that every set in the class \mathcal{A} is also in the class \mathcal{B} .

Cartesian product

Given two sets E, F we define the Cartesian (or direct) product $E \times F$ to be the set of all *ordered* pairs (x, y) whose first element $x \in E$ and whose second element $y \in F$. This clearly extends immediately to the product $E_1 \times E_2 \times \dots \times E_n$ of any finite number of sets. In particular it is immediate that \mathbf{R}^n , Euclidean n -space, is the Cartesian product

of n copies of \mathbf{R} . For an infinite indexed class $\{E_i, i \in I\}$ of sets, the product $\prod_{i \in I} E_i$ is the set of elements of the form $\{a_i, i \in I\}$ with $a_i \in E_i$ for each $i \in I$.

Exercises 1.1

1. Describe in words the following sets:

- (i) $\{t \in \mathbf{R}: 0 \leq t \leq 1\}$;
- (ii) $\{(x, y) \in \mathbf{R}^2: x^2 + y^2 \leq 1\}$;
- (iii) $\{k \in \mathbf{Z}: k = n^2 \text{ for some } n \in \mathbf{Z}\}$;
- (iv) $\{k \in \mathbf{Z}: n|k \Rightarrow n = 1 \text{ or } k\}$;
- (v) $\{\mathcal{A}: E \in \mathcal{A}\}$;
- (vi) $\{B: B \subset E\}$.

2. Show that the relation \subset is reflexive and transitive, but not in general symmetric.

3. The sets $X \times (Y \times Z)$ and $(X \times Y) \times Z$ are different but there is a natural correspondence between them.

4. Suppose x is an element of X and $A = \{x\}$. Which of the following statements are correct: $x \in A$, $x \in X$, $x \subset A$, $x \subset X$, $A \in X$, $A \subset X$, $A \subset x$?

5. Suppose $P(\alpha)$ and $Q(\alpha)$ are two propositions about the element α such that $P(\alpha) \Rightarrow Q(\alpha)$. Show that $\{\alpha: P(\alpha)\} \subset \{\alpha: Q(\alpha)\}$.

1.2 Mappings

Suppose A and B are any two sets: a *function* from A to B is a rule which, for each element in A , determines a unique element in B . We talk of the function f and use the notation $f: A \rightarrow B$ to denote a function f defined on A and taking values in B . For any $x \in A$, $f(x)$ means the value of the function f at the point x and is therefore an element of the set B : we therefore avoid the terminology (common in older text books) 'the function $f(x)$ '. The words *mapping* and *transformation* are often used as a synonym for function.

For a given function $f: A \rightarrow B$, we call A the *domain* of f and the subset of B consisting of the set of values $f(x)$ for x in A is called the *range* of f and may be denoted $f(A)$. When $f(A) = B$ we say that f is a function from A *onto* B . Given a function $f: A \rightarrow B$, by definition $f(x)$ is a uniquely determined element of B for each $x \in A$; if in addition for each y in $f(A)$ there is a *unique* $x \in A$ (we know there is at least one) with $y = f(x)$ we say that the function f is (1, 1). Another shorter way of saying this is that $f: A \rightarrow B$ is (1, 1) if and only if for $x_1, x_2 \in A$,

$$x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2).$$

Given $f: A \rightarrow B$ there is an associated $f: \mathcal{A} \rightarrow \mathcal{B}$, where \mathcal{A} is the class of all subsets of A and \mathcal{B} is the class of all subsets of B , defined by

$$f(E) = \{y \in B: \exists x \in E \text{ with } y = f(x)\}$$

for each $E \subset A$. (the symbol \exists should be read, 'there exists': i.e. the set described by $\{x \in E: y = f(x)\}$ is not empty). There is also a function $f^{-1}: \mathcal{B} \rightarrow \mathcal{A}$ defined by

$$f^{-1}(F) = \{x \in A: f(x) \in F\},$$

for each $F \subset B$. The set $f^{-1}(F)$ is called the *inverse image* of F under f . Note that if $y \in B - f(A)$, then the inverse image $f^{-1}(\{y\})$ of the one point set $\{y\}$ is the empty set. If $f: A \rightarrow B$ is $(1, 1)$ and $y \in f(A)$, then it is clear that $f^{-1}(\{y\})$ is a one point subset of A , so that in this case (only) we can think of f^{-1} as a function from $f(A)$ to A . In particular, if $f: A \rightarrow B$ is $(1, 1)$ and onto there is a function $f^{-1}: B \rightarrow A$ called the *inverse function* of f such that $f^{-1}(y) = x$ if and only if $y = f(x)$.

Now suppose $f: A_1 \rightarrow B$, $g: A_2 \rightarrow B$ are functions such that $A_1 \supset A_2$ and $f(x) = g(x)$ for all x in A_2 : under these conditions we say that f is an *extension* of g (from A_2 to A_1) and g is the *restriction* of f (to A_2). For example, if

$$g(x) = \cos x \quad (x \in \mathbf{R});$$

$$f(x + iy) = \cos x \cosh y + i \sin x \sinh y \quad (x + iy \in \mathbf{C});$$

then $f: \mathbf{C} \rightarrow \mathbf{C}$ is an extension of $g: \mathbf{R} \rightarrow \mathbf{C}$ from \mathbf{R} to \mathbf{C} , and the usual convention of designating both f and g by 'cos' obscures the differences in their domains.

If we have two functions $f: A \rightarrow B$, $g: B \rightarrow C$ the result of applying the rule for g to the element $f(x)$ defines an element in C for all $x \in A$. Thus we have defined a function $h: A \rightarrow C$ which is called the *composition* of f and g and denoted $g \circ f$ or $g(f)$. Thus, for $x \in A$

$$h(x) = (g \circ f)x = g(f(x)) \in C.$$

Note that, if $f: A \rightarrow B$ is $(1, 1)$ and onto we could define the inverse function $f^{-1}: B \rightarrow A$ as the unique function from B to A such that

$$(f \circ f^{-1})(y) = y \quad \text{for all } y \in B,$$

$$(f^{-1} \circ f)(x) = x \quad \text{for all } x \in A.$$

Sequence

Given any set X a *finite sequence* of n points of X is a function from $\{1, 2, \dots, n\}$ to X . This is usually denoted by x_1, x_2, \dots, x_n where $x_i \in X$ is the value of the function at the integer i . Similarly, an *infinite*

sequence in X is a function from \mathbf{Z} to X (where \mathbf{Z} is the set of positive integers). This is denoted x_1, x_2, \dots , or $\{x_i\}$ ($i = 1, 2, \dots$), or just $\{x_i\}$ where x_i is the value of the function at i , and is called the i th element of the sequence. Given a sequence $\{n_i\}$ of positive integers (that is, a function $f: \mathbf{Z} \rightarrow \mathbf{Z}$ where $f(i) = n_i$) such that $n_i > n_j$ for $i > j$, and a sequence $\{x_i\}$ of elements of X (a function $g: \mathbf{Z} \rightarrow X$) it is clear that the composite function $g \circ f: \mathbf{Z} \rightarrow X$ is again a sequence. Such a sequence is called a *subsequence* of $\{x_i\}$ and is denoted $\{x_{n_i}\}$ ($i = 1, 2, \dots$). Thus $\{x_{n_i}\}$ is a subsequence of $\{x_i\}$ if $n_i \in \mathbf{Z}$ for all $i \in \mathbf{Z}$, and $i > j \Rightarrow n_i > n_j$.

We can think of a sequence as a point in the product space $\prod_{i=1}^{\infty} X_i$ where $X_i = X$ for all i . More generally a point in the product space $\prod_{i \in I} X_i$ with $X_i = X$ for $i \in I$ can be identified as a function $f: I \rightarrow X$.

Exercises 1.2

1. Suppose $f: \mathbf{R} \rightarrow \mathbf{R}$ is defined by $f(x) = \sin x$. Describe each of the following sets:

$$f^{-1}\{0\}, \quad f^{-1}\{1\}, \quad f^{-1}\{2\}, \quad f^{-1}\{y: 0 \leq y \leq \tfrac{1}{2}\}.$$

2. Suppose $f: A \rightarrow B$ is any function. Prove

- (i) $E \subset f^{-1}(f(E))$, for each $E \subset A$;
- (ii) $F \supset f(f^{-1}(F))$, for each $F \subset B$;

and give examples in which there is not equality in (i), (ii).

3. Suppose $f: A \rightarrow B$, $g: B \rightarrow C$ are functions and $h = g \circ f$: show that $h^{-1}(E) = f^{-1}[g^{-1}(E)]$ for each $E \subset C$.

4. If $A \subset B \subset C$, $f: A \rightarrow X$, $g: B \rightarrow X$, $h: C \rightarrow X$ are such that h is an extension of g and g is an extension of f , prove that f is the restriction of h to A .

5. Show that the restriction of a $(1, 1)$ mapping is $(1, 1)$.

6. Suppose $m, n \in \mathbf{Z}$, A is a set with m distinct elements and B is a set with n distinct elements. How many distinct functions are there from A to B ?

1.3 Cardinal numbers

If there is a mapping $f: A \rightarrow B$ which is $(1, 1)$ and onto, then it is reasonable to say that there are the same number of elements in A as there are in B . In fact, for finite sets, the elementary process of counting sets up such a mapping from the set being counted to the integers $\{1, 2, \dots, n\}$, and from experience we know that if the same finite set of objects is counted in different ways we always end up with

the same integer n . (This fact can also be deduced from primitive axioms about the integers.) We say that the set A is equivalent to the set B , and write $A \sim B$ if there is a mapping $f: A \rightarrow B$ which is (1, 1) and onto. It is clear that \sim is an equivalence relation between sets in the sense that it is reflexive, symmetric and transitive, and we can therefore form equivalence classes of sets with respect to this relation. Such an equivalence class of sets is called a *cardinal number*, but by noting that the equivalence class is determined by any one of its members, we see that the easiest way to specify a cardinal number is to specify a representative set. Thus any set which can be mapped (1, 1) onto the representative set will have the same cardinal. As is usual we shall use the following notation:

- the cardinal of the empty set \emptyset is 0;
- the cardinal of the set of integers $\{1, 2, \dots, n\}$ is n ;
- the cardinal of the set \mathbf{Z} of positive integers is \aleph_0 ;
- the cardinal of the set \mathbf{R} of real numbers is c .

Since \mathbf{Z} is ordered we can clearly order the cardinals of *finite* sets by saying that A has a smaller cardinal than B if A is equivalent to a proper subset of B . This definition does not work for infinite sets as the mappings

$$n \rightarrow 2^n \quad \text{or} \quad n \rightarrow n^2$$

map \mathbf{Z} onto a proper subset of \mathbf{Z} and are (1, 1). Instead we say that the cardinal of a set A is less than the cardinal of the set B if there is a subset $B_1 \subset B$ such that $A \sim B_1$ but no subset $A_1 \subset A$ such that $A_1 \sim B$.

From this definition of ordering we consider the following statements, where m, n, p denote cardinals

- (i) $m < n, n < p \Rightarrow m < p$;
- (ii) at *most* one of the relations $m < n, m = n, n < m$ holds so that $m \leq n, n \leq m \Rightarrow m = n$.
- (iii) at *least* one of the relations $m < n, m = n, n < m$ holds.

Now (i) follows easily from the definition, for let M, N, P be sets with cardinals m, n, p and suppose $N_1 \subset N, P_1 \subset P$ with $M \sim N_1, N \sim P_1$. The mapping $f: N \rightarrow P_1$ when restricted to N_1 gives an equivalence $N_1 \sim P_2 \subset P_1$ so that $M \sim P_2 \subset P$. Further if $P \sim M_1 \subset M$ the mapping $g: M \rightarrow N_1$ when restricted to M_1 shows $P \sim M_1 \sim N_2 \subset N$ which contradicts $n < p$. (ii) can also be deduced from the definition (see exercise 1.3 (5)), though this requires quite a complicated argument: (ii) is known as the Schröder-Bernstein theorem. However, the truth of (iii)—that all cardinals are comparable—cannot be proved without

the use of an additional axiom (known as the axiom of choice) which we will discuss briefly in §1.6. If we assume the axiom of choice or something equivalent, then (iii) is also true.

A set of cardinal \aleph_0 is said to be *enumerable*. Thus such a set $A \sim \mathbb{Z}$ so that the elements of A can be 'enumerated' as a sequence a_1, a_2, \dots in which each element of A occurs once and only once. A set which has a cardinal $m \leq \aleph_0$ is said to be *countable*. Thus E is countable if there is a subset $A \subset \mathbb{Z}$ such that $E \sim A$, and a set is countable if it is either finite or enumerable.

Given any infinite set B we can choose, by induction, a sequence $\{b_i\}$ of distinct elements in B and if B_1 is the set of elements in $\{b_i\}$ the cardinal of B_1 is \aleph_0 . Hence if m is an infinite cardinal we always have $m \geq \aleph_0$. By using the equivalence

$$b_i \leftrightarrow b_{2i}$$

between B_1 and the proper subset $B_2 \subset B_1$ where B_2 contains the even elements of $\{b_i\}$ and the identity mapping

$$b \leftrightarrow b \quad \text{for } b \in B - B_1$$

we have an equivalence between $B = B_1 \cup (B - B_1)$ and $B_2 \cup (B - B_1)$, a proper subset of B . This shows that any infinite set B contains a proper subset of the same cardinal.

In order to see that some infinite sets have cardinal $> \aleph_0$ it is sufficient to recall that the set $\{x \in \mathbb{R} : 0 < x < 1\}$ cannot be arranged as a sequence.† Now $\pi^{-1} \tan^{-1} x + \frac{1}{2} = f(x)$, $x \in \mathbb{R}$ defines a mapping $f: \mathbb{R} \rightarrow (0, 1)$ which is (1, 1) and onto so that \mathbb{R} has the same cardinal as the interval $(0, 1)$ and we have $c > \aleph_0$. It is worth remarking that a famous unsolved problem of mathematics concerns the existence or otherwise of cardinals m such that $c > m > \aleph_0$. The axiom that no such exist, that is that $m > \aleph_0 \Rightarrow m \geq c$ is known as the *continuum hypothesis*.

The fact that there are infinitely many different infinite cardinals follows from the next theorem, which compares the cardinal of a set E with the cardinal of the class of subsets of E .

Theorem 1.1. *For any set E , the class $\mathcal{C} = \mathcal{C}(E)$ of all subsets of E has a cardinal greater than that of E .*

Proof. For sets E of finite cardinal n , one can prove directly that the cardinal of $\mathcal{C}(E)$ is 2^n , and an induction argument easily yields $n < 2^n$ for $n \in \mathbb{Z}$. However, the case of finite sets E is included in the general proof, so there is nothing gained by this special argument.

† See, for example, J. C. Burkill, *A First Course in Mathematical Analysis* (Cambridge, 1962).

Suppose \mathcal{Q} is the class of one points sets $\{x\}$ with $x \in E$. Then $\mathcal{Q} \subset \mathcal{C}$ and $E \sim \mathcal{Q}$ because of the mapping $x \leftrightarrow \{x\}$. Therefore it is sufficient to prove by (ii) above, that \mathcal{C} is equivalent to no subset $E_1 \subset E$. Suppose then that $\phi: \mathcal{C} \rightarrow E_1$ is $(1, 1)$ and onto and let $\chi: E_1 \rightarrow \mathcal{C}$ denote the inverse function. Let A be the subset of E_1 defined by

$$A = \{x \in E_1, \quad x \notin \chi(x)\}.$$

Then $A \in \mathcal{C}$ so that $\phi(A) = x_0 \in E_1$. Now if $x_0 \in A$, $\chi(x_0) = A$ does not contain x_0 which is impossible, while if $x_0 \notin A$, then x_0 is not in $\chi(x_0)$ so that $x_0 \in A$. In either case we have a contradiction. \blacksquare

It is possible to build up systematically an arithmetic of cardinals. This will only be needed for finite cardinals and \aleph_0 in this book, so we restrict the results to these cases and discuss them in the next section.

Exercises 1.3

1. Show that $(0, 1] \sim (0, 1)$ by considering f defined by

$$\begin{aligned} f(x) &= \frac{3}{2} - x, & \text{for } \frac{1}{2} < x \leq 1; \\ &= \frac{3}{4} - x, & \text{for } \frac{1}{4} < x \leq \frac{1}{2}; \\ &= \frac{3}{8} - x, & \text{for } \frac{1}{8} < x \leq \frac{1}{4}; \\ &= \frac{3}{2^n} - x, & \text{for } \frac{1}{2^n} < x \leq \frac{1}{2^{n-1}} \quad (n = 1, 2, \dots). \end{aligned}$$

Deduce that all intervals (a, b) , $(a, b]$, $[a, b]$ or $[a, b)$ with $a < b$ have the same cardinal c .

2. Every function $f: [a, b] \rightarrow \mathbf{R}$ which is monotonic, i.e.

$$a \leq x_1 < x_2 \leq b \Rightarrow f(x_1) \leq f(x_2),$$

is discontinuous at the points of a countable subset of $[a, b]$.

Hint. Consider the sets of points x where the size of the discontinuity $d(x) = f(x+0) - f(x-0)$ satisfies $1/(n+1) < d(x) \leq 1/n$ and prove this is finite for all n in \mathbf{Z} .

3. Show that $\mathbf{R}^2 \sim \mathbf{R}$.

Hint.

$$(\cdot a_1 a_2 a_3 \dots, \cdot b_1 b_2 b_3 \dots) \leftrightarrow \cdot a_1 b_1 a_2 b_2 \dots$$

defines a $(1, 1)$ mapping between pairs of decimal expansions and single expansions of numbers in $(0, 1)$. Modify this mapping to eliminate the difficulty caused by the fact that decimal expansions are not quite unique.

4. Prove that a finite set E of cardinal m has 2^m distinct subsets.

5. Suppose $A_1 \subset A$, $B_1 \subset B$, $A_1 \sim B$ and $A \sim B_1$. Construct a mapping to show that $A \sim B$.

Hint. Suppose $f: A \rightarrow B_1$, $g: B \rightarrow A_1$ are (1, 1) and onto. Say x (in either A or B) is an ancestor of y if and only if y can be obtained from x by successive applications of f and g . Decompose A into 3 sets A_o, A_e, A_i according as to whether the element x has an odd, even or infinite number of ancestors and decompose B similarly. Consider the mapping which agrees with f on A_e and A_i , and with g^{-1} on A_o .

1.4 Operations on subsets

For two sets A, B we define the *union* of A and B (denoted $A \cup B$) to be the set of elements in either A or B or both. The *intersection* of A and B (denoted $A \cap B$) is the set of elements in both A and B .

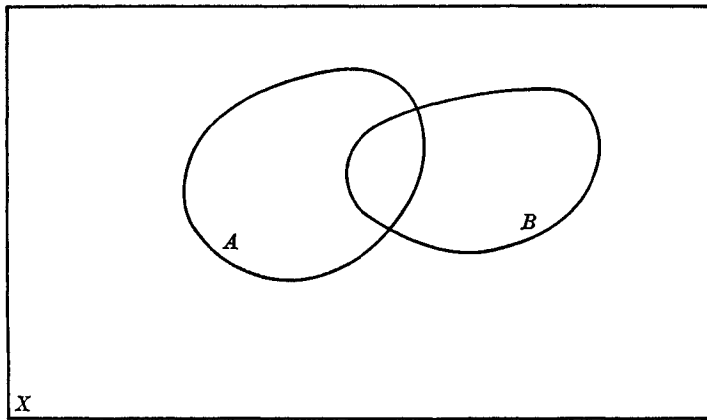


Fig. 1

If $A \subset X$, the *complement* of A with respect to X (denoted $X - A$) is the set of those elements in X which are not in A . We also use $(A - B)$ to denote the set of elements in A which are not in B for arbitrary sets A, B . For any two sets A, B the *symmetric difference* (denoted $A \triangle B$) is $(A - B) \cup (B - A)$, that is the set of elements which are in one of A, B but not in both. Note that $A \triangle B = B \triangle A$.

These finite operations on sets are best illustrated by means of a Venn diagram. In this some figure (like a rectangle) denotes the whole space X and suitable geometrical figures inside denote the subsets A, B , etc. It is well known that drawing does not prove a theorem, but the reader is advised to illustrate the results of the next paragraph by means of suitable Venn diagrams (see Figure 1).

The operations \cup, \cap, \triangle satisfy algebraic laws, some of which are

listed below. We assume the reader is familiar with these, so proofs are omitted.

- (i) $A \cup B = B \cup A$, $A \cap B = B \cap A$;
- (ii) $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$;
- (iii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
- (iv) $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$;
- (v) if $A \subset X$, then $A \cup X = X$, $A \cap X = A$;
- (vi) if $A \subset X$, $B \subset X$, then $X - (A \cup B) = (X - A) \cap (X - B)$,
 $X - (A \cap B) = (X - A) \cup (X - B)$;
- (vii) $A \cup B = (A \triangle B) \triangle (A \cap B)$, $A - B = A \triangle (A \cap B)$.

A similarity between the laws satisfied by \cap , \cup and the usual algebraic laws for multiplication and addition can be observed (in fact the older notation for these operations is product and sum) but the differences should also be noted: in particular the distributive laws, (iii) above, are different in the algebra of sets. (vi) above will be generalized and proved as a lemma—it is known as de Morgan's law.

Given a class \mathcal{C} of subsets A , the *union* $\bigcup \{A; A \in \mathcal{C}\}$ is the set of elements which are in at least one set A belonging to \mathcal{C} and the *intersection* $\bigcap \{A; A \in \mathcal{C}\}$ is the set of elements which are in every set A of \mathcal{C} . If the class \mathcal{C} is indexed so that \mathcal{C} consists precisely of the sets A_α , ($\alpha \in I$), then we use the notations $\bigcup_{\alpha \in I} A_\alpha$, $\bigcap_{\alpha \in I} A_\alpha$ for the union and intersection of the class. In particular when \mathcal{C} is finite or enumerable it is usual to assume that it is indexed by $\{1, 2, \dots, n\}$ or \mathbb{Z} respectively and the notation is

$$\bigcup_{i=1}^n A_i, \quad \bigcap_{i=1}^n A_i, \quad \bigcup_{i=1}^{\infty} A_i, \quad \bigcap_{i=1}^{\infty} A_i.$$

When the class \mathcal{C} is empty, that is $I = \emptyset$, we adopt the conventions

$$\bigcup_{\alpha \in I} E_\alpha = \emptyset, \quad \bigcap_{\alpha \in I} E_\alpha = X, \quad \text{the whole space.}$$

This ensures that certain identities are valid without restriction on I .

Lemma. Suppose E_α , $\alpha \in I$ is a class of subsets of X , and E_1 is one set of the class, then

- (i) $\bigcap_{\alpha \in I} E_\alpha \subset E_1 \subset \bigcup_{\alpha \in I} E_\alpha$;
- (ii) $X - \bigcup_{\alpha \in I} E_\alpha = \bigcap_{\alpha \in I} (X - E_\alpha)$;
- (iii) $X - \bigcap_{\alpha \in I} E_\alpha = \bigcup_{\alpha \in I} (X - E_\alpha)$.

Proof. (i) This is immediate from the definition.

(ii) Suppose $x \in X - \bigcup_{\alpha \in I} E_\alpha$, then $x \in X$ and x is not in $\bigcup_{\alpha \in I} E_\alpha$, that is x is not in any E_α , $\alpha \in I$ so that $x \in X - E_\alpha$ for every α in I , and $x \in \bigcap_{\alpha \in I} (X - E_\alpha)$. Conversely if $x \in \bigcap_{\alpha \in I} (X - E_\alpha)$, then for every $\alpha \in I$, x is in X but not in E_α , so $x \in X$ but x is not in $\bigcup_{\alpha \in I} E_\alpha$; that is, $x \in X - \bigcup_{\alpha \in I} E_\alpha$.

(iii) Similar to (ii). **■**

Two sets A, B are said to be *disjoint* if they have no elements in common; that is, if $A \cap B = \emptyset$. A *disjoint class* is a class \mathcal{C} of sets such that any two distinct sets of \mathcal{C} are disjoint. The union of a disjoint class is sometimes called a *disjoint union*.

Lemma. Given a finite or enumerable union of sets $\bigcup_{i=1}^p E_i$ (where p can be $+\infty$), there are subsets $F_i \subset E_i$ such that the sets F_i are disjoint and $\bigcup_{i=1}^p E_i = \bigcup_{i=1}^p F_i$.

Proof. We write out the details for $p = \infty$. Only obvious changes are needed for $p \in \mathbb{Z}$. Put $C = \bigcup_{i=1}^p E_i$ and define $F_1 = E_1$,

$$F_n = E_n - \bigcup_{i=1}^{n-1} E_i \quad (n = 2, 3, \dots).$$

Then $F_n \subset E_n$ for all n , and if $i > j$, F_i and E_j are disjoint, so that F_i, F_j must be disjoint. Further if $x \in C$, and n is the smallest integer (which exists because \mathbb{Z} is well ordered) such that $x \in E_n$; then $x \in E_n$ but not to E_i for $i < n$. Thus $x \in F_n$ and so $x \in \bigcup_{i=1}^{\infty} F_i$. Thus

$$C \subset \bigcup_{i=1}^{\infty} F_i, \text{ and the reverse inclusion is immediate. } \mathbf{■}$$

Theorem 1.2. The union of a countable class of countable sets is a countable set.

Proof. By the process of the above lemma we can replace the countable union by a countable disjoint union of sets which are subsets of those in the original class—each of which is therefore countable. Each countable set can be enumerated as a finite or infinite sequence. So we have

$$C = \bigcup_{i=1}^{\infty} E_i \quad \text{a disjoint union,}$$

$$E_i = \{x_{ij}\} \quad (j = 1, 2, \dots),$$

where the infinite union may be a finite one and some (or all) of the sequences $\{x_{ij}\}$ may be finite. Put $F_n = \{x_{ij} : i + j = n\}$, then F_n is a

finite set containing at most $(n+1)$ elements. The sets F_n are disjoint, and C can be enumerated by first enumerating F_1 , then F_2 , and so on. If $F_n = \emptyset$ for $n \geq N$ then C is finite; otherwise it is enumerable. \blacksquare

Corollary. *The set \mathbf{Q} of rational numbers is enumerable.*

Proof. $\mathbf{Q} = \bigcup_{n=1}^{\infty} E_n$, where E_n is the set of real numbers of the form p/n where p is an integer. E_n is enumerable since

$$0, +1, -1, +2, -2, \dots, +p, -p, \dots$$

is an enumeration of the set of integers. \blacksquare

For a sequence E_1, E_2, \dots of sets, we put

$$\limsup E_i = \bigcap_{n=1}^{\infty} \left(\bigcup_{i=n}^{\infty} E_i \right), \quad \liminf E_i = \bigcup_{n=1}^{\infty} \left(\bigcap_{i=n}^{\infty} E_i \right);$$

and if $\{E_i\}$ is such that $\limsup E_i = \liminf E_i$ we say that the sequence converges to the set $E = \limsup E_i = \liminf E_i$. For any sequence $\{E_i\}$, $\limsup E_i$ is the set of those elements which are in E_i for infinitely many i and $\liminf E_i$ is the set of those elements which are in all but a finite number of the sets E_i .

A sequence $\{E_i\}$ is said to be *increasing* if, for each positive integer n , $E_n \subset E_{n+1}$; it is said to be *decreasing* if, for each positive integer n , $E_n \supset E_{n+1}$. A monotone sequence of sets is one which is either increasing or decreasing. Note that any monotone sequence converges to a limit for

(i) If $\{E_i\}$ is increasing,

$$\bigcup_{i=n}^{\infty} E_i = \bigcup_{i=1}^{\infty} E_i, \quad \bigcap_{i=n}^{\infty} E_i = E_n \quad \text{for all } n,$$

so that $\limsup E_i = \liminf E_i = \bigcup_{i=1}^{\infty} E_i$; while

(ii) If $\{E_i\}$ is decreasing,

$$\bigcup_{i=n}^{\infty} E_i = E_n, \quad \bigcap_{i=n}^{\infty} E_i = \bigcap_{i=1}^{\infty} E_i \quad \text{for all } n,$$

so that $\limsup E_i = \liminf E_i = \bigcap_{i=1}^{\infty} E_i$.

Indicator function

Given a subset E of a space X , the function $\chi_E: X \rightarrow \mathbf{R}$ defined by

$$\chi_E(x) = \begin{cases} 1 & \text{for } x \in E, \\ 0 & \text{for } x \in X - E \end{cases}$$

is called the indicator function of E (many books use the term ‘characteristic function’ for χ_E , but we will avoid this term because characteristic function has a different meaning in probability theory). The correspondence between subsets of X and indicator functions is clearly (1, 1) for

$$E = \{x: \chi_E(x) = 1\},$$

and we will use indicator functions as a convenient tool for carrying out operations on sets.

Exercises 1.4

1. Prove each of the following set identities:

$$A - B = A - (A \cap B) = A \cup B - B,$$

$$A \cap (B - C) = A \cap B - A \cap C,$$

$$(A - B) - C = A - (B \cup C),$$

$$A - (B - C) = (A - B) \cup (A \cap C),$$

$$(A - B) \cap (C - D) = A \cap C - B \cup D,$$

$$E \triangle (F \triangle G) = (E \triangle F) \triangle G,$$

$$E \cap (F \triangle G) = (E \cap F) \triangle (E \cap G),$$

$$E \triangle \emptyset = E, E \triangle X = X - E,$$

$$E \triangle E = \emptyset, E \triangle (X - E) = X,$$

$$E \triangle F = (E \cup F) - (E \cap F).$$

2. With respect to which of the operations Δ , \cup , \cap does the class of all subsets of X form a group?

3. Show that $E \subset F$ if and only if $X - F \subset X - E$.

4. Prove that $A \triangle B = C \triangle D$ if and only if $A \triangle C = B \triangle D$, by showing that either equality is equivalent to the statement that every point of X is in 0, 2 or 4 of the sets A, B, C, D .

5. Show that if $I_1 \subset I_2$, then

$$\bigcap_{\alpha \in I_1} E_\alpha \supset \bigcap_{\alpha \in I_2} E_\alpha, \quad \bigcup_{\alpha \in I_1} E_\alpha \subset \bigcup_{\alpha \in I_2} E_\alpha.$$

6. A real number is said to be algebraic if it is a zero of a polynomial $a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$ where the coefficients a_i are integers. Defining the ‘height’ of a polynomial to be the integer

$$h = n + |a_n| + |a_{n-1}| + \dots + |a_0|,$$

show that there are only finitely many polynomials of height h , and deduce that the set of all algebraic real numbers is enumerable. Deduce that the

set of transcendental numbers (real numbers which are not algebraic) has cardinal $> \aleph_0$.

7. Show that in \mathbf{R}^n , the set \mathbf{Q}^n of points (x_1, x_2, \dots, x_n) , where each co-ordinate x_i is rational, is an enumerable set.

Further, the class of all spheres with centres at points of \mathbf{Q}^n and rational radii is an enumerable class.

8. Show that any sequence of disjoint sets converges to \emptyset . Show that $\{E_n\}$ is a convergent sequence if and only if there is no point x of X such that each of $x \in E_n$, $x \in X - E_n$ holds for infinitely many n .

Suppose

$$E_n = \begin{cases} \{x: 0 < x \leq 1 - (1/n)\} & n \text{ odd,} \\ \{x: (1/n) \leq x < 1\} & n \text{ even;} \end{cases}$$

show that $\{E_n\}$ converges but is not monotone.

9. For any sequence $\{E_n\}$ of sets prove

(i) $\limsup E_n$, $\liminf E_n$ are unaltered by the omission or alteration of any finite number of sets in the sequence.

(ii) for any set F ,

$$F - \limsup E_n = \liminf (F - E_n),$$

$$F - \liminf E_n = \limsup (F - E_n).$$

10. If $E_n = A$ for n even, $E_n = B$ for n odd, show that $\limsup E_n = A \cup B$, $\liminf E_n = A \cap B$.

11. Can an uncountable union of distinct sets be countable?

12. If $\{E_n\}$ is a sequence of sets and

$$D_1 = E_1, \quad D_n = D_{n-1} \triangle E_n \quad \text{for } n = 2, 3, \dots,$$

show that the sequence $\{D_n\}$ converges to a limit if and only if $\lim E_n = \emptyset$.

13. Show that $\chi_E(x) \leq \chi_F(x)$ for all x in X if and only if $E \subset F$. Suppose $A = E \cup F$, $B = E \cap F$, $C = E \triangle F$: show that

$$\chi_B = \chi_E \cdot \chi_F, \quad \chi_A = \chi_E + \chi_F - \chi_B, \quad \chi_C = |\chi_E - \chi_F|.$$

Generalise the first two of these identities to finite unions and intersections.

14. If χ_n is the indicator function of E_n ($n = 1, 2, \dots$) and $A = \limsup E_n$, $B = \liminf E_n$, show that, for all x in X ,

$$\chi_A(x) = \limsup_{n \rightarrow \infty} \chi_n(x), \quad \chi_B(x) = \liminf_{n \rightarrow \infty} \chi_n(x).$$

1.5 Classes of subsets

Up to the present our operations have been defined on the class \mathcal{C} of all subsets of a given set X . This class is too large for many pur-

poses and it is usual to restrict attention to subclasses of \mathcal{C} . However it is important that the subclasses considered have sufficient structure, and we now define various types of class starting with the simplest.

1. *Semi-ring*

A class \mathcal{S} of subsets such that

$$(i) \quad \emptyset \in \mathcal{S};$$

$$(ii) \quad A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S};$$

$$(iii) \quad A, B \in \mathcal{S} \Rightarrow A - B = \bigcup_{i=1}^n E_i, \text{ where the } E_i \text{ are disjoint sets in } \mathcal{S}, \text{ is}$$

called a semi-ring. (Note that many authors, following Von Neumann, who first defined the concept, have an additional condition in the definition of a semi-ring—instead of (iii) they assume that if $A, B \in \mathcal{S}$ and $B \subset A$ there is a finite class C_0, C_1, \dots, C_n of sets of \mathcal{S} such that $B = C_0 \subset C_1 \subset \dots \subset C_n = A$ and $D_i = C_i - C_{i-1} \in \mathcal{S}$ for $i = 1, 2, \dots, n$. This stronger condition causes complications and we weaken it since it is unnecessary.) An important example of a semi-ring of subsets of \mathbf{R} is the class $\mathcal{P} = \mathcal{P}^1$ of finite intervals $(a, b]$ which are open on the left and closed on the right. Similarly, \mathcal{P}^n consisting of the rectangles in \mathbf{R}^n of the form $\{(x_1, x_2, \dots, x_n): a_i < x_i \leq b_i\}$ is a semi-ring in \mathbf{R}^n .

2. *Ring*

This is any non-empty class \mathcal{R} of subsets such that

$$A, B \in \mathcal{R} \Rightarrow A \cap B \in \mathcal{R} \quad \text{and} \quad A \Delta B \in \mathcal{R}.$$

Since $\emptyset = A \Delta A$, $A \cup B = (A \Delta B) \Delta (A \cap B)$, and $A - B = A \Delta (A \cap B)$ we see that a ring is a class of sets closed under the operations of union, intersection, and difference and $\emptyset \in \mathcal{R}$. Thus a ring is certainly also a semi-ring. As examples the system $\{\emptyset, X\}$ is a ring as is the class of all subsets of X . However, the class \mathcal{P} of half-open intervals in \mathbf{R} is not a ring, for it is not closed under the operation of difference.

3. *Field (or algebra)*

Any class \mathcal{A} of subsets of X which is a ring and contains X is called a field. Thus a ring is a field if and only if it is closed under the operation of taking the complement. The class of all finite subsets of a space X is a ring, but is not a field unless X is finite. In \mathbf{R} the class of all bounded subsets is a ring but not a field.

4. *Sigma ring (σ -ring)*

A ring \mathcal{R} is called a σ -ring if it is closed under countable unions, i.e. if

$$A_i \in \mathcal{R} \quad (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{R}.$$

Now put $A = \bigcup_{i=1}^{\infty} A_i$ and use the identity $\bigcap_{i=1}^{\infty} A_i = A - \bigcup_{i=1}^{\infty} (A - A_i)$ to see that a σ -ring is also closed under countable intersections. Hence if \mathcal{R} is a σ -ring and $\{A_n\}$ is a sequence of sets from \mathcal{R} then $\limsup A_n$ and $\liminf A_n$ both belong to \mathcal{R} .

5. *Sigma-field (σ -field, Borel field, σ -algebra)*

Any class \mathcal{F} of sets which contains the whole space X and is a σ -ring is called a σ -field. Alternatively, a σ -field can be defined as a field which is closed under countable unions. For any space X , the class of all countable subsets will be a σ -ring, but will only be a σ -field if X is countable.

6. *Monotone class*

Any class \mathcal{M} of subsets such that, for any monotone sequence $\{E_n\}$ of sets in \mathcal{M} we have $\lim E_n \in \mathcal{M}$ is called a monotone class. It is clear that a σ -ring is a monotone class, and any monotone class which is a ring is also a σ -ring since

$$E_i \in \mathcal{M} \Rightarrow \bigcup_{i=1}^n E_i \in \mathcal{M},$$

and $\bigcup_{i=1}^n E_i$ is monotone so that $\bigcup_{i=1}^{\infty} E_i = \lim_{i \rightarrow \infty} \bigcup_{i=1}^n E_i$ is in \mathcal{M} .

We now use the term z -class to denote any one of the types 2, 3, 4, 5, 6 above (but not a semi-ring), and we consider a collection of z -classes.

Lemma. If \mathcal{C}_α , for $\alpha \in I$ is a z -class, then $\mathcal{C} = \bigcap_{\alpha \in I} \mathcal{C}_\alpha$ is a z -class.

Proof. Each of these z -classes is defined in terms of closure with respect to specified operations. Since each \mathcal{C}_α is closed with respect to operations, the resulting subset will be in \mathcal{C}_α for all $\alpha \in I$ and therefore in \mathcal{C} , so that \mathcal{C} is also a z -class. \blacksquare

Note. The intersection of a collection of semi-rings need not be a semi-ring.

Theorem 1.3. Given any class \mathcal{C} of subsets of X there is a unique z -class \mathcal{S} containing \mathcal{C} such that, if \mathcal{Q} is any other z -class containing \mathcal{C} we must have $\mathcal{Q} \supset \mathcal{S}$.

Remark. The z -class \mathcal{S} obtained in this theorem is called the z -class generated by \mathcal{C} . It is clearly the smallest z -class of subsets which contains \mathcal{C} .

Proof. The class of all subsets of X is a z -class containing \mathcal{C} . Put $\mathcal{S} = \bigcap \{ \mathcal{Q} : \mathcal{Q} \supset \mathcal{C} \text{ and } \mathcal{Q} \text{ is a } z\text{-class} \}$. Then \mathcal{S} is a z -class by the lemma and it clearly satisfies the conditions of the theorem. \blacksquare

In certain special cases one can specify the nature of the z -class generated by a given class.

Theorem 1.4. *The ring $\mathcal{R}(\mathcal{S})$ generated by a semi-ring \mathcal{S} consists precisely of the sets which can be expressed in the form*

$$E = \bigcup_{k=1}^n A_k$$

of a finite disjoint union of sets of \mathcal{S} .

Proof. (i) The ring $\mathcal{R}(\mathcal{S})$ certainly must contain all sets of this form, since it has to be closed under finite unions.

(ii) To see that the system \mathcal{Q} of sets of this type form a ring suppose

$$A = \bigcup_{k=1}^n A_k, \quad B = \bigcup_{k=1}^m B_k$$

and put $C_{ij} = A_i \cap B_j \in \mathcal{S}$. Then since the sets C_{ij} are disjoint and

$$A \cap B = \bigcup_{i=1}^n \bigcup_{j=1}^m C_{ij}$$

the system \mathcal{Q} is closed under intersections. Now from the definition of a semi-ring, an induction argument shows that

$$A_i = \bigcup_{i=1}^m C_{ij} \cup \bigcup_{k=1}^{r_i} D_{ik}, \quad (i = 1, \dots, n);$$

$$B_j = \bigcup_{i=1}^n C_{ij} \cup \bigcup_{k=1}^{s_j} E_{kj}, \quad (j = 1, 2, \dots, m);$$

where the finite sequences $\{D_{ik}\} (k = 1, \dots, r_i)$ and $\{E_{kj}\} (k = 1, \dots, s_j)$ consist of disjoint sets in \mathcal{S} . It follows now that

$$A \triangle B = \bigcup_{i=1}^n \left(\bigcup_{k=1}^{r_i} D_{ik} \right) \cup \bigcup_{j=1}^m \left(\bigcup_{k=1}^{s_j} E_{kj} \right)$$

so that the system \mathcal{Q} is also closed under the operation of taking the symmetric difference. \blacksquare

Example. We have already seen that \mathcal{P} , the class of intervals $(a, b]$ in \mathbf{R} , is a semi-ring. The generated ring is the class \mathcal{C} of finite unions of disjoint half-open intervals. \mathcal{C} is called the class of *elemen-*

itary figures in \mathbf{R} . Similarly, the elementary figures in \mathbf{R}^n form the class \mathcal{E}^n of finite disjoint unions of half-open rectangles from \mathcal{P}^n .

The next theorem is often important in proving that a given class is a σ -ring.

Theorem 1.5. *If \mathcal{R} is any ring, the monotone class $\mathcal{M}(\mathcal{R})$ generated by \mathcal{R} is the same as the σ -ring $\mathcal{S}(\mathcal{R})$ generated by \mathcal{R} .*

Corollary. *Any monotone class \mathcal{M} which contains a ring \mathcal{R} contains the σ -ring $\mathcal{S}(\mathcal{R})$ generated by \mathcal{R} .*

Proof. Since a σ -ring is always a monotone class and $\mathcal{S}(\mathcal{R}) \supset \mathcal{R}$ we must have $\mathcal{S}(\mathcal{R}) \supset \mathcal{M}(\mathcal{R})$, denoted by \mathcal{M} .

Hence it is sufficient to show that \mathcal{M} is a σ -ring, and this will follow if we can prove that \mathcal{M} is a ring. For any set F , let $\mathcal{Q}(F)$ be the class of sets E for which $E - F$, $F - E$, $E \cup F$ are all in \mathcal{M} . Then if $\mathcal{Q}(F)$ is not empty it is easy to check that it is a monotone class. It is clear that $\mathcal{Q}(F) \supset \mathcal{R}$ for any $F \in \mathcal{R}$ so that $\mathcal{Q}(F) \supset \mathcal{M}$. Hence, $E \in \mathcal{M}$, $F \in \mathcal{R} \Rightarrow E \in \mathcal{Q}(F) \Rightarrow F \in \mathcal{Q}(E)$ by the symmetry of the definition of the class \mathcal{Q} , and it follows that $\mathcal{M} \subset \mathcal{Q}(E)$ since \mathcal{Q} is a monotone class. But the truth of this for every $E \in \mathcal{M}$ implies that \mathcal{M} is a ring. \blacksquare

In §1.2 we discussed mappings $f: X \rightarrow Y$ and saw that any such mapping induced a set mapping f^{-1} on the class of all subsets of Y . If f^{-1} is restricted to a special class \mathcal{C} of subsets in Y , then the image of \mathcal{C} under f^{-1} will be a class of subsets in X . The interesting thing is that the structure of the class \mathcal{C} is often preserved by such a mapping f^{-1} .

Theorem 1.6. *Suppose \mathcal{C} is a z -class of subsets of Y , $f: X \rightarrow Y$ is any mapping and $f^{-1}(\mathcal{C})$ denotes the class of subsets of X of the form $f^{-1}(E)$, $E \in \mathcal{C}$. Then $f^{-1}(\mathcal{C})$ is a z -class of subsets of X .*

Proof. It is easy to check that the mapping $f^{-1}: \mathcal{C} \rightarrow f^{-1}(\mathcal{C})$ commutes with each of the set operations union, symmetric difference, countable union and monotone limit. The closure of \mathcal{C} with respect to any of these operations therefore implies the closure of $f^{-1}(\mathcal{C})$ with respect to the same operation. \blacksquare

Exercises 1.5

1. Give an example of two semi-rings $\mathcal{S}_1, \mathcal{S}_2$ whose intersection is not a semi-ring.

2. Prove that any finite field is also a σ -field.

3. If \mathcal{R} is a ring of sets and we define operations $\odot =$ multiplication and $\oplus =$ addition by

$$E \odot F = E \cap F, \quad E \oplus F = E \Delta F$$

show that \mathcal{R} becomes a ring in the algebraic sense.

4. If \mathcal{R} is a ring and \mathcal{C} is the class of all subsets E of X such that either E or $(X - E)$ is in \mathcal{R} , show that \mathcal{C} is a field.

5. What is the ring $\mathcal{R}(\mathcal{C})$ generated by each of the following classes:

- (i) for a single fixed E , $\mathcal{C} = \{E\}$;
- (ii) for a single E , \mathcal{C} is class of all subsets of E ;
- (iii) \mathcal{C} is class of all sets with precisely 2 points?

6. Prove that if A is any subset of a space X , $A \neq \emptyset$ or X , then the σ -field $\mathcal{F}(A)$ generated by the set A is the class $\{\emptyset, A, X - A, X\}$.

7. If \mathcal{C} is a non-empty class of sets show that every set in the σ -ring generated by \mathcal{C} is a subset of a countable union of sets of \mathcal{C} .

8. For each of the following classes \mathcal{C} describe the σ -field, σ -ring and monotone class generated by \mathcal{C} .

(i) P is any permutation of the points of X , i.e. any transformation from X to itself which is $(1, 1)$ and onto, and \mathcal{C} is the class of subsets of X left invariant by P .

(ii) X is \mathbf{R}^3 , Euclidean 3-space, \mathcal{C} is the class of all cylinders in X , i.e. sets E such that $(x, y, z_1) \in E \Rightarrow (x, y, z_2) \in E$ for all $z_2 \in \mathbf{R}$.

(iii) $X = \mathbf{R}^2$, the plane, \mathcal{C} is class of all sets which are subsets of a countable union of horizontal lines.

9. Suppose X is the set of rational numbers in $0 \leq x \leq 1$, and let \mathcal{Q} be the set of intervals of the form $\{x \in X; a < x \leq b\}$ where $0 \leq a \leq b \leq 1$; $a, b \in X$. Show that \mathcal{Q} is a semi-ring and every set in \mathcal{Q} is either empty or infinite.

Show that the σ -ring generated by \mathcal{Q} contains all subsets of X .

10. Given a function $f: X \rightarrow Y$, and a class of subsets \mathcal{A} of X , $f(\mathcal{A})$ will denote the class of subsets of Y of the form $f(A)$, $A \in \mathcal{A}$.

What is the relation between $f(A - B)$ and $f(A) - f(B)$? Give an example in which $f(A \cap B) \neq f(A) \cap f(B)$. Show that it is possible to have a ring \mathcal{A} such that $f(\mathcal{A})$ is not a ring.

Give an example of a mapping $f: X \rightarrow Y$ and a semi-ring \mathcal{S} in Y such that $f^{-1}(\mathcal{S})$ is not a semi-ring. For any class \mathcal{N} of sets in Y show that

$$\mathcal{R}(f^{-1}(\mathcal{N})) = f^{-1}(\mathcal{R}(\mathcal{N})), \quad \mathcal{F}(f^{-1}(\mathcal{N})) = f^{-1}(\mathcal{F}(\mathcal{N})),$$

where $\mathcal{R}(\mathcal{C})$ is the ring generated by \mathcal{C} , and $\mathcal{F}(\mathcal{C})$ is the σ -field generated by \mathcal{C} .

1.6 Axiom of choice

Any non-empty set A contains at least one element x , and in the ordinary process of logic one can choose a particular element from a non-empty set. By using the principle of induction it follows that one can choose an element from each of a sequence of non-empty sets, but difficulty arises if one has to make the simultaneous choice of an

element from each set of a non-countable class \mathcal{C} . The assumption that such a choice is possible can be formulated in the following equivalent forms, known as the axiom of choice:

(1) Given a non-empty class \mathcal{C} of disjoint non-empty sets E_α , there is a set $G \subset \bigcup \{E_\alpha : E_\alpha \in \mathcal{C}\}$ such that $G \cap E_\alpha$ is a single point set for each $E_\alpha \in \mathcal{C}$.

(2) For a non-empty class \mathcal{C} of non-empty sets E_α , there is a function (called a choice function) $f: \mathcal{C} \rightarrow \bigcup \{E_\alpha : E_\alpha \in \mathcal{C}\}$ such that, for each E_α in \mathcal{C} , $f(E_\alpha) \in E_\alpha$.

The difficulty in proofs using the axiom of choice is that only the existence of a choice function is postulated, and if \mathcal{C} is uncountable, one has no information about its nature. However, we will find it convenient at times to use this axiom (or something equivalent). It has recently been shown that both the axiom of choice, and its negative, are consistent with the other axioms of set theory, so that one has to postulate this as an axiom. Although part of our theory will be valid without this axiom we will not trouble to discover how much and we will use the axiom of choice throughout when it is convenient.

There are a large number of other apparently different axioms which turn out to be logically equivalent to the axiom of choice. We will formulate just two of these, as they will be convenient later. Various new concepts will be needed before we can state them precisely.

Partial ordering

Suppose V is a set with elements a, b, \dots and $<$ is a relation defined between some but not necessarily all pairs $a, b \in V$ such that

- (i) $<$ is transitive, i.e. $a < b, b < c \Rightarrow a < c$;
- (ii) $<$ is reflexive, i.e. $a < a$ for all a in V ;
- (iii) $a < b, b < a \Rightarrow a = b$;

then V is said to be partially ordered by the relation $<$. V is said to be *simply* (or *totally*) *ordered* if,

- (iv) for each pair $a, b \in V$ at least one of $a < b, b < a$ is valid.

Any partial ordering in a set V induces automatically a partial ordering in every subset of V . If $W \subset V$ and the induced ordering in W is a simple ordering, then W is said to be a chain in V .

For example, in \mathbf{R} the usual \leq relation defines a total ordering of \mathbf{R} . However, in \mathbf{R}^2 , if we say $(x_1, y_1) < (x_2, y_2)$ if and only if $y_1 \leq y_2$ and $x_1 \leq x_2$ we have an example of a partial ordering which is not simple. A more useful example is the class \mathcal{C} of all subsets of a fixed set X with $A < B$ meaning $A \subset B$.

A chain W in a partially ordered set V is called a *maximal* chain if it is not possible to obtain a larger chain by the addition of an element in $(V - W)$. We can now state

Kuratowski's lemma. *Every partially ordered set V contains a maximal chain.*

This means that there is a totally ordered subset $W \subset V$ such that for every $x \in V - W$, there is some element $y \in W$ such that neither of $x < y$, $y < x$ is true.

For a partially ordered set V , the element a is said to be an *upper bound* for the subset $C \subset V$ if $c < a$ for every $c \in C$. The element a is said to be the least upper bound or *supremum* of the subset C if

- (i) a is an upper bound for C ;
- (ii) if b is an upper bound for C , then $a < b$.

It is easy to check that, in any partially ordered set V , it is impossible for two distinct elements a_1, a_2 to satisfy the above conditions (i), (ii) so that the supremum of a set C is unique when it exists. However, even when a set V is totally ordered, not all its subsets need have a supremum. With the usual ordering \mathbb{R} has the property that any non-empty subset C which is bounded above has a supremum (this is known as the *least upper bound axiom*), but \mathbb{Q} does not have this property.

Finally, we say that the element $m \in V$ is a maximal element of V if $m < a \Rightarrow m = a$. We can now state

Zorn's lemma. *If V is partially ordered and each chain W in V has a supremum, then V has a maximal element.*

Both Zorn's lemma and Kuratowski's lemma can be deduced from the axiom of choice,† but we will not give the details as these are complicated and outside the mainstream of our argument. However, the next theorem shows that, if we assume Zorn's lemma as an additional axiom, then both the axiom of choice and Kuratowski's lemma will be valid. This means that, in our subsequent work, we will assume whichever of these three results happens to be most convenient.

Theorem 1.6. *The statements of (A) Kuratowski's lemma and (B) Zorn's lemma are equivalent. Either of them implies (C) the axiom of choice.*

Proof. $(A) \Rightarrow (B)$. Suppose W is a maximal chain in V , then by the hypothesis of (B) there is a supremum m for W so that $a < m$ for all $a \in W$. If m is not a maximal element of V , then there is a $b \in V$ such that $b \neq m$ and $m < b$. Then b is not in W as this would imply $b < m$,

† For a discussion of these and other axioms equivalent to axiom of choice see, for example, J. L. Kelley, *General Topology* (Van Nostrand, 1955).

and $b = m$. Hence we may add b to the chain W and the new set obtained is still a chain. This would contradict the fact that W is a maximal chain.

$(B) \Rightarrow (A)$. The chains in V form a class \mathcal{C} which is partially ordered by inclusion. If now \mathcal{W} is a chain in \mathcal{C} with elements W (each of which is a chain in V), then the union $\bigcup \{W : W \in \mathcal{W}\}$ is a chain in V so that it is an element of \mathcal{C} which can only be the supremum of \mathcal{W} . Hence by hypothesis \mathcal{C} contains a maximal element, i.e. V contains a maximal chain.

$(B) \Rightarrow (C)$. We now suppose given a class \mathcal{N} of sets E_α . There are clearly some subsets (in fact any finite subset) $\mathcal{N}_1 \subset \mathcal{N}$ on which it is possible to define a choice function $g: \mathcal{N}_1 \rightarrow \bigcup \{E_\alpha : E_\alpha \in \mathcal{N}_1\}$ such that $g(E_\alpha) \in E_\alpha$. The set V of all such functions g is therefore non-empty and it is partially ordered if we say $g_1 < g_2$ if g_1 is defined on \mathcal{N}_1 , g_2 is defined on \mathcal{N}_2 , $\mathcal{N}_1 \subset \mathcal{N}_2$ and $g_1(E_\alpha) = g_2(E_\alpha)$ for $E_\alpha \in \mathcal{N}_1$ (i.e. g_2 is an extension of g_1). If now W is a chain in V containing functions g_i defined on \mathcal{N}_i , the supremum of W is the function defined on $\bigcup \mathcal{N}_i$ which has the value $g_i(E_\alpha)$ on any set $E_\alpha \in \mathcal{N}_i$. If we now assume (B) it follows that the set V has a maximal element f . Then this function f must be defined on all the sets E_α , for otherwise if f is not defined on E_1 we could choose an element $x_1 \in E_1$, put $f(E_1) = x_1$ and this would be a proper extension of f and therefore contradict the fact that f is maximal. **】**

Exercises 1.6

1. Show that Z is partially ordered if $a < b$ means that a is a divisor of b .

2. Suppose α is a decomposition of the non-empty set X into disjoint subsets; $X = \bigcup A_i$ all the A_i disjoint. Show that the collection of such decompositions is partially ordered if $\alpha < \beta$ means that β is a refinement of α , i.e. if β is the decomposition $X = \bigcup B_j$ then each B_j is a subset of some A_i .

3. A partially ordered set V is said to be *well ordered* if each non-empty subset $W \subset V$ has a least element, i.e. there is a $w_0 \in W$ such that $w_0 < w$ for all $w \in W$. Show that, if V is well ordered, then it is simply ordered, and by considering the natural ordering of \mathbf{R} show that there exist simply ordered sets which are not well ordered.

4. Assuming Zorn's lemma, show that any set X can be well ordered.

Hint. Consider the class \mathcal{C} of well ordered subsets $V \subset X$ with the partial ordering $V_1 < V_2$ if: (i) $V_1 \subset V_2$, (ii) the ordering in V_1 is the same as that induced by the ordering in V_2 , (iii) V_1 is an initial segment of V_2 in the sense that $a \in V_1, b \in V_2, b < a \Rightarrow b \in V_1$. Show that each chain in \mathcal{C} has a supremum and show that the maximal element V_0 in \mathcal{C} must be X .

2

POINT SET TOPOLOGY

2.1 Metric space

In the first chapter we were concerned with abstract sets where no structure in the set was assumed or used. In practice, most useful spaces do have a structure which can be described in terms of a class of subsets called 'open'. By far the most convenient method of obtaining this class of open sets is to quantify the notion of nearness for each pair of points in the space. A non-empty set X together with a 'distance' function $\rho: X \times X \rightarrow \mathbf{R}$ is said to form a *metric space* provided that

- (i) $\rho(y, x) = \rho(x, y) \geq 0$ for all $x, y \in X$;
- (ii) $\rho(x, y) = 0$ if and only if $x = y$;
- (iii) $\rho(x, y) \leq \rho(x, z) + \rho(y, z)$ for all $x, y, z \in X$.

The real number $\rho(x, y)$ should be thought of as the distance from x to y . Note that it is possible to deduce conditions (i), (ii) and (iii) from a smaller set of axioms: this has little point as all the conditions agree with the intuitive notion of distance. Condition (iii) for ρ is often called the triangle inequality because it says that the lengths of two sides of a triangle sum to at least that of the third. Condition (ii) ensures that ρ distinguishes distinct points of X , and (i) says that the distance from y to x is the same as the distance from x to y . When we speak of a metric space X we mean the set X together with a particular ρ satisfying conditions (i), (ii) and (iii) above. If there is any danger of ambiguity we will speak of the metric space (X, ρ) .

In the set \mathbf{R} of real numbers, it is not difficult to check (i), (ii) and (iii) for the usual distance function

$$\rho(x, y) = |x - y|,$$

and similarly in \mathbf{R}^n , $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$

$$\rho(x, y) = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{\frac{1}{2}}$$

(one always assumes the positive square root) the conditions for a metric are satisfied. Thus \mathbf{R} and \mathbf{R}^n are metric spaces with the usual Euclidean distance for ρ .

Open sphere

In a metric space (X, ρ) , if $x \in X$, $r > 0$, then

$$S(x, r) = \{y: \rho(x, y) < r\};$$

the set consisting of those points of X whose distance from x is less than r is called an open sphere (spherical neighbourhood) centre x , radius r . Clearly, in \mathbf{R}^n , $S(x, r)$ is the inside of the usual Euclidean n -sphere centre x , radius r (for $n = 2$, the 'sphere' is the interior of a circle while for $n = 1$ it reduces to the interval $(x - r, x + r)$).

Open set

A subset E of a metric space X is said to be *open* if, for each point x in E there is an $r > 0$ such that the open sphere $S(x, r) \subset E$. Note that the open spheres defined above are examples of open sets since

$$y \in S(x, r) \Rightarrow \rho(x, y) = r_1 < r,$$

so that, for $0 < r_2 \leq r - r_1$, $S(y, r_2) \subset S(x, r)$.

Theorem 2.1. *In a metric space X , the class \mathcal{G} of open sets satisfies*

- (i) $\emptyset, X \in \mathcal{G}$;
- (ii) $A_1, A_2, \dots, A_n \in \mathcal{G} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{G}$;
- (iii) $A_\alpha \in \mathcal{G}$ for α in $I \Rightarrow \bigcup_{\alpha \in I} A_\alpha \in \mathcal{G}$.

Proof. (i) Since any statement about the elements of \emptyset is true, $\emptyset \in \mathcal{G}$, and it is clear that $S(x, r) \subset X$ for any $x \in X$, $r > 0$ so certainly $X \in \mathcal{G}$.

(ii) If $x \in \bigcap_{i=1}^n A_i$, then $x \in A_i$ for $i = 1, \dots, n$ and each A_i is open so there are real numbers $r_i > 0$ for which $S(x, r_i) \subset A_i$. If we put $r = \min_{1 \leq i \leq n} r_i$, then $0 < r \leq r_i$ so that $S(x, r) \subset S(x, r_i) \subset A_i$ for $i = 1, \dots, n$; and $S(x, r) \subset \bigcap_{i=1}^n A_i$.

(iii) For any $x \in \bigcup_{\alpha \in I} A_\alpha$, there must be a particular α in I such that $x \in A_\alpha$. Since this A_α is open, there is an $r > 0$ such that

$$S(x, r) \subset A_\alpha \subset \bigcup_{\alpha \in I} A_\alpha. \blacksquare$$

Remark. The condition (ii) says that \mathcal{G} is closed for finite intersections, while (iii) says it is closed under arbitrary unions. One

cannot extend (ii) to give closure for infinite intersections for, in \mathbf{R} the intervals $(0, 1 + (1/n))$ are open sets, but

$$\bigcap_{n=1}^{\infty} \left(0, 1 + \frac{1}{n}\right) = \{x: 0 < x \leq 1\} = (0, 1]$$

is not open as it contains no open sphere centre 1.

It is more general to start with a set X and a class \mathcal{G} of subsets of X satisfying (i), (ii), (iii) of theorem 2.1 and to call these 'the open sets' in X . Such a class \mathcal{G} and set X are said to form a *topological space*, and \mathcal{G} is said to determine the topology in X . A topological space (X, \mathcal{G}) is said to be *metrisable* if there is a distance function ρ defined on it which determines the class \mathcal{G} for its open sets. Most topological spaces (X, \mathcal{G}) of interest satisfy the rather weak conditions which are sufficient to ensure metrisability, so that little is lost by assuming in the first place that we have a metric space (X, ρ) . Of course two different metrics ρ_1, ρ_2 on a set X may define the same class \mathcal{G} of open sets, so that even when a topological space is metrisable, the metric ρ is not uniquely determined—see exercise 2.4 (1).

In this chapter we will define most of the further concepts which depend on the topology of X in terms of the class \mathcal{G} of open sets in X : this means that the definitions will make sense either in a metric space (X, ρ) or in a topological space (X, \mathcal{G}) . However, when it simplifies the proof, we will assume that X has a metric ρ determining \mathcal{G} and use this metric, so that some theorems will be stated and proved for metric spaces even though they are true more generally.

Closed set

A subset E of X is said to be closed if $(X - E)$ is open. If we apply this definition, with de Morgan's laws, to the conditions (i), (ii), (iii) of theorem 2.1 satisfied by the class of open sets, we see that the class \mathcal{C} of closed sets satisfies

$$(i) \quad \emptyset, X \in \mathcal{C};$$

$$(ii) \quad A_1, A_2, \dots, A_n \in \mathcal{C} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{C};$$

$$(iii) \quad A_\alpha \in \mathcal{C}, \alpha \text{ in } I \Rightarrow \bigcap_{\alpha \in I} A_\alpha \in \mathcal{C};$$

so that the class \mathcal{C} is closed for finite unions and arbitrary intersections.

In a metric space (X, ρ) , for $x \in X, r > 0$ the set

$$\bar{S}(x, r) = \{y: \rho(x, y) \leq r\}$$

is called the *closed sphere* centre x , radius r . It is always a closed set according to our definition for

$$y \in G = X - \bar{S}(x, r) \Rightarrow \rho(x, y) = r_1 > r$$

so that $S(y, r_2) \subset G$ for $0 < r_2 \leq r_1 - r$.

Neighbourhood

In a topological space (X, \mathcal{G}) , any open set containing $x \in X$ is said to be a neighbourhood of x .

Limit point of a set

Given a subset E of X , a point $x \in X$ is said to be a limit point (or point of accumulation) of E if every neighbourhood of x contains a point of E other than x . Note that the point x may or may not be in E . In a metric space it is easy to see that x is a limit point of E only if every neighbourhood N of x contains infinitely many points of E : for, if N contains only the points x_1, x_2, \dots, x_n of E (all different from x), then $S(x, r)$ where $r = \min_{1 \leq i \leq n} \rho(x, x_i)$ is a neighbourhood of x which contains no point of E other than x .

Lemma. A set $E \subset X$ is closed if and only if E contains all its limit points.

Proof. Suppose E is closed, then $X - E = G$ is open, so that if $x \in G$ there is a neighbourhood N of x with $N \subset G$. This means that N contains no point of E so that x is not a limit point of E . Conversely, if E is a set which contains its limit points and $x \in G = X - E$, then x is not a limit point of E so there is a neighbourhood N_x of x containing no point of E . Since N_x is open, so is $H = \bigcup_{x \in G} N_x$. But $N_x \subset G$ for all $x \in G$ so $H \subset G$, and every point x of G is in the corresponding N_x so $H \supset G$. Thus $H = G$ and G is open. \blacksquare

Closure

For any set $E \subset X$, the closure of E , denoted by \bar{E} , is the intersection of all the closed subsets of X which contain E . It is immediate that \bar{E} is a closed set, and $E = \bar{E}$ if and only if E is closed. Further since a closed set contains its limit points, \bar{E} must contain all the limit points of E : in fact

$$\bar{E} = E \cup E',$$

where E' is the set of limit points of E , known as the derived set of E ; for if $x \notin E \cup E'$, there is a neighbourhood N of x which contains no point of $E \cup E'$ so that $(X - N)$ is a closed set containing E and $x \notin \bar{E}$.

Limit of a sequence

Given a sequence $\{x_i\}$ of points in a metric space (X, ρ) we say that the sequence converges to the point $x \in X$ if each neighbourhood of x contains all but a finite number of points of the sequence. Thus $\{x_i\}$ converges to x if given $\epsilon > 0$, there is an integer N such that

$$i \geq N \Rightarrow \rho(x, x_i) < \epsilon.$$

We then write $x = \lim_{i \rightarrow \infty} x_i$ or $x = \lim x_i$ and say that x is the limit of the sequence $\{x_i\}$. Note that, in a metric space, the limit of a sequence is unique—see exercise 2.1 (7).

In a metric space X , given a point x and a set E , the *distance from x to E* , denoted by $d(x, E)$ is defined by

$$d(x, E) = \inf\{\rho(x, y) : y \in E\}.$$

This is always defined since $\{\rho(x, y) : y \in E\}$ is a set of non-negative real numbers. If $E \subset S(x, r)$ for some open sphere, we say that E is bounded and define the diameter of E , denoted $\text{diam}(E)$, by

$$\text{diam}(E) = \sup\{\rho(x, y) : x, y \in E\}.$$

If E is not bounded then the set $\{\rho(x, y) : x, y \in E\}$ is not bounded above and we put $\text{diam}(E) = +\infty$. Note that $\text{diam}(E)$ is finite if and only if E is bounded. Finally, if E, F are two subsets of X , we define $d(E, F)$ by

$$\begin{aligned} d(E, F) &= \inf\{\rho(x, y) : x \in E, y \in F\} \\ &= \inf\{d(x, F) : x \in E\} \end{aligned}$$

and call $d(E, F)$ the distance from E to F . Note that if $E \cap F \neq \emptyset$, then $d(E, F) = 0$ but there is no converse to this statement.

Remark

Many readers will be familiar with the concepts of this section for \mathbf{R} and \mathbf{R}^2 . Usually the proofs given in these special cases can be generalised to a general metric space, and often even to a topological space. The reader who has difficulty in working in an abstract situation should visualise the argument in the plane \mathbf{R}^2 , but not use any of the special properties of \mathbf{R}^2 .

Exercises 2.1

1. In any set X , the class \mathcal{C} of all subsets of X satisfies the conditions for a class of open sets. Show that this topology can be generated by the metric

$$\rho(x, y) = 1 \quad \text{for } x \neq y, \quad x, y \in X.$$

(This is called the *discrete* topology in X .) At the opposite end of the scale, the *indiscrete* topology in X is that for which $\mathcal{G} = \{\emptyset, X\}$: in this case the space is non-metrisable if X contains at least two points.

2. Show that in a metric space X containing at least 2 points

(i) single point sets $\{x\}$ are closed, but they are open only if

$$d(x, X - \{x\}) > 0;$$

(ii) finite sets are closed;

(iii) any open set G is the union of the class of open spheres contained in G ;

(iv) any open set G is the union of the class of closed spheres contained in G .

3. In the topological space X , given a set $E \subset X$ a point x is said to be an *interior point* of E if there is a neighbourhood N of x with $N \subset E$. Prove

(i) the set E^0 consisting of the interior points of E is open;

(ii) E is open if and only if $E = E^0$.

4. In a topological space X show that

$$\overline{A_1 \cup A_2 \cup \dots \cup A_n} = \overline{A_1} \cup \overline{A_2} \cup \dots \cup \overline{A_n},$$

but this does not extend to arbitrary unions. Give an example in which $\overline{E \cap F} \neq \overline{E} \cap \overline{F}$.

5. If X is a 2-point space $\{x_1, x_2\}$, $\rho(x_1, x_2) = 1$ show that (X, ρ) is a metric space in which the closure $\overline{S(x_1, 1)}$ of the open sphere $S(x_1, 1)$ is not the same as the closed sphere $\bar{S}(x_1, 1)$. However, if X is a normed linear space (see §2.6) then $\bar{S}(x, r) = \overline{S(x, r)}$.

6. Suppose $A \subset \mathbb{R}$ and A is closed and bounded below. Show that the infimum of A is an element of A .

7. In a metric space, suppose $\{x_n\}$ is a sequence converging to x and E is the set of points in this sequence. Show

(i) every subsequence $\{x_{n_i}\}$ converges to x ;

(ii) either x is the only limit point of E or there is an integer N such that $x_n = x$ for $n \geq N$. Deduce that a sequence $\{x_n\}$ cannot converge to two different limit points.

8. Suppose E is the set of points in a sequence $\{x_n\}$ in a metric space and x is a limit point of E . Show there is a subsequence $\{x_{n_i}\}$ which converges to x .

9. For any set E in a metric space, show that

$$\bar{E} = \{x: d(x, E) = 0\}.$$

10. If E, F are subsets of a metric space X , $x, y \in X$, show

- (i) $\rho(x, y) \geq |d(x, E) - d(y, E)|$;
- (ii) $\rho(x, y) \leq d(x, E) + \text{diam}(E)$, if $y \in E$;
- (iii) $|\rho(x, y_1) - \rho(x, y_2)| \leq \text{diam}(E)$, if $y_1, y_2 \in E$;
- (iv) $d(E, F) = d(F, E)$.

Is d a metric on the space of subsets of X ?

11. In \mathbf{R} show that a bounded open set is uniquely expressible as a countable union of disjoint open intervals.

Hint. For each $x \in E$, put $a = \inf\{y: (y, x) \subset E\}$, $b = \sup\{y: (x, y) \subset E\}$; and show that the open interval $I_x = (a, b)$ contains x , is contained in E and is such that any open interval I satisfying $x \in I \subset E$ satisfies $I \subset I_x$. Deduce that for $x_1, x_2 \in E$, either $I_{x_1} = I_{x_2}$ or $I_{x_1} \cap I_{x_2} = \emptyset$, so that $E = \bigcup_{x \in E} I_x$ is a disjoint union. Enumerate the intervals I_x by considering those of length greater than $1/n$ ($n = 1, 2, \dots$).

In \mathbf{R}^n ($n \geq 2$) show that a bounded open set can be expressed as a disjoint union of a countable number of half-open rectangles in \mathcal{P}^n (but that this expression is never unique). Show that in general an open set in \mathbf{R}^n ($n \geq 2$) cannot be expressed as a disjoint union of open spheres, or of open rectangles.

2.2 Completeness and compactness

In a metric space (X, ρ) a sequence $\{x_n\}$ is said to be a *Cauchy sequence* if given $\epsilon > 0$, there is an integer N such that

$$n, m \geq N \Rightarrow \rho(x_n, x_m) < \epsilon.$$

It is immediate that any sequence $\{x_n\}$ in a metric space which converges to a point $x \in X$, is a Cauchy sequence.

Complete metric space

A metric space (X, ρ) is said to be complete if, for each Cauchy sequence $\{x_n\}$ in X , there is a point $x \in X$ such that $x = \lim x_n$. For example, the set \mathbf{Q} of rationals is a metric space with the usual distance, but it is not complete for $\sqrt{2} \notin \mathbf{Q}$, but one can easily define a Cauchy sequence $\{x_n\}$ of rationals which converges to $\sqrt{2}$ (in \mathbf{R}), and this sequence cannot converge to any rational. One of the important properties of the space \mathbf{R} is that it is complete. This property is equivalent to the assumption that, in the usual ordering, every non-void subset of \mathbf{R} which is bounded above has a supremum or least upper bound. We now give a proof of the completeness of \mathbf{R} by a method which will turn out to be useful in more complicated situations.

Lemma. *The space \mathbf{R} is complete.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in \mathbf{R} . Define a sequence of integers $\{n_i\}$ by $n_0 = 1$; if n_{i-1} is defined, let n_i be such that $n_i > n_{i-1}$ and $n, m \geq n_i \Rightarrow |x_n - x_m| < 1/2^i$. Then the series

$$\sum_{i=1}^{\infty} (x_{n_i} - x_{n_{i-1}})$$

is absolutely convergent, and therefore convergent,† say to y . But

$$\sum_{i=1}^p (x_{n_i} - x_{n_{i-1}}) = x_{n_p} - x_1,$$

so the subsequence $\{x_{n_p}\} (p = 1, 2, \dots)$ must converge to $x = x_1 + y$. Given $\epsilon > 0$, choose integers $P, N \geq n_P$ such that

$$p \geq P \Rightarrow |x - x_{n_p}| < \frac{1}{2}\epsilon,$$

$$n, m \geq N \Rightarrow |x_n - x_m| < \frac{1}{2}\epsilon.$$

Now, if $m \geq N$, we can take $n_p \geq N$ with $p \geq P$ to obtain

$$|x - x_m| \leq |x - x_{n_p}| + |x_{n_p} - x_m| < \epsilon,$$

so the sequence $\{x_n\}$ must converge to x . ■

Covering systems

A class \mathcal{C} of subsets of X is said to *cover* the set $E \subset X$ or form a *covering* for E , if $E \subset \bigcup \{S : S \in \mathcal{C}\}$. If all the sets of \mathcal{C} are open, and \mathcal{C} covers E , then we say that \mathcal{C} is an *open covering* of E .

Compact set

A subset E of X is said to be *compact* if, for each open covering \mathcal{C} of E , there is a finite subclass $\mathcal{C}_1 \subset \mathcal{C}$ such that \mathcal{C}_1 covers E . For example, the celebrated Heine–Borel theorem states that any finite closed interval $[a, b]$ is compact. Though this is proved in most elementary text-books we include a proof which starts from the least upper bound property.

Lemma. *If a, b are real numbers, the closed interval*

$$[a, b] = \{x : a \leq x \leq b\}$$

is compact.

† The fact that absolute convergence implies convergence is a consequence of the least upper bound axiom for \mathbf{R} , that is, it follows from the assumption that every non-empty subset of \mathbf{R} which is bounded above has a supremum. (See, for example, Burkill, *A First Course in Mathematical Analysis*, Cambridge, 1962.)

Proof. Let \mathcal{C} be any open cover of $[a, b]$ and let c be the supremum of the set of x in $[a, b]$ for which some finite subfamily $\mathcal{C}_1 \subset \mathcal{C}$ covers $[a, x]$. (This set is non-void since it contains a .) Choose a set $G \in \mathcal{C}$ with $c \in G$ and choose a point $d \in (a, c)$ such that the closed interval $[d, c] \subset G$. Then there is a finite subfamily covering $[a, d]$ and the addition of G to this family gives a finite subfamily covering $[a, c]$. But unless $c = b$, since G is open, we have covered by a finite subfamily the interval $[a, e]$ for some $e > c$ which contradicts the choice of c . \square

It is also possible to prove directly that any closed rectangle in \mathbf{R}^n is compact, but we will be able to deduce this from theorem 2.6. We can use this to show that, in \mathbf{R}^n , every closed bounded set is compact. This will follow from the following:

Lemma. *If E is compact, and F is a closed subset of E , then F is compact.*

Proof. Suppose \mathcal{C} is an open covering for F . Then \mathcal{C} , together with $(X - F)$, which is open, forms an open covering for E . This has a finite subcovering \mathcal{C}_1 of E and $\mathcal{C} \cap \mathcal{C}_1$ must be a finite subclass of \mathcal{C} which covers F . \square

It is not true in a general metric space that closed bounded subsets are compact—see exercise 2.2 (3). However, we can prove:

Lemma. *In a metric space X , every compact subset is closed and bounded.*

Proof. If E is not closed, there is a point $x_0 \in X$ which is a limit point of E but is not in E . For every $x \in E$, put $S_x = S(x, r)$ with $r = \frac{1}{2}\rho(x, x_0)$. Then the collection of all such open spheres covers E , but every finite subclass $S(x_1, r_1), \dots, S(x_n, r_n)$ has a void intersection with $S(x_0, r)$, where $r = \min_{1 \leq i \leq n} r_i$ and so cannot cover E , for $S(x_0, r)$ contains points of E . On the other hand, if E is not bounded, the class of open spheres of radius 1 and centres in E covers E , but no finite subclass can cover E . \square

Whenever the whole space X is compact, we talk of a compact space. The above lemma shows that \mathbf{R}^n is not compact because it is not bounded. A space X is said to be *locally compact* if every point x in X has a neighbourhood N such that \bar{N} is compact. It is clear that \mathbf{R}^n is locally compact.

There are various other properties in a topological space which are equivalent to compactness under suitable conditions.

Weierstrass property

A set E is said to have property (W) if every infinite subset of E has at least one limit point.

Finite intersection property

A class \mathcal{A} of subsets of E is said to have the finite intersection property if every finite intersection $\bigcap_{i=1}^n A_i$, where $A_i \in \mathcal{A}$, ($i = 1, 2, \dots, n$) is non-void.

Theorem 2.2. (i) *A closed subset E of a topological space X is compact if and only if every class \mathcal{A} of closed subsets of E with the finite intersection property has a non-void intersection.*

(ii) *In a metric space X , a subset E is compact if and only if it has property (W).*

Remark. In a general topological space, property (W) is equivalent to sequential compactness—the property that every countable open covering has a finite subcovering. A space in which arbitrary open coverings can be replaced by countable subcoverings is called *Lindelöf*. Thus in any Lindelöf space, property (W) is equivalent to compactness.

Proof. (i) Suppose E is compact and F_α , $\alpha \in I$ is a class of closed subsets of E with $\bigcap_{\alpha \in I} F_\alpha$ void. Then the class of sets $G_\alpha = X - F_\alpha$, $\alpha \in I$ is an open covering of E . Choose a finite subcovering

$$G_{\alpha_1}, G_{\alpha_2}, \dots, G_{\alpha_n}; \quad \text{then} \quad \bigcap_{i=1}^n F_{\alpha_i} = \emptyset$$

so that the class of sets F_α , $\alpha \in I$ has not got the finite intersection property. This proves that compactness implies that any class \mathcal{A} of closed subsets with the finite intersection property has a non-void intersection. Conversely suppose a closed set E is such that any class \mathcal{A} of closed subsets with the finite intersection property has a non-void intersection, and suppose G_α , $\alpha \in I$ is an open covering of E , so that $\bigcap_{\alpha \in I} (X - G_\alpha) = \emptyset$. If E is closed $E \cap (X - G_\alpha)$, $\alpha \in I$ is a family of closed subsets of E , so there must be a finite set $\alpha_1, \alpha_2, \dots, \alpha_n$ such that $\bigcap_{i=1}^n (X - G_{\alpha_i}) \cap E = \emptyset$. This means that $G_{\alpha_1}, \dots, G_{\alpha_n}$ form a finite subcovering for E .

(ii) Suppose first that E has not got property (W). Let A be an infinite subset of E with no limit point. If A is not enumerable, choose an enumerable subset B of A . Then B is closed and $(X - B)$ is open. Enumerate B as a sequence of distinct points $\{x_i\}$, and for each x_i choose a neighbourhood N_i which contains x_i but no other point of B .

Then the sequence $\{N_i\}$ together with the set $(X - B)$ form an open covering of E , which has no finite subcovering as none of the open sets N_i can be omitted without 'uncovering' the corresponding x_i .

Conversely suppose E has property (W) . Then there is a finite class \mathcal{C}_n of spherical neighbourhoods of radius $1/n$ which covers E ; for otherwise we could find an infinite subset of E all of whose points were distant more than $1/n$ apart and such a subset can have no limit point. Let $\mathcal{C} = \bigcup_{n=1}^{\infty} \mathcal{C}_n$; so that \mathcal{C} is countable. Now if G is any open set,

for each x in $G \cap E$ we can find a sphere $S \in \mathcal{C}$ containing x with $S \subset G$: for we can first choose $\delta > 0$ so that $S(x, \delta) \subset G$ and if $n > 2/\delta$, the sphere of \mathcal{C}_n which contains x will be contained in $S(x, \delta)$. Given any open covering \mathcal{D} of E , carry out the above process for each set D of \mathcal{D} which intersects E , and each x in $D \cap E$, and let $\mathcal{C}' \subset \mathcal{C}$ be the countable collection of open spheres obtained. For each $S \in \mathcal{C}'$, choose one set $D \in \mathcal{D}$ with $D \supset S$, and let \mathcal{D}' be the countable class of sets so obtained. Then, since \mathcal{C}' covers E , the class \mathcal{D}' is a countable subcovering. This means that, if we assume property (W) , open coverings can be replaced by a countable subcovering.

Now suppose $E \subset \bigcup_{i=1}^{\infty} G_i$ where the sets G_i are open. Then, if there is no finite subcovering, for each integer n we can find a point

$$x_n \in E - \bigcup_{i=1}^n G_i$$

and the sequence $\{x_n\}$ must form an infinite set, so that there is a limit point $x_0 \in E$. But $x_0 \in G_k$ for some k , and G_k is open and therefore is a neighbourhood of x_0 ; this means we can find an $n > k$ such that

$$x_n \in G_k \subset \bigcup_{i=1}^n G_i,$$

which is a contradiction. **]**

Compactification

Many operations can be carried out more easily in compact spaces than in non-compact spaces. Given a non-compact space X a useful trick is to enlarge it to a topological space $X^* \supset X$ which is compact and such that the system \mathcal{G} of open sets in X is obtained by taking the intersection $X \cap G$ with X of sets G which are open in X^* . This device is known as the compactification of X . For example, \mathbf{R} is not compact, but if we adjoin two points $+\infty, -\infty$ to form the space \mathbf{R}^* of *extended*

real numbers we can show that \mathbf{R}^* is compact if we call a set $E \subset \mathbf{R}^*$ open if $E \cap \mathbf{R}$ is open and

if $+\infty \in E$, there is a neighbourhood $\{x: a < x \leq +\infty\} \subset E$;

if $-\infty \in E$, there is a neighbourhood $\{x: -\infty \leq x < b\} \subset E$,

where $a, b \in \mathbf{R}$. Note that the extended real number system \mathbf{R}^* is simply ordered if we put $-\infty < x < +\infty$ for all $x \in \mathbf{R}$.

In general, a non-compact topological space X may be compactified in many different ways. The simplest method is to adjoin a single point ∞ (which can be thought of as a point at infinity) to give the space $X^* = X \cup \{\infty\}$ and say that a subset G of X^* is open if either

(i) $G \subset X$ and G is open in X ; or

(ii) $\infty \in G$ and $(X^* - G)$ is a closed compact subset of X . It is not difficult to verify that this collection of 'open' sets defines a topology in X^* in which X^* is compact. This process is called the *one-point compactification* of X . It is familiar in the theory of the complex plane, where it is usual to add a single point at infinity (with neighbourhoods of the form $|z| > R$) to make the resulting 'closed plane' compact. Note that, if G^* is the class of open sets in X^* , G is the class of sets of the form $X \cap E$, $E \in G^*$.

There are other, more sophisticated, methods of compactifying a topological space X , but we will not require these in the present book.

Exercises 2.2

1. If (X, ρ) is a compact metric space, show that

(i) X is complete;

(ii) for each $\epsilon > 0$, X can be covered by a finite class of open spheres of radius ϵ .

2. If (X, ρ) is a complete metric space which, for each $\epsilon > 0$, can be covered by a finite number of spheres of radius ϵ , show that X is compact.

3. The open interval $X = (0, 1) \subset \mathbf{R}$ with the usual metric

$$\rho(x_1, x_2) = |x_1 - x_2|$$

is a metric space. In (X, ρ) the set X is closed and bounded. Show that X is not compact (and therefore not complete by example 2).

4. Construct a covering of the closed interval $[0, 1]$ by a family of closed intervals such that there is no finite subcovering.

5. If A, B are compact subsets of a metric space X , show that there are points $x_0 \in A$, $y_0 \in B$ for which

$$\rho(x_0, y_0) = d(A, B).$$

Hint. Take sequences $\{x_i\}$ in A , $\{y_i\}$ in B with $\rho(x_i, y_i) < d(A, B) + 1/i$, and apply property (W) to find convergent subsequences.

6. Give the details of the proof that the process of adjoining a point ∞ used to give the one-point compactification does yield a compact set X^* . If this process is applied to a space X which is already compact, show that the one point set $\{\infty\}$ is then both open and closed in X^* .

2.3 Functions

In Chapter 1 we defined the notion of a function $f: X \rightarrow Y$. When X and Y are topological spaces it is natural to enquire how the function f is related to these topologies. In particular do points which are 'close' in X map into points which are close in Y ? We make this precise first for metric spaces.

Continuous function

If (X, ρ_X) , (Y, ρ_Y) are metric spaces, a function $f: X \rightarrow Y$ is said to be continuous at $x = a$ if, given $\epsilon > 0$ there is a $\delta > 0$ such that

$$\rho_X(x, a) < \delta \Rightarrow \rho_Y(f(x), f(a)) < \epsilon.$$

If $E \subset X$, we say f is continuous on E if it is continuous at each point of E . In particular $f: X \rightarrow Y$ is said to be continuous (or continuous on X) if it is continuous at each point of X .

Lemma. If (X, ρ_X) , (Y, ρ_Y) are metric spaces a function $f: X \rightarrow Y$ is continuous if and only if $f^{-1}(G)$ is an open set in X for each open set G in Y .

Proof. Suppose first that f is continuous and G is an open set in Y . If $f^{-1}(G)$ is void, then it is open. Otherwise, let $a \in f^{-1}(G)$, $f(a) \in G$ so that there is an $\epsilon > 0$ for which the sphere $S(f(a), \epsilon) \subset G$. But then we can find a $\delta > 0$ such that

$$\rho_X(x, a) < \delta \Rightarrow f(x) \in S(f(a), \epsilon) \subset G$$

so that the sphere $S(a, \delta) \subset f^{-1}(G)$.

Conversely consider f at a point a of X . For each $\epsilon > 0$, $S(f(a), \epsilon) = H$ is an open set in Y , so that if $f^{-1}(H)$ is open, we can find a $\delta > 0$ for which $S(a, \delta) \subset f^{-1}(H)$, that is such that

$$\rho_X(x, a) < \delta \Rightarrow \rho_Y(f(x), f(a)) < \epsilon. \blacksquare$$

If (X, \mathcal{G}) , (Y, \mathcal{H}) are topological spaces, the function $f: X \rightarrow Y$ is said to be *continuous* if $f^{-1}(H) \in \mathcal{G}$ for every H in \mathcal{H} . The lemma just proved shows that when the topologies in X , Y are determined by

metrics this definition agrees with the one first given for mappings from one metric space to another.

Now if $f: X \rightarrow Y$ is continuous and E is a closed subset of Y , it follows that $f^{-1}(E)$ is closed in X . One has to be careful about the implications in the reverse direction. In general, it is not true for a continuous $f: X \rightarrow Y$, that A open in $X \Rightarrow f(A)$ open in Y . There is one important result of this kind which is valid:

Theorem 2.3. *If $f: X \rightarrow Y$ is continuous, and A is a compact subset of X , then $f(A)$ is compact in Y .*

Proof. Suppose $G_\alpha, \alpha \in I$ forms an open covering of $f(A)$. Then $f^{-1}(G_\alpha)$ is open for each α and the class $f^{-1}(G_\alpha), \alpha \in I$ must cover A . Since A is compact, there is a finite subcovering $f^{-1}(G_1), \dots, f^{-1}(G_n)$ which covers A , and this implies that G_1, \dots, G_n cover $f(A)$. **】**

Corollary. *If $f: X \rightarrow \mathbf{R}$ is continuous, and A is compact, the set $f(A)$ is bounded and the function f attains its bounds on A at points in A .*

Proof. $f(A)$ is compact, and so it must be closed and bounded. Hence $\sup\{x: x \in f(A)\}$ and $\inf\{x: x \in f(A)\}$ exist and belong to the set $f(A)$. Hence there are points $x_1, x_2 \in A$ for which $f(A) \subset [f(x_1), f(x_2)]$. **】**

Remark. The reader will recognise this corollary as a generalisation of the elementary theorem that a continuous function $f: [a, b] \rightarrow \mathbf{R}$ is bounded and attains its bounds.

It is important to notice that, in a metric space X , the distance function $\rho(x, y)$ is continuous for each fixed y considered as a function from X to \mathbf{R} . Further, for a fixed set A , $d(x, A)$ defines a continuous function from X to \mathbf{R} since

$$\rho(x_1, x_2) \geq |d(x_1, A) - d(x_2, A)|.$$

This means that if E is compact, F is any set, the function $d(x, F)$ for $x \in E$ attains its lower bound so that there is an x_0 in E with

$$d(x_0, F) = d(E, F).$$

Now, if F is also compact

$$d(x_0, F) = \inf\{\rho(x_0, y): y \in F\}$$

is the lower bound on a compact set of another continuous function, so that there is a y_0 in F such that

$$d(x_0, F) = \rho(x_0, y_0) = d(E, F).$$

Thus we have proved a further corollary to theorem 2.3—which could have been proved by a different argument (see exercise 2.2 (5)).

Corollary. If E, F are two compact subsets of a metric space (X, ρ) , there are points $x_0 \in E, y_0 \in F$ such that

$$\rho(x_0, y_0) = d(E, F).$$

Uniformly continuous function

A mapping $f: X \rightarrow Y$ from the metric space (X, ρ_X) to the metric space (Y, ρ_Y) is said to be uniformly continuous on the subset $A \subset X$ if given $\epsilon > 0$, there is a $\delta > 0$ for which

$$x, y \in A, \rho_X(x, y) < \delta \Rightarrow \rho_Y(f(x), f(y)) < \epsilon. \quad (2.3.1)$$

Clearly a function which is uniformly continuous on A is certainly continuous at each point of A , but the point of the condition (2.3.1) is that one can make $f(x)$ close to $f(y)$ in Y simply by making x close to y simultaneously for all $x, y \in A$. The choice of δ in (2.3.1) does not depend on x or y . In general, uniform continuity does not follow from continuity, but there is an important case in which it does:

Theorem 2.4. If X, Y are metric spaces, and $f: X \rightarrow Y$ is continuous on A where A is a compact subset of X , then f is uniformly continuous on A .

Proof. Given $\epsilon > 0$, for each $x \in A$, there is a $\delta_x > 0$ such that

$$\xi \in S(x, \delta_x) \cap A \Rightarrow f(\xi) \in S(f(x), \tfrac{1}{2}\epsilon).$$

For $x \in A$, the class of spheres $S_x = S(x, \tfrac{1}{2}\delta_x)$ form an open covering of A . Choose a finite subcovering S_{x_1}, \dots, S_{x_n} and put $\delta = \tfrac{1}{2} \min(\delta_{x_1}, \dots, \delta_{x_n})$. Then if $\rho_X(\xi, \eta) < \delta$, $\xi, \eta \in A$, there must be a sphere S_{x_i} which contains ξ , and $S(x_i, \delta_{x_i})$ will then contain η . This implies

$$\rho_Y(f(\xi), f(\eta)) \leq \rho_Y(f(\xi), f(x_i)) + \rho_Y(f(\eta), f(x_i)) < \epsilon. \blacksquare$$

Remark. The reader will recognize the above theorem as a generalisation of the result that a continuous function $f: [a, b] \rightarrow \mathbf{R}$ is uniformly continuous.

Exercises 2.3

1. Consider the function $f: (0, \infty) \rightarrow \mathbf{R}$ given by $f(x) = \min(1, 1/x)$, for $x > 0$.

Show that it is continuous. Find the image $f(E)$ of

- (i) the set $E = (\tfrac{1}{2}, 1)$;
- (ii) the set Z of positive integers;

(i) shows that E can be open, $f(E)$ not open; (ii) shows that E can be closed, $f(E)$ not closed.

2. The function

$$f: (0, 1) \rightarrow \mathbf{R} \text{ given by } f(x) = \frac{1}{x(1-x)}$$

is continuous on $(0, 1)$, but not bounded and not uniformly continuous, so theorems 2.3, 2.4 fail if the set is not closed. To see they also fail if the set is closed but not compact, examine

$$g: \mathbf{R} \rightarrow \mathbf{R} \text{ given by } g(x) = \exp(x).$$

3. In the argument of the proof to theorem 2.4 why could we not have put

$$\delta = \frac{1}{2} \inf \{ \delta_x : x \in A \}$$

before first restricting to a finite subset?

4. Suppose A is compact and $\{f_r\}$ is a monotone sequence of continuous functions $f_r: A \rightarrow \mathbf{R}$ converging to a continuous $f: A \rightarrow \mathbf{R}$. Show that the convergence must be uniform, and give an example to show that the condition that A be compact is essential.

5. Prove Lebesgue's covering lemma, which states that if \mathcal{C} is an open cover of a compact set A in a metric space (X, ρ) , then there is a $\delta > 0$, such that the sphere $S(x, \delta)$ is contained in a set of \mathcal{C} for each $x \in X$.

2.4 Cartesian products

We have already defined the direct product of two arbitrary sets X, Y as the set of ordered pairs (x, y) with $x \in X, y \in Y$. If (X, \mathcal{G}) (Y, \mathcal{H}) are topological spaces, then there is a natural method of defining a topology in $X \times Y$. Let \mathcal{A} be the class of *rectangle* sets $G \times H$ with $G \in \mathcal{G}, H \in \mathcal{H}$ and let \mathcal{C} be the class of sets in $X \times Y$ which are unions of sets in \mathcal{A} (finite or infinite unions): it is immediate that \mathcal{C} satisfies the conditions (i), (ii), (iii) of theorem 2.1. This class \mathcal{C} of 'open' sets in $X \times Y$ is said to define the *product topology*. This definition extends in an obvious manner to finite products $X_1 \times X_2 \times \dots \times X_n$, and it is also possible to extend it to an arbitrary product of topological spaces—though we will not have occasion to consider a topology for infinite product spaces.

Theorem 2.5. *If (X_i, ρ_i) ($i = 1, \dots, n$) are metric spaces then*

$$\rho((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1 \leq i \leq n} \rho_i(x_i, y_i)$$

defines a metric in the Cartesian product $X_1 \times \dots \times X_n$ which generates the product topology.

Proof. It is clear that $\rho(x, y) = 0$ if and only if $\rho_i(x_i, y_i) = 0$ for each i

which implies $x_i = y_i$, $1 \leq i \leq n$ or $x = y$. Thus in order to show that ρ is a metric it is sufficient to prove the triangle inequality. But

$$\rho_i(x_i, y_i) \leq \rho_i(x_i, z_i) + \rho_i(y_i, z_i) \quad (i = 1, \dots, n)$$

since the ρ_i are all metrics, so that

$$\begin{aligned} \max_{1 \leq i \leq n} \rho_i(x_i, y_i) &\leq \max_{1 \leq i \leq n} \{\rho_i(x_i, z_i) + \rho_i(y_i, z_i)\} \\ &\leq \max_{1 \leq i \leq n} \rho_i(x_i, z_i) + \max_{1 \leq i \leq n} \rho_i(y_i, z_i). \end{aligned}$$

Now in the product space, the open sphere centre x , radius r has the form

$$\begin{aligned} \{(y_1, \dots, y_n) : \rho_i(x_i, y_i) < r, \quad 1 \leq i \leq n\} \\ = S(x_1, r) \times S(x_2, r) \times \dots \times S(x_n, r) \end{aligned}$$

that is, it is the product of spheres in each of the component spaces. Thus the open spheres are open sets in the product topology and since every open set in the topology of a metric ρ is a union of the open spheres contained in it, each such set must be open in the product topology.

Conversely if G is an open set in the product topology and $x \in G$, there must be open sets $G_i \subset X_i$ such that $x \in G_1 \times \dots \times G_n \subset G$. Choose $r_i > 0$ such that $S(x_i, r_i) \subset G_i$ and put $r = \min_{1 \leq i \leq n} r_i$. Then

$r > 0$ and $S(x, r) \subset S(x_1, r_1) \times \dots \times S(x_n, r_n) \subset G$. Thus any set G open in the product topology is also open in the topology of the metric ρ . \blacksquare

Remark. The metric ρ defined in this theorem is by no means the only one which generates the product topology—see exercise 2.4 (1, 2).

Theorem 2.6. *If X, Y are compact topological spaces, then $X \times Y$ is compact in the product topology.*

Remark. The proof which follows extends immediately to finite Cartesian products of compact sets. Actually the theorem is true for arbitrary products, and in this more general form is due to Tychonoff.

Proof. Suppose first that $R_\alpha = G_\alpha \times H_\alpha$, $\alpha \in I$ is a covering of $X \times Y$ by open rectangles. Then if x_0 is a fixed point of X and I_{x_0} is the set of indices α for which $(x_0, y) \in G_\alpha \times H_\alpha$ for some $y \in Y$, the class H_α , $\alpha \in I_{x_0}$ forms an open covering of the compact set Y . Hence, there is a finite set $J_{x_0} \subset I_{x_0}$ such that R_α , $\alpha \in J_{x_0}$ covers the set $\{x_0\} \times Y$. But if we put $A_{x_0} = \bigcap_{\alpha \in J_{x_0}} G_\alpha$, A_{x_0} is open, contains x_0 , and the finite class R_α , $\alpha \in J_{x_0}$ must cover all of $A_{x_0} \times Y$. For each $x_0 \in X$, form such an open set A_{x_0} : the class of all sets of this form is an open covering of the

compact set X , and so has a finite subcovering $A_{x_1}, A_{x_2}, \dots, A_{x_n}$. It follows that $R_\alpha, \alpha \in \bigcup_{i=1}^n J_{x_i}$ is a finite subcovering of $X \times Y$.

It remains to show that, in testing for compactness, it is sufficient to consider coverings by open rectangles $G \times H$ with G open in X , H open in Y . Suppose then that every covering of $X \times Y$ by open rectangles has a finite subcovering, and consider an arbitrary open covering $G_\alpha, \alpha \in I$. Each point $(x, y) \in X \times Y$ is an element of an open set G_α and therefore there is an open rectangle $R_{x,y}$ with $(x, y) \in R_{x,y} \subset G_\alpha$. The class of open rectangles $R_{x,y}, (x, y) \in X \times Y$ clearly covers $X \times Y$ and so we can find a finite subcovering R_1, R_2, \dots, R_n . The corresponding sets G_1, \dots, G_n then form a finite subclass of the original covering class which covers $X \times Y$. \blacksquare

Corollary. Any bounded closed set in \mathbf{R}^n is compact.

Proof. The usual topology in \mathbf{R}^n is given by the distance function

$$\rho(x, y) = \left\{ \sum_{i=1}^n (x_i - y_i)^2 \right\}^{\frac{1}{2}},$$

while theorem 2.5 shows that the product topology is given by the distance function

$$\tau(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|.$$

Now $\tau(x, y) \leq \rho(x, y) \leq \sqrt{n} \tau(x, y)$ for all $x, y \in \mathbf{R}^n$ so that the topology of the usual metric ρ is the product topology.

If E is bounded, there is a real number K such that E is a subset of the Cartesian product of n intervals $[-K, K]$. Since each of these intervals is compact, the product is compact and therefore E is compact if it is closed. \blacksquare

Exercises 2.4

1. If ρ_1, ρ_2 are two metrics in X such that $c\rho_1(x, y) \leq \rho_2(x, y) \leq k\rho_1(x, y)$ for all $x, y \in X$ where $c > 0, k > 0$; show that ρ_1 and ρ_2 generate the same topology in X .

2. Show that, if $X \times Y$ has the product topology, then the projection function $p: X \times Y \rightarrow X$ defined by $p(x, y) = x$ is continuous.

In a space X , if $\mathcal{G}_1, \mathcal{G}_2$ are two collections of 'open' sets defining topologies and $\mathcal{G}_1 \supset \mathcal{G}_2$ we say that the topology given by \mathcal{G}_2 is *coarser* than that given by \mathcal{G}_1 . Show that the product topology in $X \times Y$ is the coarsest topology for which projections are continuous. (For an arbitrary Cartesian product $\prod_{\alpha \in I} X_\alpha$ of topological spaces $(X_\alpha, \mathcal{G}_\alpha)$ the projection p_β can be defined as $p_\beta(x_\alpha, \alpha \in I) = x_\beta \in X_\beta$ for any $\beta \in I$. One method of defining the product

topology in the Cartesian product space is to say that it is the coarsest topology for which each of the projections is continuous.)

3. Suppose $X \times Y$ has the product topology and $A \subset X$, $B \subset Y$. Show that

$$\overline{A \times B} = \overline{A} \times \overline{B},$$

and prove that the product of closed sets is closed.

2.5 Further types of subset

In a topological space X , a subset $E \subset X$ is said to be *nowhere dense* if the closure \overline{E} of E contains no non-void open set. If E is nowhere dense, and G is any non-void open set, the intersection $G \cap (X - \overline{E})$ is a non-void open subset disjoint from \overline{E} , and therefore from E . Conversely if \overline{E} contains a non-void open set H then every non-void open subset of H is a neighbourhood of each of its points, and therefore contains points of E . Thus E is nowhere dense if and only if every non-void open set in X contains a non-void open set disjoint from E .

Category

A subset $E \subset X$ is said to be of the *first category* (in X) if there is a sequence $\{E_n\}$ of nowhere dense subsets of X such that $E = \bigcup_{i=1}^{\infty} E_i$. A set $E \subset X$ which cannot be expressed as a countable union of nowhere dense sets is said to be of the *second category*.

Before proving that complete metric spaces are necessarily of the second category, it is convenient to prove a lemma which again generalises a well-known result in \mathbf{R} (about a decreasing sequence of closed intervals).

Lemma (Cantor). *In a complete metric space, given $\{A_n\}$ a decreasing sequence of non-empty closed sets such that $\text{diam}(A_n) \rightarrow 0$, $\bigcap_{n=1}^{\infty} A_n$ is a one point set.*

Proof. For each integer n , choose a point $x_n \in A_n$. Then given $\epsilon > 0$ we can choose N so that

$$n \geq N \Rightarrow \text{diam}(A_n) < \epsilon,$$

and, since $A_N \supset A_n$ for $n \geq N$,

$$n, m \geq N \Rightarrow x_n, x_m \in A_N \Rightarrow \rho(x_n, x_m) < \epsilon;$$

so that $\{x_n\}$ is a Cauchy sequence. Since the space is complete, there is a point x_0 such that $x_n \rightarrow x_0$ as $n \rightarrow \infty$. For each n , since A_n is

closed and $x_i \in A_n$ for $i \geq n$ we must have $x_0 \in A_n$ so that $x_0 \in \bigcap_{n=1}^{\infty} A_n$.

Now $\text{diam} \left(\bigcap_{n=1}^{\infty} A_n \right) \leq \text{diam} (A_n)$ for each n so that $\text{diam} \left(\bigcap_{n=1}^{\infty} A_n \right) = 0$ and the set $\bigcap_{n=1}^{\infty} A_n$ cannot contain more than one point. **■**

Theorem 2.7 (Baire). *Every complete metric space X is of the second category.*

Proof. It is sufficient to show that if $\{A_n\}$ is any sequence of nowhere dense sets there are points $x \in X - \bigcup_{n=1}^{\infty} A_n$. Starting with such a sequence $\{A_n\}$, since we can find a non-void open set disjoint from A_1 there is a sphere $S(x_1, r_1)$ with $0 < r_1 < 1$ such that $\bar{S}(x_1, r_1) \cap A_1 = \emptyset$. Suppose we have found spheres $S(x_1, r_1) \supset S(x_2, r_2) \supset \dots \supset S(x_{n-1}, r_{n-1})$ with $0 < r_{n-1} < 1/(n-1)$ such that $\bar{S}(x_i, r_i) \cap A_i = \emptyset$ for $1 \leq i \leq n-1$. Then $S(x_{n-1}, r_{n-1})$ is a non-void open set in X and A_n is nowhere dense. There must therefore be an open subset disjoint from A_n so that we can find a sphere $S(x_n, r_n)$ with

$$0 < r_n < 1/n, \quad S(x_n, r_n) \subset S(x_{n-1}, r_{n-1}) \quad \text{and} \quad \bar{S}(x_n, r_n) \cap A_n = \emptyset.$$

By the last lemma, there is a unique point $x_0 \in \bigcap_{n=1}^{\infty} \bar{S}(x_n, r_n)$ and this point x_0 cannot be in A_n for any integer n . This means that $X \neq \bigcup_{n=1}^{\infty} A_n$. **■**

Corollary. \mathbf{R}^n is of the second category. For $a < b$, the interval $[a, b] \subset \mathbf{R}$ is of the second category.

Dense subset

If A, E are subsets of a topological space X , we say that A is *dense in E* if $\bar{A} \supset E$. This means that any open set G which contains a point of E also contains a point of A . In particular A is dense in X if $\bar{A} = X$, that is, if every non-void open set contains a point of A .

Separable space

A topological space X is said to be separable if there is a countable set $E \subset X$ such that E is dense in X . This implies the existence of a sequence $\{x_n\}$ in X such that every non-void open subset contains a point of the sequence. It is immediate that \mathbf{R} and \mathbf{R}^n are separable as the set of points with rational coordinates is dense and countable. Further, every compact metric space X is separable, since X can be

covered by a finite class \mathcal{C}_n of open spheres of radius $1/n$ for $n = 1, 2, \dots$ and the (countable) set consisting of the centres of all these spheres is clearly dense in X .

Borel sets and Borelian sets

In any topological space X , we will call the σ -field \mathcal{B} generated by the open sets the class of Borel sets, and the σ -field \mathcal{K} generated by the compact sets the class of Borelian sets. (One must remark that some authors use the term Borel sets for \mathcal{K} .)

In a metric space the compact sets are closed, and therefore in \mathcal{B} so that $\mathcal{K} \subset \mathcal{B}$. If $X = \bigcup_{i=1}^{\infty} K_i$ is a countable union of compact sets (in this case we can say that X is σ -compact), then $\mathcal{K} = \mathcal{B}$. In order to prove this it is sufficient to show that each open set $G \in \mathcal{K}$: but if G is open, $E = X - G$ is closed and so $E \cap K_i$ is compact for each i and this implies $E = \bigcup_{i=1}^{\infty} E \cap K_i \in \mathcal{K}$ and $G = X - E \in \mathcal{K}$. Now Euclidean n -space \mathbf{R}^n is the union of the closed spheres $\bar{S}(0, k)$ ($k = 1, 2, \dots$) each of which is compact, so \mathbf{R}^n is σ -compact. This means that, in \mathbf{R}^n the Borel sets and the Borelian sets are the same.

Note that, by our definition, the class \mathcal{B}^n of Borel sets in \mathbf{R}^n is the σ -field generated by the open sets in \mathbf{R}^n . It is convenient to see that \mathcal{B}^n can also be obtained as the σ -field generated by a simpler class of sets.

Lemma. *The class \mathcal{P}^n of half-open intervals in \mathbf{R}^n generates the σ -field \mathcal{B}^n of Borel sets in \mathbf{R}^n .*

Proof. Let \mathcal{F}^n be the σ -field generated by \mathcal{P}^n . Each set in \mathcal{P}^n ,

$$\{(x_1, x_2, \dots, x_n): a_i < x_i \leq b_i, \quad i = 1, 2, \dots, n\}$$

can clearly be obtained as a countable intersection

$$\bigcap_{k=1}^{\infty} \left\{ (x_1, \dots, x_n): a_i < x_i < b_i + \frac{1}{k}, \quad i = 1, \dots, n \right\}$$

of open rectangles, and is therefore in \mathcal{B}^n . Hence $\mathcal{B}^n \supset \mathcal{P}^n$ so that $\mathcal{B}^n \supset \mathcal{F}^n$.

On the other hand, each open set G in \mathbf{R}^n is a union of those rectangles of \mathcal{P}^n whose boundary points a_i, b_i are all rational. Since there are only countably many such sets, each G is a countable union of sets in \mathcal{P}^n and so $\mathcal{P}^n \supset \mathcal{G}$. This implies that $\mathcal{F}^n \supset \mathcal{B}^n$. \square

It is sometimes useful to be able to describe sets which can be obtained from a given class \mathcal{C} by a countable operation. We say that

E is a \mathcal{C}_σ set if it is possible to find sets E_1, E_2, \dots in \mathcal{C} such that $E = \bigcup_{i=1}^{\infty} E_i$; and E is a \mathcal{C}_δ set if $E = \bigcap_{i=1}^{\infty} E_i$ for a sequence $\{E_n\}$ in \mathcal{C} . In particular, if \mathcal{G} is the class of open sets in a space X , we see that $\mathcal{B} \supset \mathcal{G}_\delta \supset \mathcal{G}$, $\mathcal{G}_\sigma = \mathcal{G}$. Similarly, if \mathcal{F} is the class of closed sets, $\mathcal{B} \supset \mathcal{F}_\sigma \supset \mathcal{F}$, and $\mathcal{F}_\delta = \mathcal{F}$.

Perfect set

A subset E of a topological space X is said to be perfect if E is closed, and each point of E is a limit point of E . For example, in \mathbf{R}^n , any closed sphere $\bar{S}(X, r)$, $r > 0$ is perfect and, in particular, the closed interval $[a, b]$ is perfect in \mathbf{R} for any $a < b$. It is obvious that finite sets in a metric space cannot be perfect. In fact more is true—see exercise 2.5 (7).

Exercises 2.5

1. Show that, in \mathbf{R}^n , any countable set is of the first category. Give a category argument for the existence of irrational numbers.
2. Show that the class \mathcal{N} of nowhere dense subsets of X is a ring, and the class \mathcal{C} containing all sets of the first category is the generated σ -ring.
3. Show that in a complete metric space, a set of the first category contains no non-empty open set. Deduce that every non-empty open set is of the second category.
4. If G is an open set in a topological space, prove that $(\bar{G} - G)$ is nowhere dense.
5. In \mathbf{R} show that the class \mathcal{P}_0 of half-open intervals with rational end-points generates the σ -field \mathcal{B} of all Borel sets. Similarly in \mathbf{R}^n , show that \mathcal{P}_0^n generates the Borel sets \mathcal{B}^n .
6. Show that a set E is perfect if and only if $E = E'$, where E' is the set of limit points of E .
7. Show that any non-empty perfect subset of a complete metric space is non-countable. *Hint.* Use theorem 2.7 and the fact that a closed subset of a complete metric space is complete.

2.6 Normed linear space

There are many abstract sets which have an algebraic structure as well as a topology. Thus if, in the set X there is a binary operation $+$ (called addition) and an operation in which elements of X can be

multiplied by elements of the real number field \mathbf{R} to give elements in X we say that X is a *real linear space* if for all $x, y, z \in X$, $a, b \in \mathbf{R}$;

- (i) $x + y = y + x$;
- (ii) $x + (y + z) = (x + y) + z$;
- (iii) $x + y = x + z \Rightarrow y = z$;
- (iv) $a(x + y) = ax + ay$;
- (v) $(a + b)x = ax + bx$;
- (vi) $a(bx) = (ab)x$;
- (vii) $1 \cdot x = x$.

It follows from these axioms that X has a unique zero element $0 = 0 \cdot y$ for all $y \in X$, and that subtraction can be defined in X by $x - y = x + (-1)y$.

In the present book we will only consider linear spaces over \mathbf{R} . Most of our results can be extended, though sometimes with a little difficulty, to linear spaces over the number field \mathbf{C} . We will not carry out this extension, nor do we consider any more general number fields.

It is immediate that \mathbf{R}^n is a real linear space with vector addition and scalar multiplication for the two operations. The properties of linear spaces are studied at length in elementary courses on linear algebra.† We will not require many of these, but will develop the properties of linear independence when they are needed in Chapter 8.

Norm

If in a real linear space X there is a function $n: X \rightarrow \mathbf{R}$ satisfying

- (i) $n(0) = 0$, $n(x) > 0$ if $x \neq 0$;
- (ii) $n(x + y) \leq n(x) + n(y)$ for all $x, y \in X$;
- (iii) $n(ax) = |a|n(x)$ for $a \in \mathbf{R}$, $x \in X$,

we say that X is a *normed linear space*. We will in this case use the usual notation $\|x\|$ for the value $n(x)$ of the norm function n at x .

In any normed linear space X ,

$$\rho(x, y) = \|x - y\| = \rho(x - y, 0)$$

defines a metric, and in the topology determined by this metric, the algebraic operations are continuous in the sense that

- (i) $x + y$ is continuous in the product topology of $X \times X$;
- (ii) ax is continuous in the product topology of $X \times \mathbf{R}$.

† See, for example, G. Birkoff and S. MacLane. *A Survey of Modern Algebra*, (Macmillan, 1941).

It follows in particular that

$$(iii) \ a \in \mathbf{R}, \lim x_n = 0 \Rightarrow \lim (ax_n) = 0;$$

$$(iv) \ x \in X, a_n \in R, \lim a_n = 0 \Rightarrow \lim (a_n x) = 0.$$

(The reader is advised to check (i)–(iv) using the axioms.)

Special normed linear spaces will be studied in Chapters 7 and 8. At this stage we consider a few important examples of such spaces and examine the topological structure imposed by the norm.

M. Consider the set of those functions $x: [0, 1] \rightarrow \mathbf{R}$ which are bounded. Define, for $t \in [0, 1]$

$$(x + y)(t) = x(t) + y(t),$$

$$(ax)(t) = ax(t)$$

and check that this makes **M** a linear space. If we put

$$\|x\| = \sup_{0 \leq t \leq 1} |x(t)|$$

it is not hard to check that the conditions for a norm are also satisfied, so we have a normed linear space.

C. The set of those functions $x: [0, 1] \rightarrow \mathbf{R}$ which are continuous is a subset **C** of **M**. Since this subset **C** is closed under the operations of addition and scalar multiplication, it must be a normed linear space with the same norm

$$\|x\| = \sup_{0 \leq t \leq 1} |x(t)|.$$

s. The set of all sequences of real numbers $\{x_i\}$ is a linear space if we put

$$\{x_i\} + \{y_i\} = \{x_i + y_i\},$$

$$a\{x_i\} = \{ax_i\}.$$

Since for x, y real we have

$$\frac{|x + y|}{1 + |x + y|} \leq \frac{|x|}{1 + |x|} + \frac{|y|}{1 + |y|}$$

it follows that $\rho(\{x_i\}, \{y_i\}) = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|x_i - y_i|}{1 + |x_i - y_i|}$

defines a metric in **s**.

m. This is the set of all bounded sequences of real numbers with the same linear structure as **s**. However this time it is more convenient to use the norm

$$\|\{x_i\}\| = \sup_i |x_i|,$$

to make **m** into a normed linear space.

c. This is the set of convergent sequences of real numbers with the same norm and linear structure as m .

Each of the above spaces has a topology defined by the norm. We now obtain a few of the topological properties of these spaces, leaving the reader to determine the remainder.

Lemma. *The space C is complete.*

Proof. If $\{x_n\}$ is a Cauchy sequence in C , then for each $t \in [0, 1]$ $\{x_n(t)\}$ is a Cauchy sequence in \mathbf{R} which must converge to a real number $x_0(t)$. For each $\epsilon > 0$, there is an integer N such that, if

$$y_m(t) = |x_N(t) - x_m(t)| \quad (m \geq N),$$

then $\|y_m\| < \frac{1}{2}\epsilon$; that is,

$$0 \leq y_m(t) < \frac{1}{2}\epsilon \quad \text{for each } t \text{ in } [0, 1].$$

If we now let $m \rightarrow \infty$, it follows that

$$|x_N(t) - x_0(t)| \leq \frac{1}{2}\epsilon \quad \text{for all } t \text{ in } [0, 1]$$

so that, if $n \geq N$, $t \in [0, 1]$

$$|x_n(t) - x_0(t)| \leq |x_N(t) - x_n(t)| + |x_N(t) - x_0(t)| < \epsilon$$

and $\|x_n - x_0\| \rightarrow 0$ as $n \rightarrow \infty$. This means that x_0 is the uniform limit of a sequence of continuous functions and must therefore be continuous; that is, $x_0 \in C$.]

Lemma. *The space M is not separable.*

Proof. For each $s \in [0, 1]$, let x_s be the function given by

$$x_s(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq s, \\ 1 & \text{for } s < t \leq 1. \end{cases}$$

Then if $r, s \in [0, 1]$ and $r \neq s$ we must have $\|x_r - x_s\| = 1$. Now any dense set in M has to contain a point y_s such that $\|y_s - x_s\| < \frac{1}{2}$ for each $s \in [0, 1]$; and we cannot have $y_r = y_s$ with $r \neq s$ for then

$$1 = \|x_r - x_s\| \leq \|x_r - y_r\| + \|y_r - y_s\| + \|y_s - x_s\| < 1.$$

This means that any set dense in M must contain at least c points, and therefore M is not separable.]

Lemma. *The space c is not locally compact.*

Proof. A metric space X is locally compact if, for each $z \in X$, there is an $\epsilon > 0$ such that the closed sphere $\bar{S}(z, \epsilon)$ is compact. Now put

$$x_i^k = \begin{cases} 1 & \text{for } i = k, \\ 0 & \text{for } i \neq k, \end{cases}$$

and for each integer k ,

$$x^k = \{x_i^k\} (i = 1, 2, \dots) \in \mathcal{C} \quad \text{and} \quad k \neq j \Rightarrow \|x^k - x^j\| = 1.$$

Given $z = \{z_i\} \in \mathcal{C}$ and $\epsilon > 0$, put

$$z^k = z + \epsilon x^k$$

and all the points z^k are in $\bar{S}(z, \epsilon)$. But

$$\|z^k - z^j\| = \epsilon \quad \text{if} \quad k \neq j$$

so that $\{z^k\} (k = 1, 2, \dots)$ forms an infinite set in $\bar{S}(z, \epsilon)$ with no limit point, and $\bar{S}(z, \epsilon)$ cannot be compact. **]**

Exercises 2.6

1. Show that s is bounded but not compact. If $x = \{x_i\} \in s$, and

$$E = \{y: |y_i| \leq |x_i|\},$$

show that E is compact in s , but show that s is not locally compact.

2. Show that each of the spaces M , \mathcal{C} , m , s is complete.

3. Show that each of the spaces C , \mathcal{C} , s is separable, but that m is not separable.

Hint for C. Consider the set of functions which take rational values at each of a finite set of rationals in $[0, 1]$ and are defined by linear interpolation between these points.

4. $C^*(X)$ denotes the space of functions $f: X \rightarrow \mathbf{R}$ which are continuous and bounded. Show that $C^*(\mathbf{R})$ is not separable by considering continuous functions which take the values $+1$, -1 on disjoint subsets Z_1 , Z_2 of the set Z of positive integers and are defined elsewhere by linear interpolation. (The distance between any two such functions is 2, and there are \mathfrak{c} of them.)

However, let \mathcal{D} be the subset of $C^*(\mathbf{R})$ consisting of those functions f for which $\lim_{x \rightarrow +\infty} f(x)$, $\lim_{x \rightarrow -\infty} f(x)$ both exist. Prove that \mathcal{D} is separable.

5. Let l_2 be the subset of s such that $\sum_{i=1}^{\infty} x_i^2$ converges. In the linear structure of s show that l_2 is a linear space and that

$$\|x\| = \sum_{i=1}^{\infty} x_i^2$$

defines a norm. In the topology of this norm show that l_2 is separable.

The subset $\{x: |x_i| < 1/i\}$ of l_2 is known as the Hilbert cube: prove it is compact.

Hint. Starting with an infinite sequence in the Hilbert cube pick a subsequence in which the first coordinate converges, then successive subsequences in which the 2nd, 3rd, ..., n th coordinate converges. Show that the sequence to which these coordinates converge is in the Hilbert cube and is a limit point of the original set.

2.7 Cantor set

We now digress briefly from the study of general situations and consider the definition and properties of a special subset of \mathbf{R} first considered by Cantor. This set and associated functions will be useful in the sequel to provide counter-examples to several conjectures which are plausible but false.

If we denote the open interval $((3r-2)/3^n, (3r-1)/3^n)$ by $E_{n,r}$, put

$$G_n = \bigcup_{r=1}^{3^{n-1}} E_{n,r} \quad G = \bigcup_{n=1}^{\infty} G_n;$$

it is clear that G is an open subset of $[0, 1]$. Its complement

$$C = [0, 1] - G$$

is called the Cantor set. From its definition C is closed.

Lemma. *The Cantor set C is nowhere dense and perfect.*

Proof. If we express points $x \in [0, 1]$ in the form

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} \quad (a_i = 0, 1, 2), \quad (2.7.1)$$

then the set G_n above is the set of x for which $a_n = 1$. Hence, the set C consists of precisely those real numbers which have a representation in the form (2.7.1) with each $a_i = 0$ or 2 . Given a point $x_1 \in C$, altering the n th term a_n (replacing $0, 2$ by $2, 0$ respectively) gives a new point x_2 in C such that

$$|x_1 - x_2| = 2 \cdot 3^{-n}.$$

This shows that every point of C is a limit point of C , so that C is perfect.

If H is any open set with $H \cap [0, 1]$ not void then $H \cap [0, 1]$ contains an open interval I of length $\delta > 0$. If $\delta > 3^{1-n}$, then I must contain an interval $E_{n,r}$ so that H contains an open set disjoint from C . This proves that C is nowhere dense. ■

From the above lemma and example 2.5 (7) we can deduce that C is not countable. However, one can prove that C must have the same cardinal as the continuum $[0, 1]$ by considering the following mapping:

$$\text{if } x \in [0, 1], \text{ put } x = \sum_{i=1}^{\infty} \frac{b_i}{2^i} \quad (b_i = 0 \text{ or } 1),$$

where the sequence $\{b_i\}$ does not satisfy $b_i = 1$ for $i \geq N$. Put

$$f(x) = \sum_{i=1}^{\infty} \frac{a_i}{3^i} \quad \begin{cases} a_i = 0 & \text{if } b_i = 0, \\ a_i = 2 & \text{if } b_i = 1. \end{cases}$$

Then $f: [0, 1] \rightarrow C$ is $(1, 1)$ and maps $[0, 1]$ onto a (proper) subset of C . Since $C \subset [0, 1]$ the cardinal of C is c .

We can think of f as a function on $[0, 1]$ to $[0, 1]$. It is easy to see that f is monotonic, that is,

$$x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$$

so that, for each $y \in [0, 1]$,

$$f^{-1}[0, y] = [0, z]. \quad (2.7.2)$$

If z is defined by (2.7.2), then we say that

$$z = g(y).$$

This defines $g: [0, 1] \rightarrow [0, 1]$ as a monotonic function which is clearly constant on each of the sets $E_{n,r}$. In fact

$$\frac{3r-2}{3^n} \leq y \leq \frac{3r-1}{3^n} \Rightarrow g(y) = \frac{2r-1}{2^n}.$$

The function g is continuous and monotonic increasing, for

$$0 \leq y_1 - y_2 < 3^{-n-1} \Rightarrow 0 \leq g(y_1) - g(y_2) < 2^{-n}.$$

Since the function g is constant in each $E_{n,r}$ it follows that it is differentiable with zero derivative at each point of G . One can easily see that g increases at each point of C —and in fact the ‘upper derivative’ at points of C is $+\infty$.

Note that there is nothing magical about the integer 3 used in the construction of C . Similar constructions using expansions to a different base will give sets with similar properties.

Exercises 2.7

1. If $x = \sum_{i=1}^{\infty} \frac{c_i}{10^i}$ ($c_i = 0, 1, \dots, 9$) is a decimal expansion of real numbers in $[0, 1]$ and T is the set of such x for which $c_i \neq 7$, show that T is perfect and nowhere dense.

2. Construct a set which is dense in $[0, 1]$ and yet the union of a countable class of nowhere dense perfect sets.

3. Show that the function $g: [0, 1] \rightarrow [0, 1]$ defined above satisfies a Lipschitz condition of order $\alpha = \log 2 / \log 3$, but not of any order $\beta > \alpha$. (A function $h: I \rightarrow \mathbf{R}$ is said to satisfy a Lipschitz condition of order α at $x_0 \in I$ if $|h(x) - h(x_0)| \leq K |x - x_0|^\alpha$ for $x \in I$ and some suitable $K \in \mathbf{R}$.)

SET FUNCTIONS

3.1 Types of set function

We consider only functions $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$, where \mathcal{C} is a non-empty class of sets. Thus μ is a rule which determines, for each $E \in \mathcal{C}$, a unique element $\mu(E)$ which is either a real number or $\pm\infty$. We always assume that \mathcal{C} contains the empty set \emptyset . \mathbf{R}^* denotes the compactification of the real number field \mathbf{R} by the addition of two points $+\infty$, $-\infty$, while \mathbf{R}^+ will denote the set of non-negative real numbers together with $+\infty$. It is not possible to arrange for \mathbf{R}^* to be an algebraic field extending \mathbf{R} , though we will preserve as many of the algebraic properties of \mathbf{R} as possible by adopting the convention that, for any $a \in \mathbf{R}$,

$$-\infty < a < +\infty, \quad -\infty \pm a = -\infty, \quad +\infty \pm a = +\infty;$$

$$\text{if } a > 0, \quad a(+\infty) = +\infty, \quad a(-\infty) = -\infty;$$

$$\text{if } a = 0, \quad a(\pm\infty) = 0;$$

$$\text{if } a < 0, \quad a(+\infty) = -\infty, \quad a(-\infty) = +\infty;$$

$$(+\infty)(+\infty) = +\infty, \quad (+\infty)(-\infty) = -\infty, \quad (-\infty)(-\infty) = +\infty.$$

Thus the operation of dividing by $(\pm\infty)$ is not allowed and

$$(+\infty) + (-\infty)$$

is not defined in \mathbf{R}^* . All these definitions are natural except for the convention $0(\pm\infty) = 0$.

Arbitrary set functions $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ are not of much interest. We adopt conditions on μ which correspond to our intuitive idea of mass for a physical object (we generalise the notion to allow negative masses). The first property we define corresponds to the notion that the mass of a pair of disjoint objects is the sum of the masses of the individual objects.

Additive set function

A set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is said to be (finitely) additive if

$$(i) \quad \mu(\emptyset) = 0,$$

(ii) for every finite collection E_1, E_2, \dots, E_n of disjoint sets of \mathcal{C} such that $\bigcup_{i=1}^n E_i \in \mathcal{C}$ we have

$$\mu\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mu(E_i). \quad (3.1.1)$$

In this definition condition (i) is almost redundant for it will be implied by (ii) provided there is at least one $E \in \mathcal{C}$ with $\mu(E)$ finite. Note also that we do not in the definition assume that \mathcal{C} is closed under finite disjoint unions so that, in testing a given set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ to see whether or not it is additive, we can only use sets $E_i \in \mathcal{C}$ which are disjoint and have their union in \mathcal{C} . However, the definition is taken to imply that the right-hand side of (3.1.1) has a unique meaning in \mathbf{R}^* so that in particular there are no sets $E, F \in \mathcal{C}$ such that $E \cap F = \emptyset$, $E \cup F \in \mathcal{C}$, $\mu(E) = +\infty$, $\mu(F) = -\infty$.

The natural domain of definition for an additive set function μ is a ring since, if \mathcal{C} is a ring,

$$E_i \in \mathcal{C} \quad (i = 1, 2, \dots, n) \Rightarrow \bigcup_{i=1}^n E_i \in \mathcal{C}.$$

For a ring \mathcal{C} it is worth noticing that $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is additive if and only if $\mu(\emptyset) = 0$ and

$$E, F \in \mathcal{C}, \quad E \cap F = \emptyset \Rightarrow \mu(E \cup F) = \mu(E) + \mu(F),$$

since in this case the general result (3.1.1) can be obtained from the result for two sets by a simple induction argument.

When \mathcal{C} is a finite class of sets it is easy to give examples of set functions defined on \mathcal{C} which are additive. We now give a number of less trivial examples which will be useful for illustrating our later definitions. In each case the reader should check that $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is additive.

Example 1. Ω any space with infinitely many points, \mathcal{C} the class of all subsets of Ω . Define μ by

$$\begin{aligned} \mu(E) &= \text{number of points in } E, \text{ if } E \text{ is finite;} \\ \mu(E) &= +\infty, \text{ if } E \text{ is infinite.} \end{aligned}$$

Example 2. Ω any topological space, \mathcal{C} the class of all subsets of Ω . Put

$$\begin{aligned} \mu(E) &= 0, \text{ if } E \text{ is of the first category in } \Omega; \\ \mu(E) &= +\infty, \text{ if } E \text{ is of the second category in } \Omega. \end{aligned}$$

Example 3. $\Omega = \mathbf{R}$, \mathcal{C} the class of all finite intervals of \mathbf{R} . For $E = [a, b]$ or $[a, b)$ or $(a, b]$ or (a, b) , put

$$\mu(E) = b - a.$$

Example 4. Ω is any space with at least two distinct points n, s ; \mathcal{C} is the class of all subsets of Ω .

$\mu(E) = 0$, if E contains neither or both of n, s ;

$\mu(E) = 1$, if E contains n but not s ;

$\mu(E) = -1$, if E contains s but not n .

Example 5. $\Omega = (0, 1]$, the set of real numbers x with $0 < x \leq 1$, \mathcal{C} the class of half-open intervals $(a, b]$ where $0 \leq a \leq b \leq 1$.

$$\mu(a, b] = b - a \quad \text{if } a \neq 0;$$

$$\mu(0, b] = +\infty.$$

Example 6. Ω is any infinite space, \mathcal{C} the class of all its subsets. Let $x_1, x_2, \dots, x_n, \dots$ be an enumerable sequence of distinct points of Ω , and suppose p_1, p_2, \dots is a sequence of real numbers such that $\sum_{i=1}^{\infty} p_i$ either converges absolutely or is properly divergent to $+\infty$ or $-\infty$ (the case $\sum p_i$ convergent, $\sum |p_i|$ divergent is not allowed: why?). Put

$$\mu(E) = \sum_{x_i \in E} p_i,$$

where the sum extends over all integers $i = 1, 2, \dots$ for which $x_i \in E$. Any set function which can be defined as in example 6 is called *discrete*. Note that example (4) can be thought of as a special case of example (6).

Although it is not sufficient to restrict our attention to set functions $\mu: \mathcal{C} \rightarrow \mathbf{R}$ which are finite valued, the condition of additivity which is usually assumed prevents μ from taking both the values $+\infty, -\infty$ at least when \mathcal{C} is a ring. This is one of the results in the next theorem.

Theorem 3.1. Suppose $\tau: \mathcal{C} \rightarrow \mathbf{R}^*$ is an additive set function defined on a ring \mathcal{C} and $E, F \in \mathcal{C}$. Then

(i) if $E \supset F$ and $\tau(F)$ is finite

$$\tau(E - F) = \tau(E) - \tau(F);$$

(ii) if $E \supset F$ and $\tau(F)$ is infinite

$$\tau(E) = \tau(F);$$

(iii) if $\tau(E) = +\infty$, then $\tau(F) \neq -\infty$.

Proof. (i) Since \mathcal{C} is a ring, $E - F \in \mathcal{C}$ and additivity implies, since $F \cap (E - F) = \emptyset$,

$$\tau(E) = \tau(E - F) + \tau(F). \quad (3.1.2)$$

Subtracting the finite real number $\tau(F)$ gives the result.

(ii) If $\tau(F) = +\infty$, then (3.1.2) can only have a meaning if $\tau(E - F) \neq -\infty$, and this implies $\tau(E) = +\infty$. The case $\tau(F) = -\infty$ is similar.

(iii) Since $E \cap F$, $E - F$, $F - E$ are disjoint sets of \mathcal{C}

$$\tau(E) = \tau(E \cap F) + \tau(E - F) = +\infty,$$

$$\tau(F) = \tau(E \cap F) + \tau(F - E) = -\infty$$

could only have meaning if $\tau(E \cap F)$ is finite. But this would imply

$$\tau(E - F) = +\infty, \tau(F - E) = -\infty, \text{ and then, since } E \Delta F \in \mathcal{C},$$

$$\tau(E \Delta F) = \tau(E - F) + \tau(F - E) = +\infty + (-\infty)$$

which is impossible. \blacksquare

Our definition of additivity means that for $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ to be additive any set $E_0 \in \mathcal{C}$ which can be split into a finite number of disjoint subsets in \mathcal{C} must be such that $\mu(E_0)$ is the same as the sum of the values of μ on the 'pieces'. We often want this to be true for a dissection of E_0 into a countably infinite collection of subsets in \mathcal{C} .

σ -additive set function

A set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is said to be σ -additive (sometimes called completely additive, or countably additive) if

(i) $\mu(\emptyset) = 0$,

(ii) for any disjoint sequence E_1, E_2, \dots of sets of \mathcal{C} such that

$$E = \bigcup_{i=1}^{\infty} E_i \in \mathcal{C},$$

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E_i). \quad (3.1.3)$$

As before the condition (i) is redundant if μ takes any finite values. Since we may assume that all but a finite number of the sequence $\{E_i\}$ are void it is clear that any set function which is σ -additive is also additive. To see that the converse is not true it is sufficient to consider example (5) on p. 53. Put

$$E = (0, 1], \quad E_n = \left(\frac{1}{n+1}, \frac{1}{n} \right] \quad (n = 1, 2, \dots);$$

then $\{E_n\}$ is a disjoint sequence in \mathcal{C} whose union E is in \mathcal{C} but

$$+\infty = \mu(E) \neq 1 = \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) = \sum_{n=1}^{\infty} \mu(E_n).$$

Notice further that even when \mathcal{C} is a ring it does not follow that

$$E_i \in \mathcal{C} \quad (i = 1, 2, \dots) \Rightarrow E = \bigcup E_i \in \mathcal{C};$$

so that in testing (3.1.3) we can only use those sets $E \in \mathcal{C}$ which can be split into a countable sequence of disjoint subsets in \mathcal{C} . In particular if \mathcal{C} is a finite class of sets then additivity for $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ implies σ -additivity. We also interpret (3.1.3) to mean that the right-hand side is uniquely defined and independent of the order of the sets E_i ;

thus if μ is σ -additive, and we have such a decomposition $E = \bigcup_{i=1}^{\infty} E_i$, we cannot have $\mu(E_i) = +\infty$, $\mu(E_j) = -\infty$, nor can the series in (3.1.3) converge conditionally.

It is easy to check that each of the set functions in examples (1), (2), (4), (6) on p. 52 is σ -additive, and the set function of example (3) is also σ -additive though the proof of this fact is non-trivial. This proof will be given in detail in § 3.4, as it is an essential step in the definition of Lebesgue measure in \mathbf{R} .

Measure

Any non-negative set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ which is σ -additive is called a measure on \mathcal{C} , ($\mathbf{R}^+ = \{x \in \mathbf{R}^*: x \geq 0\}$).

We should remark that there is not general agreement in the literature as to which set functions ought to be called measures. According to our definition the set functions in examples (1), (2) and (3) are measures, those in (4) and (6) are not because μ can take negative values while the set function in (5) is not because it is not σ -additive.

The natural domain of definition of a measure, or indeed of any σ -additive set function, is a σ -ring since then

$$E_i \in \mathcal{C} \quad (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{C}.$$

However, we will not restrict our consideration to σ -additive set functions already defined on a σ -ring.

Given a set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ where \mathcal{C} is a ring it is usually quite easy to check whether or not μ is additive for one only has to check (3.1.1) for $n = 2$. In order to check that it is also σ -additive it is useful to have a characterisation of σ -additive μ in terms of a continuity condition for monotone sequences of sets. Since we have seen already (theorem 3.1) that such set functions cannot attain both values $+\infty$, $-\infty$ there will be no loss of generality in assuming that

$$-\infty < \mu(E) \leq +\infty \quad \text{for all } E \in \mathcal{C}.$$

Continuity

Suppose \mathcal{R} is a ring and $\mu: \mathcal{R} \rightarrow \mathbf{R}^*$ is additive with $\mu(E) > -\infty$ for all $E \in \mathcal{R}$. Then for any $E \in \mathcal{R}$ we say that:

(i) μ is continuous from below at E if

$$\lim_{n \rightarrow \infty} \mu(E_n) = \mu(E) \quad (3.1.4)$$

for every monotone increasing sequence $\{E_n\}$ of sets in \mathcal{R} which converges to E ;

(ii) μ is continuous from above at E if (3.1.4) is satisfied for any monotone decreasing sequence $\{E_n\}$ in \mathcal{R} with limit E which is such that $\mu(E_n) < \infty$ for some n ;

(iii) μ is continuous at E if it is continuous at E from below and from above (when $E = \emptyset$ the first requirement is trivially satisfied).

Theorem 3.2. Suppose \mathcal{R} is a ring and $\mu: \mathcal{R} \rightarrow \mathbf{R}^*$ is additive with $\mu(E) > -\infty$ for all $E \in \mathcal{R}$.

(i) If μ is σ -additive, then μ is continuous at E for all $E \in \mathcal{R}$;

(ii) if μ is continuous from below at every set $E \in \mathcal{R}$, then μ is σ -additive;

(iii) if μ is finite and continuous from above at \emptyset , then μ is σ -additive.

Proof. (i) If $\mu(E_n) = +\infty$ for $n = N$ and $\{E_n\}$ is monotone increasing then $\mu(E) = +\infty$ and $\mu(E_n) = +\infty$ for $n \geq N$ by theorem 3.1 (ii), where $E = \lim E_n$. Thus in this case $\mu(E_n) \rightarrow \mu(E)$ as $n \rightarrow \infty$. On the other hand, if $\mu(E_n) < \infty$ for all n and $\{E_n\}$ increases to E , then

$$E = E_1 \cup \bigcup_{n=1}^{\infty} (E_{n+1} - E_n)$$

is a disjoint decomposition of E and

$$\begin{aligned} \mu(E) &= \mu(E_1) + \sum_{n=1}^{\infty} \mu(E_{n+1} - E_n) \\ &= \mu(E_1) + \lim_{N \rightarrow \infty} \sum_{n=1}^N \mu(E_{n+1} - E_n) = \lim_{N \rightarrow \infty} \mu(E_N), \end{aligned}$$

since μ is additive on the ring \mathcal{R} . Thus μ is continuous from below at E . Now suppose $\{E_n\}$ decreases to E and $\mu(E_N) < +\infty$. Put

$$F_n = E_N - E_n \quad \text{for } n \geq N.$$

Then, by theorem 3.1 (ii), $\mu(F_n) < \infty$ and the sequence $\{F_n\}$ is monotone increasing to $E_N - E$. Hence, as $n \rightarrow \infty$,

$$\mu(F_n) \rightarrow \mu(E_N - E) = \mu(E_N) - \mu(E).$$

But $\mu(F_n) = \mu(E_N) - \mu(E_n)$ so that $\mu(E_n) \rightarrow \mu(E)$ as $n \rightarrow \infty$, since $\mu(E_N)$ is finite, and μ is also continuous from above at E .

(ii) Suppose $E \in \mathcal{R}$, $E_i \in \mathcal{R}$ ($i = 1, 2, \dots$) are such that $E = \bigcup_{i=1}^{\infty} E_i$ and the sets E_i are disjoint. put

$$F_n = \bigcup_{i=1}^n E_i \in \mathcal{R} \quad (n = 1, 2, \dots),$$

and $\{F_n\}$ is a monotone increasing sequence of sets in \mathcal{R} which converges to $E \in \mathcal{R}$. If μ is continuous from below at E

$$\sum_{i=1}^n \mu(E_i) = \mu(F_n) \rightarrow \mu(E) \quad \text{as } n \rightarrow \infty$$

so that

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E_i),$$

and μ is σ -additive.

(iii) In the notation of (ii) put

$$G_n = E - F_n \in \mathcal{R} \quad (n = 1, 2, \dots).$$

Then $\{G_n\}$ is a monotone decreasing sequence converging to \emptyset and, for $n = 1, 2, \dots$

$$\mu(E) = \sum_{i=1}^n \mu(E_i) + \mu(G_n).$$

If μ is finite and continuous from above at \emptyset we must have

$$\mu(G_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

so that again

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E_i). \quad \blacksquare$$

Remark 1. In our definition of continuity from above we only require to have $\mu(E_n) \rightarrow \mu(E)$ for those sequences $\{E_n\}$ which decrease to E for which $\mu(E_n)$ is finite for some n . To see that we could not relax this finiteness condition, consider example (2) on p. 52 which we have already seen to be σ -additive with $\Omega = (0, 1)$. Then if

$$E_n = \left(0, \frac{1}{n}\right) \quad (n = 1, 2, \dots)$$

we have a sequence decreasing to \emptyset such that $\mu(E_n) = +\infty$ for all n since E_n is of the second category.

Remark 2. The condition that μ be finite cannot be omitted in theorem 3.2 (iii). Consider example (5) on p. 53 which we saw was additive but not σ -additive. Actually the class \mathcal{C} of sets on which μ is defined is a semi-ring rather than a ring, but its definition can easily

be extended to the ring of finite disjoint unions of sets in \mathcal{C} by using theorem 3.4. It is easy to check that it will remain continuous from above at \varnothing , but not σ -additive.

Part (iii) of theorem 3.2 will prove very useful in practice, especially for finite valued set functions $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ which are non-negative and additive on a ring \mathcal{R} . In order to prove that such a μ is a measure it is sufficient to show that, if $\{E_n\}$ is any sequence of sets in \mathcal{R} decreasing to \varnothing ,

$$\mu(E_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.1.5)$$

If (3.1.5) is false for some such sequence $\{E_n\}$ then, since $\mu(E_n)$ is monotone decreasing we must have

$$\mu(E_n) \rightarrow \delta > 0 \quad \text{as } n \rightarrow \infty. \quad (3.1.6)$$

If we can establish a contradiction by assuming (3.1.6), then (3.1.5) will be proved and we will have deduced that μ is a measure.

When we come to consider particular set functions one of our objectives will be to define μ on as large a class \mathcal{C} as possible. We will also want μ to be σ -additive. It would be desirable to define μ on the class of all subsets of Ω , but unfortunately this is not possible if Ω is not countable and μ is to have an interesting structure. In particular it has been shown, using the continuum hypothesis, that it is impossible to define a measure μ on all subsets of the real line such that (a) sets consisting of a single point have zero measure (this eliminates discrete set functions like examples (1), (4), (6) on pp. 52–3); (b) every set of infinite measure has a subset of finite positive measure (this eliminates example (2)); (c) the measure of the whole space is not zero. In practice the method used is to define μ with desired properties on a restricted class of sets \mathcal{C} (as in examples (3) or (5)) and then extend the definition to a larger class $\mathcal{D} \supset \mathcal{C}$.

Extension

Given two classes $\mathcal{C} \subset \mathcal{D}$ of subsets of Ω and set functions $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$, $\nu: \mathcal{D} \rightarrow \mathbf{R}^*$ we say that ν is an extension of μ if, for all $E \in \mathcal{C}$

$$\nu(E) = \mu(E);$$

under the same conditions we say that μ is the *restriction* of ν to \mathcal{C} .

It is sometimes appropriate (as in probability theory) to work with set functions μ which are finite. However, most of the theorems which can be proved for finite σ -additive set functions can also be obtained with a condition slightly weaker than finiteness.

σ -finite set function

A set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is said to be σ -finite if, for each $E \in \mathcal{C}$, there is a sequence of sets $C_i (i = 1, 2, \dots) \in \mathcal{C}$ such that $E \subset \bigcup_{i=1}^{\infty} C_i$ and $\mu(C_i)$ is finite for all i .

In our examples, on p. 52, the set functions in (3), (4), and (5) are all finite, (1) gives a σ -finite measure if and only if Ω is countable, (2) is not σ -finite if Ω is of the second category, (6) is finite if $\Sigma|p_i|$ converges and otherwise it is σ -finite.

Sometimes it is useful to relax the condition of additivity in order to be able to define μ on the class of all subsets. The most common example of this is in the concept of outer measure.

Outer measure

If \mathcal{C} is the class of all subsets of Ω , then $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ is called an outer measure on Ω if

- (i) $\mu(\emptyset) = 0$;
- (ii) μ is monotone in the sense that $E \subset F \Rightarrow \mu(E) \leq \mu(F)$;
- (iii) μ is countably subadditive in the sense that for any sequence $\{E_i\}$ of sets,

$$E \subset \bigcup_{i=1}^{\infty} E_i \Rightarrow \mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i). \quad (3.1.7)$$

Note that every measure on the class of all subsets of a space Ω is an outer measure on Ω . However, it is not difficult to give examples of outer measures which are not measures.

Example 7. Ω any space with more than one point. Put

$$\mu(\emptyset) = 0, \quad \mu(E) = 1 \quad \text{for all } E \neq \emptyset.$$

In this book we do not study the properties of outer measures for their own sake, but we will use them as a tool to extend the definition of measures.

Exercises 3.1

1. If $\Omega = [0, 1)$ and \mathcal{C} consists of the 6 sets

$$\emptyset, \quad \Omega, \quad [0, \tfrac{1}{2}), \quad [0, \tfrac{1}{4}), \quad [0, \tfrac{3}{4}), \quad [\tfrac{1}{4}, \tfrac{3}{4}),$$

$$\mu(\emptyset) = 0, \quad \mu[0, \tfrac{1}{4}) = 2, \quad \mu[0, \tfrac{1}{2}) = 2,$$

$$\mu[0, \tfrac{3}{4}) = 4, \quad \mu[\tfrac{1}{4}, \tfrac{3}{4}) = 2, \quad \mu(\Omega) = 4,$$

show that μ is additive on \mathcal{C} . Can μ be extended to an additive set function on the ring generated by \mathcal{C} ?

2. Show that if \mathcal{R} is any finite ring of subsets of Ω and μ is additive on \mathcal{R} then μ is σ -additive on \mathcal{R} .

3. A set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ is said to be monotone if $\mu(\emptyset) = 0$ and $E \subset F, E, F \in \mathcal{C} \Rightarrow \mu(E) \leq \mu(F)$. Show that monotone set functions are non-negative, and if \mathcal{C} is a ring, show that an additive non-negative set function is monotone. Of the set functions in examples (1)–(7), which are monotone?

4. \mathbf{Z} is the space of positive integers and $\sum_{n=1}^{\infty} a_n$ is a convergent series of positive terms.

If E is a finite subset of \mathbf{Z} , put $\tau(E) = \sum_{n \in E} a_n$; if E is an infinite subset of \mathbf{Z} , put $\tau(E) = +\infty$.

Show that τ is additive, but not σ -additive on the class of all subsets of \mathbf{Z} .

5. \mathbf{Z} is the space of positive integers; for $E \subset \mathbf{Z}$ let $r_n(E)$ be the number of integers in E which are not greater than n . Let \mathcal{C} be the class of subsets E for which

$$\lim_{n \rightarrow \infty} \frac{r_n(E)}{n} = \tau(E)$$

exists. Show that τ is finitely additive, but not σ -additive on \mathcal{C} , but that \mathcal{C} is not even a semi-ring.

6. If μ is finitely additive on a ring \mathcal{R} ; $E, F, G \in \mathcal{R}$ show

$$\begin{aligned} \mu(E) + \mu(F) &= \mu(E \cup F) + \mu(E \cap F), \\ \mu(E) + \mu(F) + \mu(G) + \mu(E \cap F \cap G) \\ &= \mu(E \cup F \cup G) + \mu(E \cap F) + \mu(F \cap G) + \mu(G \cap E). \end{aligned}$$

State and prove a relationship of this kind for n subsets of \mathcal{R} .

7. Suppose \mathcal{S} is a σ -ring of subsets of Ω , μ is a measure on \mathcal{S} . Show that the class of sets $E \in \mathcal{S}$ with $\mu(E)$ finite forms a ring, and the class with $\mu(E)$ σ -finite forms a σ -ring.

8. If E is a set in \mathcal{S} of σ -finite μ -measure (where μ is a measure on \mathcal{S}) and $\mathcal{D} \subset \mathcal{S}$ where \mathcal{D} is a class of disjoint subsets of E show that the subclass of those $D \in \mathcal{D}$ for which $\mu(D) > 0$ is countable.

9. State and prove a version of theorem 3.2 (i) for set functions μ defined on a semi-ring \mathcal{C} .

10. To show that the finiteness condition in the definition of ‘continuous from above’ in theorem 3.2 (i) cannot be relaxed, consider any infinite space Ω and put

$$\tau(E) = \text{number of points in } E, \text{ if } E \text{ is finite;}$$

$$\tau(E) = +\infty, \text{ if } E \text{ is infinite.}$$

Then τ is a measure on the class of all subsets of Ω , but for any sequence $\{E_n\}$ of infinite sets which decreases to \emptyset , we do not have $\lim \tau(E_n) = 0$.

11. Suppose \mathcal{P} is the semi-ring of half-open intervals $(a, b]$, Q is the set of rationals in $(0, 1]$ and \mathcal{P}_Q is the semi-ring of sets of the form $(a, b] \cap Q$. Put

$$\mu\{(a, b] \cap Q\} = b - a \quad \text{if } 0 \leq a \leq b \leq 1.$$

Show that μ is additive on \mathcal{P}_Q and is continuous above and below at every set in \mathcal{P}_Q , but is not σ -additive. This shows that theorem 3.2 (ii), (iii) is not true for semi-rings.

12. Show that if μ is an outer measure on Ω , E_0 any fixed subset then $\mu_0(E) = \mu(E \cap E_0)$ defines another outer measure on Ω .

13. Show that if μ, ν are outer measures on Ω , so is τ defined by

$$\tau(E) = \max[\mu(E), \nu(E)].$$

14. Suppose $\{\phi_n\}$ is a sequence of σ -additive set functions defined on a σ -ring \mathcal{S} and that $\lim \phi_n(E) = \phi(E)$ exists for all E in \mathcal{S} . Show that ϕ is finitely additive on \mathcal{S} . If either

(i) $\phi_n(E) \rightarrow \phi(E)$ uniformly on \mathcal{S} with $\phi(E) > -\infty$ for all $E \in \mathcal{S}$; or

(ii) $\phi_1(E) > -\infty$, $\phi_n(E)$ monotone increasing for all $E \in \mathcal{S}$;

show that ϕ is σ -additive on \mathcal{S} .

3.2 Hahn-Jordan decompositions

When discussing σ -additive set functions we will usually restrict our attention to the non-negative ones (which we call measures). The present section justifies this procedure by showing that, under reasonable conditions a 'signed' set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^*$ which is completely additive can be expressed as the difference of two measures. This means that properties of completely additive set functions can be deduced from the corresponding properties of measures. There are also versions of the decomposition theorem for finitely additive set functions, but we will not consider these.

We have already seen (theorem 3.1 (iii)) that an additive set function defined on a ring cannot take both the values $+\infty, -\infty$. If \mathcal{S} is a σ -ring and $\mu: \mathcal{S} \rightarrow \mathbf{R}^*$ is completely additive then for any sequence $\{E_i\}$ of disjoint sets in \mathcal{S} ,

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

Since $\bigcup_{i=1}^{\infty} E_i$ is independent of the order of the sets in the sequence, it follows that the series on the right-hand side must be either

absolutely convergent or properly divergent. In the case of example (6) the set function

$$\mu(E) = \sum_{x_i \in E} p_i$$

can be decomposed $\mu(E) = \mu_+(E) - \mu_-(E)$,

$$\text{where } \mu_+(E) = \sum_{x_i \in E} \max(0, p_i), \quad \mu_-(E) = - \sum_{x_i \in E} \min(0, p_i)$$

so that μ_+, μ_- are measures of which at least one is finite. Further if we put $P = \{x; \mu\{x\} > 0\}$, $N = \Omega - P$ we have $\mu_+E = \mu_+(P \cap E)$, $\mu_-E = -\mu(N \cap E)$ for all $E \subset \Omega$, so that the decomposition into the difference of two measures can also be obtained by splitting Ω into two subsets P, N such that μ is non-negative on every subset of P and non-positive on every subset of N . These two aspects of the decomposition are true in general, provided \mathcal{F} is a σ -field.

Theorem 3.3. *Given a completely additive $\tau: \mathcal{F} \rightarrow \mathbf{R}^*$ defined on a σ -field \mathcal{F} , there are measures τ_+ and τ_- defined on \mathcal{F} and subsets P, N in \mathcal{F} such that $P \cup N = \Omega$, $P \cap N = \emptyset$ and for each $E \in \mathcal{F}$,*

$$\tau_+(E) = \tau(E \cap P) \geq 0, \quad \tau_-(E) = -\tau(E \cap N) \geq 0,$$

$$\tau(E) = \tau_+(E) - \tau_-(E);$$

so that τ is the difference of two measures τ_+, τ_- on \mathcal{F} . At least one of τ_+, τ_- is finite and, if τ is finite or σ -finite so are both τ_+, τ_- .

Proof. Since τ can take at most one of the values $+\infty, -\infty$ we may assume without loss of generality that, for all $E \in \mathcal{F}$,

$$-\infty < \tau(E) \leq +\infty.$$

We first prove that, if $E \in \mathcal{F}$ and

$$\lambda(E) = \inf_{B \subset E, B \in \mathcal{F}} \tau(B), \tag{3.2.1}$$

then $\lambda(\Omega) \neq -\infty$. If this is false then there is a set $B_1 \in \mathcal{F}$ for which $\tau(B_1) < -1$. At least one of $\lambda(B_1), \lambda(\Omega - B_1)$ must be $-\infty$; since $\lambda(A \cup B) \geq \lambda(A) + \lambda(B)$ if A, B are disjoint sets of \mathcal{F} . Put A_1 equal to B_1 if $\lambda(B_1) = -\infty$ and $(\Omega - B_1)$ otherwise. Proceed by induction. For each positive integer n , choose $B_{n+1} \subset A_n$ such that

$$\tau(B_{n+1}) < -(n+1).$$

If $\lambda(B_{n+1}) = -\infty$, put $A_{n+1} = B_{n+1}$; otherwise put $A_{n+1} = A_n - B_{n+1}$. Then $\lambda(A_{n+1}) = -\infty$.

There are two possible cases:

- (i) for infinitely many integers n , $A_n = A_{n-1} - B_n$;
- (ii) for $n \geq n_0$, $A_n = B_n$.

In case (i) there is a subsequence $\{B_{n_i}\}$ of disjoint sets and

$$\tau\left(\bigcup_{i=1}^{\infty} B_{n_i}\right) = \sum_{i=1}^{\infty} \tau(B_{n_i}) \leq \sum_{i=1}^{\infty} -(n_i + 1) = -\infty,$$

so that τ takes the value

$$-\infty \quad \text{on} \quad E = \bigcup_{i=1}^{\infty} B_{n_i} \in \mathcal{F},$$

contrary to assumption. In case (ii), the set

$$E = \bigcap_{n=n_0}^{\infty} B_n \in \mathcal{F}$$

and since $\{B_n\}$ is then a decreasing sequence of sets we have

$$\tau(E) = \lim_{n \rightarrow \infty} \tau(B_n) = -\infty$$

again giving a contradiction.

Since $\tau(\emptyset) = 0$, $\lambda(\Omega) \leq 0$ so that $\lambda = \lambda(\Omega)$ is finite and we can find a sequence $\{C_n\}$ of sets in \mathcal{F} for which

$$\tau(C_n) < \lambda + 2^{-n}.$$

Now consider the set $C_n \cap C_{n+1}$. By noting that

$$C_n \cup C_{n+1} = (C_n - C_n \cap C_{n+1}) \cup (C_{n+1} - C_n \cap C_{n+1}) \cup (C_n \cap C_{n+1})$$

is a decomposition into disjoint sets, it follows that

$$\begin{aligned} \tau(C_n \cap C_{n+1}) &= \tau(C_n) + \tau(C_{n+1}) - \tau(C_n \cup C_{n+1}), \\ &< \lambda + 2^{-n} + \lambda + 2^{-n-1} - \lambda = \lambda + 2^{-n} + 2^{-n-1}. \end{aligned}$$

This argument can be repeated to the pair of sets $(C_n \cap C_{n+1})$ and C_{n+2} : by induction it can be proved that

$$\tau\left(\bigcap_{r=n}^p C_r\right) < \lambda + \sum_{r=n}^p 2^{-r} < \lambda + 2^{1-n}.$$

If we put $D_n = \bigcap_{r=n}^{\infty} C_r$, we have $D_n \in \mathcal{F}$ and, by theorem 3.2 (i),

$$\tau(D_n) \leq \lambda + 2^{1-n}.$$

But now $\{D_n\}$ is a monotone increasing sequence of sets in \mathcal{F} so that

$$N = \lim_{n \rightarrow \infty} D_n = \liminf_{n \rightarrow \infty} C_n \in \mathcal{F},$$

and

$$\tau(N) = \lambda,$$

by theorem 3.2 (i).

Finally, put $P = \Omega - N$. If $E \in \mathcal{F}$, $E \subset P$ we must have $\tau(E) \geq 0$ for, if $\tau(E) < 0$, then $\tau(E \cup N) = \tau(E) + \tau(N) < \lambda$. Also if $E \in \mathcal{F}$, $E \subset N$ we must have $\tau(E) \leq 0$ for, if $\tau(E) > 0$, then

$$\tau(N - E) = \tau(N) - \tau(E) < \lambda.$$

If we now put

$$\tau_+(E) = \tau(E \cap P), \quad \tau_-(E) = \tau(E \cap N),$$

it is clear that all the conditions of the theorem are satisfied. \square

Remark. It is usual to call the decomposition $\tau = \tau_+ - \tau_-$, of τ into the difference of two measures, the Jordan decomposition while the decomposition of Ω into positive and negative sets P and N is called the Hahn decomposition. It is not difficult to show that the Jordan decomposition is unique while the sets P , N are not uniquely determined by τ unless $\tau(E) \neq 0$ for all $E \in \mathcal{F}$ such that $\mu(E) \neq 0$ and $\mu(F) = 0$ or $\mu(E)$ for every $F \subset E$ with $F \in \mathcal{F}$. It is further clear that

$$\left. \begin{aligned} \tau_-(E) &= -\lambda(E), \\ \tau_+(E) &= \sup_{B \subset E, B \in \mathcal{F}} \tau(B) \end{aligned} \right\} \quad (3.2.2)$$

under the conditions of theorem 3.3, where $\lambda(E)$ is given by (3.2.1).

If one is given a σ -additive $\mu: \mathcal{S} \rightarrow \mathbf{R}^*$ defined on a σ -ring \mathcal{S} which is not a σ -field then it is not, in general, possible to obtain the Hahn decomposition, but the Jordan decomposition is still possible, using (3.2.1), (3.2.2) as the definition of τ_- , τ_+ .

Exercises 3.2

1. If $\phi: \mathcal{S} \rightarrow \mathbf{R}^*$ is σ -additive on a σ -ring \mathcal{S} , show that, for any $E \in \mathcal{S}$, there are sets $A \subset E$, $B \subset E$ with $A, B \in \mathcal{S}$ such that

$$\phi(A) = \inf_{C \subset E, C \in \mathcal{S}} \phi(C), \quad \phi(B) = \sup_{C \subset E, C \in \mathcal{S}} \phi(C).$$

2. Show that, given a (finitely) additive $\mu: \mathcal{R} \rightarrow \mathbf{R}^*$ defined on a σ -ring \mathcal{R} and taking finite values, there is a decomposition

$$\mu(E) = \mu_+(E) - \mu_-(E)$$

of μ into the difference of two non-negative additive set functions on \mathcal{R} .

3. The set $E_0 \in \mathcal{C}$ is said to be an *atom* of a set function $\phi: \mathcal{C} \rightarrow \mathbf{R}^*$ if $\phi(E_0) \neq 0$ and for every $E \subset E_0$, $E \in \mathcal{C}$; $\phi(E) = 0$ or $\phi(E_0)$. Write down the atoms of the set functions of examples (4) and (6) on page 53.

4. A set function $\phi: \mathcal{C} \rightarrow \mathbf{R}^*$ is said to be *non-atomic* if it has no atoms. Suppose $\phi: \mathcal{F} \rightarrow \mathbf{R}^*$ is σ -additive on the σ -field \mathcal{F} , non-atomic, and finite valued. Show that for any $A \in \mathcal{F}$, ϕ takes every real value between $-\phi_-(A)$ and $\phi_+(A)$ for some subset $E \subset A$.

3.3 Additive set functions on a ring

In order to simplify the arguments we now consider only non-negative set functions $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$. It is often possible, for a given ring \mathcal{R} to find a semi-ring $\mathcal{C} \subset \mathcal{R}$ such that \mathcal{R} is the ring generated by \mathcal{C} . We saw (see §1.5) that the sets of \mathcal{R} can then be expressed in terms of the sets of \mathcal{C} , so it is natural to ask whether in these circumstances a set function $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ can be extended to $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$. We now prove that, if μ is additive on \mathcal{C} , this is always possible and that the result is unique.

Theorem 3.4. *If $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ is a non-negative additive set function defined on a semi-ring \mathcal{C} , then there is a unique additive set function ν defined on the generated ring $\mathcal{R} = \mathcal{R}(\mathcal{C})$ such that ν is an extension of μ . ν is non-negative on \mathcal{R} , and is called the extension of μ from \mathcal{C} to $\mathcal{R}(\mathcal{C})$.*

Proof. Suppose A is any set of $\mathcal{R} = \mathcal{R}(\mathcal{C})$, then by theorem 1.4, $A = \bigcup_{k=1}^n E_k$ where the sets E_k are disjoint and $E_k \in \mathcal{C}$. Define

$$\nu(A) = \sum_{k=1}^n \mu(E_k). \quad (3.3.1)$$

Since for any $a, b \in \mathbf{R}^+$, $a + b$ is always defined, the right-hand side of (3.3.1) defines a number in \mathbf{R}^+ . ν is thus defined on \mathcal{R} provided we can show that (3.3.1) gives the same result for any two decompositions of A into disjoint subsets in \mathcal{C} .

Suppose

$$A = \bigcup_{k=1}^n E_k = \bigcup_{j=1}^m F_j,$$

where $F_j \in \mathcal{C}$ and are disjoint. Put $H_{kj} = E_k \cap F_j$. Then since \mathcal{C} is a semi-ring the sets $H_{kj} \in \mathcal{C}$, are disjoint and

$$E_k = \bigcup_{j=1}^m H_{kj} \quad (k = 1, 2, \dots, n);$$

$$F_j = \bigcup_{k=1}^n H_{kj} \quad (j = 1, 2, \dots, m);$$

so that, since μ is additive on \mathcal{C} ,

$$\sum_{k=1}^n \mu(E_k) = \sum_{k=1}^n \left(\sum_{j=1}^m \mu(H_{kj}) \right) = \sum_{j=1}^m \left(\sum_{k=1}^n \mu(H_{kj}) \right) = \sum_{j=1}^m \mu(F_j)$$

and it makes no difference which decomposition is used with (3.3.1) to define $\nu(A)$.

If A_1, A_2 are disjoint sets of \mathcal{R} , and

$$A_1 = \bigcup_{k=1}^n E_k, \quad A_2 = \bigcup_{i=1}^m F_i,$$

then $A_1 \cup A_2$ is a set of \mathcal{R} with a possible decomposition into disjoint subsets of \mathcal{C} given by

$$A_1 \cup A_2 = \bigcup_{k=1}^n E_k \cup \bigcup_{i=1}^m F_i.$$

Hence

$$\begin{aligned} \nu(A_1 \cup A_2) &= \sum_{k=1}^n \mu(E_k) + \sum_{i=1}^m \mu(F_i) \\ &= \nu(A_1) + \nu(A_2), \end{aligned}$$

since ν is uniquely determined by (3.3.1). Thus ν is finitely additive on \mathcal{R} (since \mathcal{R} is a ring). It is obvious that ν is non-negative.

Now let τ be any extension of μ from \mathcal{C} to \mathcal{R} which is additive. If $A \in \mathcal{R}$ and $A = \bigcup_{k=1}^n E_k$ is a decomposition into disjoint sets of \mathcal{C} ,

$$\begin{aligned} \tau(A) &= \sum_{k=1}^n \tau(E_k), \quad \text{since } \tau \text{ is additive;} \\ &= \sum_{k=1}^n \mu(E_k), \quad \text{since } \tau \text{ is an extension;} \\ &= \nu(A) \quad \text{by (3.3.1).} \end{aligned}$$

Thus ν is the unique additive extension of μ from \mathcal{C} to \mathcal{R} . **]**

If we start with a measure $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ on a semi-ring \mathcal{C} , then μ is clearly a non-negative finitely additive set function, and so possesses a unique additive extension to the generated ring \mathcal{R} . What can we say about this extension?

Theorem 3.5. *If $\mu: \mathcal{C} \rightarrow \mathbf{R}^+$ is a measure defined on a semi-ring \mathcal{C} , then the (unique) additive extension of μ to the generated ring $\mathcal{R}(\mathcal{C})$ is also a measure.*

Proof. In the last theorem we discovered the form of the unique additive extension ν of μ from \mathcal{C} to \mathcal{R} . It is sufficient to show that ν is σ -additive on \mathcal{R} . Suppose $E \in \mathcal{R}$, $E_k (k = 1, 2, \dots) \in \mathcal{R}$ and are disjoint, and $E = \bigcup_{k=1}^{\infty} E_k$.

Put $E = \bigcup_{r=1}^n A_r$, A_r disjoint sets of \mathcal{C} ;

$$E_k = \bigcup_{i=1}^{n_k} B_{ki}, \quad B_{ki} \text{ disjoint sets of } \mathcal{C}.$$

Put

$$C_{rki} = A_r \cap B_{ki} \quad (r = 1, 2, \dots, n; k = 1, 2, \dots; i = 1, 2, \dots, n_k),$$

then $\{C_{rki}\}$ forms a disjoint collection of sets in \mathcal{C} , and

$$A_r = \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{n_k} C_{rki}, \quad B_{ki} = \bigcup_{r=1}^n C_{rki}$$

are disjoint decompositions into sets of \mathcal{C} . Since μ is additive on \mathcal{C} ,

$$\mu(B_{ki}) = \sum_{r=1}^n \mu(C_{rki});$$

and since μ is σ -additive on \mathcal{C}

$$\mu(A_r) = \sum_{k=1}^{\infty} \sum_{i=1}^{n_k} \mu(C_{rki}).$$

Since the order of summation of double series of non-negative terms makes no difference, we have

$$\begin{aligned} \nu(E) &= \sum_{r=1}^n \mu(A_r) = \sum_{r=1}^n \left(\sum_{k=1}^{\infty} \sum_{i=1}^{n_k} \mu(C_{rki}) \right) \\ &= \sum_{k=1}^{\infty} \left(\sum_{i=1}^{n_k} \sum_{r=1}^n \mu(C_{rki}) \right) \\ &= \sum_{k=1}^{\infty} \sum_{i=1}^{n_k} \mu(B_{ki}) = \sum_{k=1}^{\infty} \nu(E_k). \blacksquare \end{aligned}$$

The above theorem gives one method of obtaining a measure on a ring—it is sufficient to define a measure on any semi-ring which generates the given ring. The extension to the generated ring is easily carried out, is unique, and gives a measure. There are circumstances in which one can define a set function μ directly on a ring so that it is easy to see that μ is non-negative and additive. Under these circumstances one can often use theorem 3.2 as a criterion for determining whether or not μ is a measure. Another useful criterion is given by the following theorem.

Theorem 3.6. Suppose $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ is non-negative and additive on a ring \mathcal{R} . Then

(i) if $E \in \mathcal{R}$, and $\{E_i\}$ is a sequence of disjoint sets of \mathcal{R} such that $E \supset \bigcup_{i=1}^{\infty} E_i$

$$\mu(E) \geq \sum_{i=1}^{\infty} \mu(E_i);$$

(ii) μ is a measure if and only if for any sequence $\{E_i\}$ of sets in \mathcal{R} such that $\bigcup_{i=1}^{\infty} E_i \supset E \in \mathcal{R}$,

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i).$$

Proof (i). For each positive integer n ,

$$E \supset \bigcup_{i=1}^n E_i \in \mathcal{R} \quad \text{so that} \quad \mu(E) \geq \mu\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mu(E_i),$$

since μ is additive. Hence

$$\mu(E) \geq \sum_{i=1}^{\infty} \mu(E_i).$$

(ii) First, suppose that μ is a measure. Put

$$F_i = E \cap E_i \quad (i = 1, 2, \dots); \quad G_1 = F_1,$$

and

$$G_n = F_n - \bigcup_{i=1}^{n-1} F_i \quad (n = 2, 3, \dots).$$

Then $\{G_n\}$ is a sequence of disjoint sets of \mathcal{R} such that

$$G_n \subset F_n \subset E_n, \quad \bigcup_{n=1}^{\infty} G_n = \bigcup_{n=1}^{\infty} F_n = E.$$

Thus
$$\mu(E) = \mu\left(\bigcup_{i=1}^{\infty} G_i\right) = \sum_{i=1}^{\infty} \mu(G_i) \leq \sum_{i=1}^{\infty} \mu(E_i),$$

since μ is σ -additive and non-negative.

Conversely if it is known that μ is additive and $E = \bigcup_{i=1}^{\infty} E_i$ is a disjoint decomposition of $E \in \mathcal{R}$ into sets in \mathcal{R} , by (i)

$$\mu(E) \geq \sum_{i=1}^{\infty} \mu(E_i);$$

and if the condition in (ii) is satisfied,

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

so that we must have
$$\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$$

and μ is a measure on \mathcal{R} . **]**

Exercises 3.3

1. If $\Omega = \{1, 2, 3, 4, 5\}$, show that \mathcal{C} consisting of \emptyset , Ω , $\{1\}$, $\{2, 3\}$, $\{1, 2, 3\}$, $\{4, 5\}$ is a semi-ring and that $0, 3, 1, 1, 2, 1$ defines a set of values for an additive set function μ on \mathcal{C} . What is the ring \mathcal{R} generated by \mathcal{C} ? Find the additive extension of μ to \mathcal{R} , and show that it is a measure.

2. Suppose \mathcal{R} is any ring of subsets, $\phi: \mathcal{R} \rightarrow \mathbf{R}^+$ is non-negative, finitely additive on \mathcal{R} , and $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ is a measure on \mathcal{R} such that, for any sequence of sets $\{A_n\}$ in \mathcal{R}

$$\mu(A_n) \rightarrow 0 \Rightarrow \phi(A_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty;$$

show that ϕ is completely additive.

3. If $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ is finitely additive on a ring \mathcal{R} and $E, F \in \mathcal{R}$ are such that $\mu(E \triangle F) = 0$, we say that $E \sim F$. Show that \sim is an equivalence relation in \mathcal{R} and that

$$E \sim F \Rightarrow \mu(E) = \mu(F) = \mu(E \cup F) = \mu(E \cap F).$$

Is the class of all sets $E \in \mathcal{R}$ for which $E \sim \emptyset$ a ring?

4. In the notation of question 3, put $\rho(E, F) = \mu(E \triangle F)$ and show that $\rho(E, F) \geq 0$, $\rho(E, F) = \rho(F, E)$, $\rho(E, F) \leq \rho(E, G) + \rho(G, F)$. If $E_1 \sim E_2$, $F_1 \sim F_2$ are all sets in \mathcal{R} , show that $\rho(E_1, F_1) = \rho(E_2, F_2)$. Does ρ define a metric in \mathcal{R} ?

3.4 Length, area and volume of elementary figures

In §1.5 we saw that:

(i) In $\mathbf{R} = \mathbf{R}^1$ (Euclidean 1-space) the class $\mathcal{P} = \mathcal{P}^1$ of half-open intervals $(a, b]$ forms a semi-ring which generates the ring \mathcal{E} of elementary figures (sets E of the form $E = \bigcup_{i=1}^n (a_i, b_i]$ with $b_i < a_{i+1}$ ($i = 1, 2, \dots, n-1$)).

(ii) In \mathbf{R}^k the half-open intervals have the form $\{(x_1, x_2, \dots, x_k): a_i < x_i \leq b_i, i = 1, 2, \dots, k\}$ and they again form a semi-ring \mathcal{P}^k which generates the ring \mathcal{E}^k of elementary figures (sets which can be expressed as a finite union of disjoint sets of \mathcal{P}^k).

Instead of using the terms length (for $k = 1$), area (for $k = 2$) and volume (for $k \geq 3$) of an interval we will use the same word 'length' in each case. Thus the 'length' of an interval of \mathcal{P}^k will be the product of the lengths of k perpendicular edges.

$$\mu(a, b] = b - a,$$

$$\mu\{(x_1, \dots, x_k): a_i < x_i \leq b_i, i = 1, 2, \dots, k\} = \prod_{i=1}^k (b_i - a_i).$$

Thus for each k we have defined a set function

$$\mu: \mathcal{P}^k \rightarrow \mathbf{R}^+$$

which has the usual physical meaning of length, area or volume. Historically this set function and its extension to a larger class of subsets of \mathbf{R}^k was the first to be studied; it leads quickly to the definition of Lebesgue measure in \mathbf{R}^k . Our object in the present section is to show that the set function obtained by extending μ from \mathcal{P}^k to \mathcal{E}^k is a measure on \mathcal{E}^k . There are essentially two distinct methods of doing this, and both will work for each k . In both it is necessary to show that μ is additive on \mathcal{P}^k so that it has a unique extension to an additive

set function in \mathcal{E}^k . Then one can either make use of the continuity theorem 3.2 to show that $\mu: \mathcal{E}^k \rightarrow \mathbb{R}^+$ is a measure on \mathcal{E}^k , or one can prove directly that μ is a measure on \mathcal{P}^k and appeal to theorem 3.5 to deduce that its extension is also a measure. We illustrate by applying the first method to the case $k = 1$, and the second method to the case $k = 2$.

$k = 1$

For each $(a, b] \in \mathcal{P}$ we put $\mu(a, b] = b - a$. It follows that μ is additive on \mathcal{P} for if $(a, b] = \bigcup_{i=1}^n (a_i, b_i]$ and the $(a_i, b_i]$ are disjoint we may assume that these intervals are ordered so that $b_i \leq a_{i+1}$ ($i = 1, 2, \dots, n-1$). It follows that we must have $a_1 = a, b_n = b$ and $b_i = a_{i+1}$ ($i = 1, 2, \dots, n-1$) so that, if $a_{n+1} = b_n$,

$$\begin{aligned} \sum_{i=1}^n \mu(a_i, b_i] &= \sum_{i=1}^n (b_i - a_i) = \sum_{i=1}^n (a_{i+1} - a_i) \\ &= (b - a) = \mu(a, b]. \end{aligned}$$

By theorem 3.4 there is a unique additive extension $\mu: \mathcal{E} \rightarrow \mathbb{R}^+$ since \mathcal{E} is the smallest ring containing the semi-ring \mathcal{P} . Since μ is finite on \mathcal{E} it will follow from theorem 3.2 (iii) that μ is a measure, if we can prove that μ is continuous from above at \emptyset .

Suppose this is false; then there is a monotone sequence $\{E_n\}$ of sets in \mathcal{E} for which $\lim E_n = \emptyset$ but $\mu(E_n) \rightarrow \delta > 0$ as $n \rightarrow \infty$. Now E_1 consists of a finite number of intervals of \mathcal{P} . Let F_1 be a set of \mathcal{E} obtained by taking away short half-open intervals of \mathcal{P} from the left-hand end of each of the intervals of E_1 in such a way that

$$F_1 \subset \bar{F}_1 \subset E_1; \quad \mu(F_1) > \mu(E_1) - \delta/2^2.$$

We now proceed by induction. Suppose we have obtained $F_n \in \mathcal{E}$ such that

$$F_n \subset \bar{F}_n \subset E_n \cap F_{n-1}$$

$$\text{and} \quad \mu(F_n) > \mu(E_n) - \sum_{r=1}^n \frac{\delta}{2^{r+1}}. \quad (3.4.1)$$

Then $F_n \cap E_{n+1} \in \mathcal{E}$ and

$$\mu(F_n \cap E_{n+1}) \geq \mu(E_{n+1}) - \mu(E_n - F_n) \geq \mu(E_{n+1}) - \sum_{r=1}^n \frac{\delta}{2^{r+1}}. \quad (3.4.2.)$$

We can again remove small half-open intervals from the left-hand end of each interval of $F_n \cap E_{n+1}$ to give a set $F_{n+1} \in \mathcal{E}$ such that

$$\mu(F_{n+1}) > \mu(E_{n+1} \cap F_n) - \delta/2^{n+2} \quad (3.4.3)$$

and

$$F_{n+1} \subset \bar{F}_{n+1} \subset E_{n+1} \cap F_n.$$

By (3.4.2) and (3.4.3) we deduce that

$$\mu(F_{n+1}) > \mu(E_{n+1}) - \sum_{r=1}^{n+1} \frac{\delta}{2^{r+1}}.$$

Thus by induction we can establish (3.4.1) for all n . Since $\mu(E_n) \geq \delta$ for all n , we have

$$\mu(F_n) > \frac{1}{2}\delta, \quad \text{for all } n$$

so that all the sets F_n are non-void. Hence $\{\bar{F}_n\}$ is a decreasing sequence of non-empty bounded closed sets. Hence $\bigcap_{n=1}^{\infty} \bar{F}_n$ is not void. But

$$\bigcap_{n=1}^{\infty} \bar{F}_n \subset \bigcap_{n=1}^{\infty} E_n = \emptyset,$$

so we obtain a contradiction.

$k = 2$

Suppose $C = \{(x, y): a < x \leq b, c < y \leq d\}$ is a set of \mathcal{P}^2 , and $\mu(C) = (b-a)(d-c)$. In order to prove that μ is additive on \mathcal{P}^2 , suppose that $C = \bigcup_{i=1}^n C_i$ is a decomposition of C into disjoint rectangles in each of which one of the sides (say $(c, d]$) remains the same. Then the other sides $(a_i, b_i]$ must be disjoint and satisfy

$$(a, b] = \bigcup_{i=1}^n (a_i, b_i]$$

so that by the corresponding result in \mathcal{P}^1 , μ is additive in this case. More generally if

$$C = \bigcup_{i=1}^n C_i, \quad C_i = \{(x, y): a_i < x \leq b_i, c_i < y \leq d_i\}$$

is a decomposition of C into a finite number of disjoint rectangles, use the infinite lines $x = a_i$, $x = b_i$, ($i = 1, 2, \dots, n$) to decompose each C_i into a finite number of pieces C_{ik} each with the same bounds for the y -coordinate. Hence

$$\sum_{i=1}^n \mu(C_i) = \sum_i \sum_k \mu(C_{ik}),$$

and we can sum the right-hand side by first summing over the rectangles whose x -coordinate is bounded by a pair of contiguous a_i , b_j and then summing over these intervals in x . Thus by repeated application of additivity in \mathcal{P}^1 we get

$$\mu(C) = \sum_{i=1}^n \mu(C_i),$$

as required. (The reader should draw a picture.)

Now suppose $C = \bigcup_{i=1}^{\infty} C_i$ is an infinite decomposition of C into disjoint sets of \mathcal{P}^2 . We must show that μ is completely additive on \mathcal{P}^2 . Since \mathcal{P}^2 is a semi-ring it follows by induction that, for each n ,

$$C - \bigcup_{i=1}^n C_i$$

can be expressed as a finite union of sets of \mathcal{P}^2 . Since μ is non-negative, this implies that

$$\mu(C) \geq \sum_{i=1}^n \mu(C_i), \quad \text{for all } n,$$

so that

$$\mu(C) \geq \sum_{i=1}^{\infty} \mu(C_i).$$

Suppose if possible that μ is not σ -additive, then there will be such a set C for which

$$\mu(C) = \sum_{i=1}^{\infty} \mu(C_i) + 2\delta \quad (\delta > 0). \quad (3.4.4)$$

We now use another form of compactness argument to obtain a contradiction. Suppose $\epsilon > 0$ is small enough to ensure that, if

$$F_0 = \{(x, y): a + \epsilon < x \leq b, c + \epsilon < y \leq d\},$$

then

$$\mu(F_0) > \mu(C) - \delta;$$

and $\epsilon_i > 0$ are small enough to ensure that, if

$$F_i = \{(x, y): a_i < x \leq b_i + \epsilon_i, c_i < y \leq d_i + \epsilon_i\},$$

then

$$\mu(F_i) < \mu(C_i) + \delta 2^{-n} \quad (i = 1, 2, \dots). \quad (3.4.5)$$

Then $\bar{F}_0 \subset C$ and $C_i \subset F_i^0$, the interior of F_i ($i = 1, 2, \dots$); so that

$$\bar{F}_0 \subset \bigcup_{i=1}^{\infty} F_i^0.$$

Since \bar{F}_0 is compact and the sets F_i^0 are open it follows that, for some integer n , we have

$$\bar{F}_0 \subset \bigcup_{i=1}^n F_i^0 \quad \text{so that} \quad F_0 \subset \bigcup_{i=1}^n F_i.$$

By the finite additivity of μ on \mathcal{P}^2 this implies

$$\mu(F_0) \leq \sum_{i=1}^n \mu(F_i)$$

so that, by (3.4.5)

$$\mu(C) - \delta < \sum_{i=1}^{\infty} \mu(C_i) + \sum_{i=1}^{\infty} \delta 2^{-i}.$$

Which contradicts (3.4.4).

Thus μ is a measure on \mathcal{P}^2 and by theorem 3.5 the unique additive extension $\mu: \mathcal{E}^2 \rightarrow \mathbf{R}^+$ is also a measure. Either form of argument clearly extends to the class of elementary figures in \mathbf{R}^k , so we have proved:

Theorem 3.7. *Suppose \mathcal{E}^k is the class of elementary figures in \mathbf{R}^k , that is, the class of those sets*

$$E = \bigcup_{i=1}^n C_i$$

where the C_i are disjoint half-open intervals in \mathbf{R}^k . If we put $\mu(C_i) = \text{length of } C_i = \text{product of lengths of the sides of the interval } C_i$ and

$$\mu(E) = \sum_{i=1}^n \mu(C_i),$$

then μ is uniquely defined on \mathcal{E}^k and is a measure.

CONSTRUCTION AND PROPERTIES OF MEASURES

4.1 Extension theorem; Lebesgue measure

Measure was defined as a non-negative σ -additive set function defined on a class of sets \mathcal{C} . In testing τ for σ -additivity we only needed

$$\tau(E) = \sum_{i=1}^{\infty} \tau(E_i)$$

for sequences $\{E_i\}$ of disjoint sets of \mathcal{C} for which $E = \bigcup_{i=1}^{\infty} E_i \in \mathcal{C}$. This is an artificial restriction as the condition of additivity does not apply to a sequence $\{E_i\}$ unless the union set $\bigcup_{i=1}^{\infty} E_i$ happens to belong to \mathcal{C} . For this reason the natural domain of definition for a measure $\tau: \mathcal{C} \rightarrow \mathbf{R}^+$ is a σ -ring, and in practice most useful measures are defined on σ -rings.

In the last chapter we considered properties of measures defined on a ring \mathcal{R} , so our first objective in the present chapter will be to prove that these can always be extended to a measure on the σ -ring \mathcal{S} generated by \mathcal{R} . This extension is unique provided the measure on \mathcal{R} is σ -finite. We introduce an (unnecessary) simplifying assumption—that the generated σ -ring \mathcal{S} is also a σ -field, i.e. that it contains the whole space Ω . Even with this simplification the main extension theorem is somewhat involved. The main idea is that of introducing an outer approximating set function, defined in terms of the measure on \mathcal{R} , and then restricting this to a class of sets on which it is σ -additive. The relevant set function turns out to be an outer measure, so it is convenient first to obtain a theorem about all outer measures.

Measurable set

Suppose μ^* is an outer measure defined for all subsets of Ω : that is, μ^* is non-negative, countably subadditive and monotone (see p. 59). A subset E is said to be measurable with respect to μ^* if, for every set $A \subset \Omega$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A - E). \quad (4.1.1)$$

It is important to stress that the concept of measurability for a set depends on the outer measure μ^* . The same set E may well be

measurable with respect to μ_1^* and non-measurable with respect to μ_2^* . It helps one's intuition to realise that (4.1.1) states that if one divides the set A using E and its complement, then the outer measure of the 'pieces' adds up correctly. Thus a set E is μ^* -measurable if and only if it breaks up no set A into two subsets on which μ^* is not additive. The measurability of E depends on what the set E does to the outer measure of all the other subsets.

The reader may find the above explanation of condition (4.1.1) still inadequate to provide the definition of measurability with much intuitive content. This is a case where the definition is justified by the result—it turns out that, for suitable outer measures, a wide class of sets is measurable and the class of measurable sets has got the right kind of structural properties. The definition is therefore justified ultimately by the elegance and usefulness of the theory which results from it.

Note that, because of the subadditivity condition on outer measures, we always have

$$\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A - E)$$

for all sets A, E . Hence E is μ^* -measurable if and only if

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A - E) \quad (4.1.2)$$

for every set $A \subset \Omega$. Since (4.1.2) is automatically satisfied for sets A with $\mu^*(A) = +\infty$, E is μ^* -measurable if and only if (4.1.2) is satisfied for every $A \subset \Omega$ with $\mu^*(A) < \infty$.

It is worth remarking that many of the early discussions of measurability use the concept of *inner measure*. If $\mu^*(\Omega) < \infty$, this can be defined for all subsets E by

$$\mu_*(E) = \mu^*(\Omega) - \mu^*(\Omega - E).$$

In this method of procedure a set E is said to be measurable if $\mu_*(E) = \mu^*(E)$. This apparently weaker definition of measurability can be shown to be equivalent to the one we have adopted provided the outer measure μ^* is *regular*. (An outer measure is said to be regular if, for every $A \subset \Omega$, there is a measurable cover $E \supset A$ such that $\mu^*(E) = \mu^*(A)$.) This means in particular that, under these circumstances, it is sufficient to use the single test set Ω for A in (4.1.1). We do not use the concept of inner measure in our development.

Theorem 4.1. *Let μ^* be an outer measure on Ω , and let \mathcal{M} be the class of sets of Ω which are measurable with respect to μ^* . Then \mathcal{M} is a σ -field and the restriction of μ^* to \mathcal{M} defines a measure on \mathcal{M} .*

Proof. We first show that any finite union of sets of \mathcal{M} is in \mathcal{M} . It is clearly sufficient to prove that $E_1 \cup E_2 \in \mathcal{M}$ for any $E_1, E_2 \in \mathcal{M}$. For any set A , since E_1 is measurable,

$$\mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A - E_1). \quad (4.1.3)$$

Now use $(A - E_1)$ as a test set for the measurable E_2

$$\begin{aligned} \mu^*(A - E_1) &= \mu^*((A - E_1) \cap E_2) + \mu^*(A - E_1 - E_2), \\ \mu^*(A - E_1) &= \mu^*((A - E_1) \cap E_2) + \mu^*(A - (E_1 \cup E_2)). \end{aligned} \quad (4.1.4)$$

But $[(A - E_1) \cap E_2] \cup (A \cap E_1) = A \cap (E_1 \cup E_2)$,

so that if we substitute (4.1.4) into (4.1.3) and use the subadditivity of μ^* , we obtain

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap E_1) + \mu^*((A - E_1) \cap E_2) + \mu^*(A - (E_1 \cup E_2)) \\ &\geq \mu^*(A \cap (E_1 \cup E_2)) + \mu^*(A - (E_1 \cup E_2)) \end{aligned}$$

so that, by (4.1.2), $E_1 \cup E_2 \in \mathcal{M}$.

Now, since $A \cap E = A - (\Omega - E)$, the equation for the measurability of $(\Omega - E)$ is the same as that for the measurability of E . Hence, $(\Omega - E)$ is measurable if and only if E is measurable.

Since

$$\bigcap_{i=1}^n E_i = \Omega - \bigcup_{i=1}^n (\Omega - E_i), \quad (\text{see } \S 1.4),$$

it follows that the class \mathcal{M} is also closed under finite intersections so that \mathcal{M} is a field. In order to show that \mathcal{M} is a σ -field it is sufficient to show that $E = \bigcup_{k=1}^{\infty} E_k \in \mathcal{M}$ for any sequence $\{E_k\}$ of sets of \mathcal{M} . There is no loss in generality in assuming that the sets E_k are disjoint for, since \mathcal{M} is a ring, any countable union can be replaced by a countable disjoint union of subsets in \mathcal{M} . Put

$$F_n = \bigcup_{k=1}^n E_k \quad (n = 1, 2, \dots),$$

and let \mathcal{H}_n be the hypothesis that, for any A ,

$$\mu^*(A \cap F_n) = \sum_{k=1}^n \mu^*(A \cap E_k).$$

Clearly \mathcal{H}_1 is true. Use $A \cap F_{n+1}$ as a test set for the measurability of F_n : then

$$\mu^*(A \cap F_{n+1}) = \mu^*(A \cap F_n) + \mu^*(A \cap E_{n+1}).$$

Hence $\mathcal{H}_n \Rightarrow \mathcal{H}_{n+1}$ so that, by induction \mathcal{H}_n is true for all positive integers n .

Since μ^* is monotonic, for each n

$$\mu^*(A \cap E) \geq \mu^*(A \cap F_n) = \sum_{k=1}^n \mu^*(A \cap E_k),$$

so that

$$\mu^*(A \cap E) \geq \sum_{k=1}^{\infty} \mu^*(A \cap E_k),$$

and the subadditivity of μ^* now implies that

$$\mu^*(A \cap E) = \sum_{k=1}^{\infty} \mu^*(A \cap E_k). \quad (4.1.5)$$

Thus, for any A , and any n ,

$$\mu^*(A) = \mu^*(A \cap F_n) + \mu^*(A - F_n) \geq \sum_{k=1}^n \mu^*(A \cap E_k) + \mu^*(A - E)$$

using \mathcal{H}_n and the monotonicity of μ^* . Thus, by (4.1.5),

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A - E),$$

and this implies $E \in \mathcal{M}$ by (4.1.2).

Now the restriction of μ^* to \mathcal{M} is a non-negative set function. Further (4.1.5) with A replaced by Ω shows that μ^* is σ -additive on \mathcal{M} and is therefore a measure on \mathcal{M} .]

We can now prove the basic extension theorem. In order to simplify the formulation we will assume that the ring \mathcal{R} of subsets of Ω is such that there is a sequence of sets $\{E_n\}$ in \mathcal{R} such that $\Omega = \bigcup_{n=1}^{\infty} E_n$.

We then say that Ω is σ - \mathcal{R} . This condition implies that the σ -ring generated by \mathcal{R} is a σ -field. Theorem 4.2 is true without this restriction, but the proof would then require the consideration of outer measures defined on a suitable class of subsets of Ω , rather than on all subsets. Since this generalisation also causes complications in the definition of the integral, and the extra generality is rarely needed, we will keep the condition that Ω be σ - \mathcal{R} .

Theorem 4.2. Suppose \mathcal{R} is a ring of subsets of Ω such that Ω is σ - \mathcal{R} and $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ is a measure defined on \mathcal{R} . Then there is an extension of μ to a measure ν defined on $\mathcal{S}(\mathcal{R})$, the σ -ring generated by \mathcal{R} . If μ is σ -finite on \mathcal{R} , then the extension is unique, and is σ -finite on \mathcal{S} .

Proof. Let \mathcal{C} be the class of all subsets of Ω . Since Ω is σ - \mathcal{R} , any $E \in \mathcal{C}$ can be covered by a countable sequence of sets of \mathcal{R} . Put

$$\mu^*(E) = \inf \sum_{i=1}^{\infty} \mu(F_i),$$

the infimum being taken over all sequences of sets $\{F_i\}$ in \mathcal{R} such that $E \subset \bigcup_{i=1}^{\infty} F_i$. It is clear that $\mu^*: \mathcal{C} \rightarrow \mathbf{R}^+$ is non-negative, monotone and

that $\mu^*(\emptyset) = 0$. Suppose now that $E \subset \bigcup_{i=1}^{\infty} E_i$. Then, if $\mu^*(E_i)$ is infinite for some i ,

$$\mu^*(E) \leq \sum_{i=1}^{\infty} \mu^*(E_i) \quad (4.1.6)$$

is immediate. If $\mu^*(E_i) < \infty$ for all i ; for any $\epsilon > 0$, choose sets F_{ik} ($k = 1, 2, \dots$) in \mathcal{R} such that

$$E_i \subset \bigcup_{k=1}^{\infty} F_{ik} \quad \text{and} \quad \sum_{k=1}^{\infty} \mu(F_{ik}) < \mu^*(E_i) + \frac{\epsilon}{2^i} \quad (i = 1, 2, \dots).$$

The countable collection $\{F_{ik}\}$ will now cover E , and

$$\mu^*(E) \leq \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \mu(F_{ik}) \leq \sum_{i=1}^{\infty} \left(\mu^*(E_i) + \frac{\epsilon}{2^i} \right).$$

Since ϵ is arbitrary, (4.1.6) now follows, and we have proved that $\mu^*: \mathcal{C} \rightarrow \mathbf{R}^+$ is an outer measure. Let \mathcal{M} be the class of subsets of Ω which are measurable with respect to μ^* .

We first want to show that $\mathcal{M} \supset \mathcal{R}$. If $E \in \mathcal{R}$ and $\mu^*(A) < \infty$ (the case $\mu^*(A) = +\infty$ is unimportant as (4.1.2) is then trivially satisfied), choose a sequence $\{E_i\}$ of sets of \mathcal{R} such that $A \subset \bigcup_{i=1}^{\infty} E_i$ and

$$\begin{aligned} \mu^*(A) + \epsilon &\geq \sum_{i=1}^{\infty} \mu(E_i) = \sum_{i=1}^{\infty} [\mu(E_i \cap E) + \mu(E_i - E)] \\ &\geq \mu^*(A \cap E) + \mu^*(A - E), \end{aligned}$$

by the subadditivity of μ^* . Since ϵ is arbitrary, we have again proved (4.1.2), so that $E \in \mathcal{M}$. By theorem 4.1, \mathcal{M} is a σ -ring, so that $\mathcal{M} \supset \mathcal{S}$, the σ -ring generated by \mathcal{R} . But the restriction of μ^* to \mathcal{M} is a measure, so that its further restriction ν to \mathcal{S} is also a measure.

If $E \in \mathcal{R}$ it is clear that $\mu^*(E) \geq \mu(E)$ because of theorem 3.6 (i), and since E is a covering of itself, $\mu^*(E) \leq \mu(E)$. Hence, for all sets $E \in \mathcal{R}$, we have $\nu(E) = \mu^*(E) = \mu(E)$, so that ν is an extension of μ from \mathcal{R} to \mathcal{S} .

If we now assume that μ^* is σ -finite on \mathcal{R} , it follows that $\Omega = \bigcup_{i=1}^{\infty} E_i$ with $\{E_i\}$ an increasing sequence of sets in \mathcal{R} and $\mu(E_i)$ finite, $i = 1, 2, \dots$. For a fixed integer n , consider the ring \mathcal{R}_n consisting of sets of the form $E_n \cap E$ with $E \in \mathcal{R}$. Suppose μ_1 and μ_2 are any two extensions of μ from \mathcal{R}_n to $\mathcal{S}_n = \mathcal{S}(\mathcal{R}_n)$. Then all the subsets in \mathcal{S}_n are contained in the set E_n so that μ_1 and μ_2 are finite on \mathcal{S}_n . Now let \mathcal{T}_n be the subclass of those sets E of \mathcal{S}_n for which $\mu_1(E) = \mu_2(E)$. Since

finite measures are continuous from above and below, it follows that \mathcal{T}_n is a monotone class. By theorem 1.5, since $\mathcal{T}_n \supset \mathcal{R}_n$, it follows that $\mathcal{T}_n \supset \mathcal{S}_n$ and we must have $\mathcal{T}_n = \mathcal{S}_n$. Thus the extension of μ to \mathcal{S}_n is unique for every n . But, for any $E \in \mathcal{S}$ we have

$$E = \lim_{n \rightarrow \infty} E \cap E_n$$

so that a further application of the continuity theorem shows that the extension of μ to \mathcal{S} must be unique. \blacksquare

Theorem 4.2 can be applied to any measure defined on a ring \mathcal{R} . In 3.4 we saw that the concept of length in \mathbf{R}^1 , area in \mathbf{R}^2 and volume in \mathbf{R}^k ($k \geq 3$) could be precisely formulated on the ring \mathcal{E}^k of elementary figures to define a measure on \mathcal{E}^k . It is clear that \mathbf{R}^k is σ - \mathcal{E}^k , and the measure is actually finite on \mathcal{E}^k . The σ -ring generated by \mathcal{E}^k is the class \mathcal{B}^k of Borel sets in \mathbf{R}^k (proved in §2.5). Thus if we apply the statement of theorem 4.2 to this measure $\mu: \mathcal{E}^k \rightarrow \mathbf{R}^+$ we obtain a unique extension to a measure $\nu: \mathcal{B}^k \rightarrow \mathbf{R}^+$ which is σ -finite on \mathcal{B}^k . It is worth noticing that in the proof of theorem 4.2 the extension was actually carried out to a class of measurable sets containing \mathcal{B}^k . This class is denoted by \mathcal{L}^k and can be shown to be larger than \mathcal{B}^k . A set $E \subset \mathbf{R}^k$ is said to be *Lebesgue measurable* if and only if it is in the class \mathcal{L}^k . In particular all Borel sets in \mathbf{R}^k are Lebesgue measurable. The set function $\nu: \mathcal{L}^k \rightarrow \mathbf{R}^+$ is called *Lebesgue measure* in k -space and should be thought of as a generalisation of the notion of k -dimensional volume to a very wide class of sets. We will examine the properties of this set function in some detail in §4.4, and it will then become clear that many of our intuitive ideas of length, area, and volume can be precisely formulated and remain valid for Lebesgue measure.

It is worth noticing that the outer measure obtained by covering as in theorem 4.2 is always a regular outer measure. For, if $\mu^*(E) < \infty$, choose sets $T_{n,r} \in \mathcal{R}$ ($r = 1, 2, \dots$) such that

$$E \subset \bigcup_{r=1}^{\infty} T_{n,r}, \quad \mu^*(E) + \frac{1}{n} > \sum_{r=1}^{\infty} \mu(T_{n,r}).$$

Then

$$A = \bigcap_{n=1}^{\infty} \bigcup_{r=1}^{\infty} T_{n,r} \supset E, \quad A \in \mathcal{S},$$

and $\mu^*(A) = \mu^*(E)$. This means that the approach through inner measure will lead to the same class of measurable sets and the same extension to this class. In particular the Lebesgue measure can be obtained by this method provided one considers subsets of a fixed bounded interval (of finite measure) in the first instance and then allows the interval to expand to the whole Euclidean space.

Exercises 4.1

1. Suppose μ^* is an outer measure on $\Omega = \lim E_k$ where $\{E_k\}$ is a monotone increasing sequence of sets. Show that if a set E is such that $E \cap E_k$ is measurable (μ^*) for all sufficiently large k , then E is measurable (μ^*).

2. Show that if μ^* is a regular outer measure on Ω and $\mu^*(\Omega) < \infty$, then a necessary and sufficient condition for E to be measurable (μ^*) is that

$$\mu^*(\Omega) = \mu^*(E) + \mu^*(\Omega - E).$$

3. In each of the following cases, show that μ^* is an outer measure, and determine the class of measurable sets

- (i) $\mu^*(\emptyset) = 0$, $\mu^*(E) = 1$ for all $E \neq \emptyset$.
- (ii) $\mu^*(\emptyset) = 0$, $\mu^*(E) = 1$ for $E \neq \emptyset$ or Ω , $\mu^*(\Omega) = 2$.
- (iii) Ω is not countable; $\mu^*(E) = 0$ if E is countable, $\mu^*(E) = 1$ if E is not countable.

4. Show that any outer measure which is (finitely) additive is σ -additive.

5. Suppose μ^* is an outer measure on Ω and E, F are two subsets, at least one of which is measurable (μ^*). Show that

$$\mu^*(E) + \mu^*(F) = \mu^*(E \cup F) + \mu^*(E \cap F).$$

6. Suppose $\{E_n\}$ is a sequence of sets in a σ -ring \mathcal{S} , and μ is a measure on \mathcal{S} . Show that

- (i) $\mu(\liminf E_n) \leq \liminf \mu(E_n)$;
- (ii) provided $\bigcup_{k=n}^{\infty} E_k$ has finite measure for some n ,

$$\mu(\limsup E_n) \geq \limsup \mu(E_n).$$

If $\sum_{n=1}^{\infty} \mu(E_n) < \infty$, show that $\mu(\limsup E_n) = 0$.

7. Show that, if μ is a discrete measure on Ω (as in example (6) of §3.1 with $p_i \geq 0$), then the operation of extending it to an outer measure and restricting this extension to the class of measurable sets as in theorem 4.2 yields nothing new.

8. Suppose \mathcal{M} is the σ -ring of μ^* -measurable sets in Ω . Then if $\{E_n\}$ is a monotone increasing sequence of sets in \mathcal{M} and A is any set

$$\mu^*(\lim_{n \rightarrow \infty} A \cap E_n) = \lim_{n \rightarrow \infty} \mu^*(A \cap E_n).$$

Prove a corresponding result for a decreasing sequence (which needs an additional condition).

9. If μ^* is a regular outer measure, show that $\mu^*(\lim A_n) = \lim \mu^*(A_n)$ for any increasing sequence $\{A_n\}$.

10. Suppose in theorem 4.2 that μ is known only to be finitely additive on \mathcal{R} ; then the same procedure yields an outer measure μ^* and a restriction

$\bar{\mu}$ of μ^* to the μ^* -measurable sets. Show that $\bar{\mu}$ is a measure but is not necessarily an extension of μ .

11. Suppose \mathcal{R} is a ring of subsets of a countable set Ω such that every set in \mathcal{R} is either empty or infinite, but the generated sigma-ring $\mathcal{S}(\mathcal{R})$ contains all subsets of Ω (see exercise 1.5(8)). Put $\mu_1(E)$ = number of points in E , $\mu_2(E) = 2\mu_1(E)$ for all subsets $E \subset \Omega$. Then μ_1, μ_2 agree on \mathcal{R} but not on $\mathcal{S}(\mathcal{R})$ so that the uniqueness assertion of theorem 4.2 requires μ to be σ -finite.

12. Suppose $h(t)$ is any continuous monotonic increasing function defined on $(0, y)$, $y > 0$ with $\lim_{t \rightarrow 0+} h(t) = 0$. If Ω is any metric space, let

$$h-m^*(E) = \lim_{\delta \rightarrow 0} \left[\inf \sum_{i=1}^{\infty} h\{\text{diam}(C_i)\} \right],$$

where the infimum is taken over all sequences $\{C_i\}$ of sets of diameter $< \delta$ which cover E (if there are no such coverings then the inf is $+\infty$). Show that $h-m^*(E)$ defines an outer measure in Ω . (It is called the Hausdorff measure with respect to $h(t)$.)

4.2 Complete measures

If we again think of measure as a mass distribution in the space Ω , it is clear that any subset of a set of zero mass should have the mass zero assigned to it. The present section seeks to make this notion precise.

Given a measure $\tau: \mathcal{C} \rightarrow \mathbf{R}^+$ we say that the class \mathcal{C} is complete with respect to τ if

$$E \subset F, \quad F \in \mathcal{C}, \quad \tau(F) = 0 \Rightarrow E \in \mathcal{C}.$$

(Note that since τ is monotone it follows that $\tau(E) = 0$.) If $\tau: \mathcal{C} \rightarrow \mathbf{R}^+$ is such that \mathcal{C} is complete with respect to τ we also say that τ is a *complete measure*.

All measures μ which are obtained (as in theorem 4.1) by restricting an outer measure μ^* to the class \mathcal{M} of sets which are measurable (μ^*) are complete measures. For, since outer measures are monotone, non-negative,

$$E \subset F, \quad \mu^*(F) = 0 \Rightarrow \mu^*(E) = 0,$$

and all sets E of zero μ^* -measure are measurable μ^* by (4.1.2) since

$$\mu^*(A) \geq \mu^*(A - E) = \mu^*(A - E) + \mu^*(A \cap E).$$

In particular Lebesgue measure defined on the class \mathcal{L}^k is a complete measure.

Given any measure μ on a σ -ring \mathcal{S} , there is a simple method of extending it to a complete measure on a larger σ -ring—called the *completion* of \mathcal{S} with respect to μ .

Theorem 4.3. *Given a measure μ on a σ -ring \mathcal{S} , let $\bar{\mathcal{S}}$ be the class of all sets of the form $E \triangle N$ where $E \in \mathcal{S}$ and $N \subset F \in \mathcal{S}$ with $\mu(F) = 0$. Then $\bar{\mathcal{S}}$ is a σ -ring and if we put*

$$\bar{\mu}(E \triangle N) = \mu(E),$$

then $\bar{\mu}: \bar{\mathcal{S}} \rightarrow \mathbf{R}^+$ is a (uniquely) defined extension of μ from \mathcal{S} to $\bar{\mathcal{S}}$, and $\bar{\mu}$ is a complete measure on $\bar{\mathcal{S}}$.

Proof. Let $E_0 = E \triangle N$, where $E \in \mathcal{S}$, $N \subset F \in \mathcal{S}$, $\mu(F) = 0$. Put $E_1 = E - F$, then $E_1 \subset E_0$, $E_1 \in \mathcal{S}$ and $\mu(E_1) = \mu(E)$. If

$$N_1 = E_0 - E_1,$$

then E_1, N_1 are disjoint and $E_0 = E_1 \cup N_1$. Further, since

$$E_0 \subset E \cup F = (E - F) \cup F,$$

we have $N_1 \subset F$ and $\mu(F) = 0$. Thus the class $\bar{\mathcal{S}}$ is the same as the class of sets $E \cup N$ with $E \in \mathcal{S}$, $N \subset F \in \mathcal{S}$, $\mu(F) = 0$ and $E \cap N = \emptyset$. A similar argument shows that $\bar{\mathcal{S}}$ is also the same as the class of sets of the form $E - N$ with $E \in \mathcal{S}$, $N \subset F \in \mathcal{S}$, $\mu(F) = 0$ and $N \subset E$.

It is now easy to check that $\bar{\mathcal{S}}$ is a ring. Suppose $E_1, E_2 \in \bar{\mathcal{S}}$; first express them as $E_1 = X_1 - N_1$, $E_2 = X_2 - N_2$, $N_1 \subset X_1$, $N_2 \subset X_2$ where $N_1 \subset F_1$, $N_2 \subset F_2$ and $\mu(F_1) = \mu(F_2) = 0$. Then

$$E_1 \cap E_2 = X_1 \cap X_2 - (N_1 \cup N_2),$$

and $X_1 \cap X_2 \in \mathcal{S}$, $N_1 \cup N_2 \subset F_1 \cup F_2 \in \mathcal{S}$, $\mu(F_1 \cup F_2) = 0$; so that $E_1 \cap E_2 \in \bar{\mathcal{S}}$. Now put

$$E_1 = X_3 \cup N_3, \quad E_2 = X_2 - N_2, \quad N_3 \cap X_3 = \emptyset, \quad N_3 \subset F_3 \quad \text{with}$$

$$\mu(F_3) = 0.$$

Then

$$E_1 - E_2 = (X_3 - X_2) \cup (N_3 - X_2) \cup (N_2 \cap E_1) = (X_3 - X_2) \cup N_5,$$

where $N_5 \subset F_3 \cup F_2$ and $\mu(F_3 \cup F_2) = 0$. Finally

$$E_1 = X_3 \cup N_3, \quad E_2 = X_4 \cup N_4, \quad \text{where} \quad X_4 \cap N_4 = \emptyset,$$

and $N_4 \subset F_4$ with $\mu(F_4) = 0$. Then

$$E_1 \cup E_2 = (X_3 \cup X_4) \cup (N_3 \cup N_4 - X_3 \cup X_4) = (X_3 \cup X_4) \cup N_6,$$

where $N_6 \subset F_3 \cup F_4$ and $\mu(F_3 \cup F_4) = 0$. Thus $\bar{\mathcal{S}}$ is closed under the finite operations of intersection, difference, union so it is a ring. To prove it is a σ -ring, put

$$E_i = X_i \cup N_i, \quad N_i \subset F_i, \quad \mu(F_i) = 0 \quad (i = 1, 2, \dots);$$

$$\text{then} \quad \bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} X_i \cup \bigcup_{i=1}^{\infty} N_i = X \cup N,$$

$$\text{where} \quad N \subset \bigcup_{i=1}^{\infty} F_i = F \quad \text{and} \quad \mu(F) = 0.$$

$$\text{Hence} \quad \bigcup_{i=1}^{\infty} E_i \in \bar{\mathcal{S}}.$$

To see that $\bar{\mu}$ is uniquely defined on $\bar{\mathcal{S}}$, let

$$E_1 \triangle N_1 = E_2 \triangle N_2$$

be two representations of the same set. Then (see exercise 1.4 (5))

$$E_1 \triangle E_2 = N_1 \triangle N_2$$

and $N_1 \triangle N_2 \subset F \in \mathcal{S}$ with $\mu(F) = 0$. Hence

$$\mu(E_1 - E_2) = \mu(E_2 - E_1) = 0,$$

$$\text{and} \quad \mu(E_1) = \mu(E_1 \cap E_2) = \mu(E_2).$$

Thus if we define $\bar{\mu}$ on $\bar{\mathcal{S}}$ by

$$\bar{\mu}(E_0) = \mu(E_1) \quad \text{if} \quad E_0 = E_1 \triangle N_1,$$

$\bar{\mu}$ is uniquely defined.

It only remains to show that $\bar{\mathcal{S}}$ is complete with respect to $\bar{\mu}$. Suppose E is any set of $\bar{\mathcal{S}}$ with $\bar{\mu}(E) = 0$. Then $E = X \cup N$ where $X \in \mathcal{S}$, $\mu(X) = 0$, $N \subset F \in \mathcal{S}$, $\mu(F) = 0$. Thus, if $G \subset E$, we have $G \subset X \cup F$ with $\mu(X \cup F) = 0$ and $X \cup F \in \mathcal{S}$; so that

$$G = \emptyset \cup G \in \bar{\mathcal{S}},$$

and $\bar{\mu}(G) = 0$. \blacksquare

We already saw that if μ was a σ -finite measure defined on a ring \mathcal{R} , then it had only one extension to a measure on the generated σ -ring \mathcal{S} . If we now complete \mathcal{S} to obtain the measure $\bar{\mu}$ defined on $\bar{\mathcal{S}}$ so that $\bar{\mathcal{S}}$ is now complete with respect to the extension $\bar{\mu}$ of μ , then we have extended μ from \mathcal{R} to $\bar{\mathcal{S}}$. Since the extension from \mathcal{S} to $\bar{\mathcal{S}}$ is also unique, it follows that there is only one extension of μ from \mathcal{R} to $\bar{\mathcal{S}}$. There is a sense in which, in general, this is as far as one can get with extensions while still preserving uniqueness, though it may be possible to extend μ further to a larger σ -field; see theorem 6.11.

It should also be noticed that in the extension theorem 4.2, the class \mathcal{M} of μ^* -measurable sets is none other than $\bar{\mathcal{P}}$ the completion of the σ -ring \mathcal{S} with respect to μ . For, in the first place, $\mathcal{M} \supset \mathcal{S}$ and \mathcal{M} is complete, hence $\mathcal{M} \supset \bar{\mathcal{P}}$. Secondly, if E is any set of \mathcal{M} such that $\mu(E) < \infty$, we can cover it by $F \in \mathcal{S}$ such that $\mu^*(F) = \mu^*(E)$. Then $F - E \in \mathcal{M}$ and has zero measure, so that it can be covered by a $G \in \mathcal{S}$ with $\mu(G) = 0$, and

$$E = (F - G) \cup (E \cap G) \in \bar{\mathcal{P}}.$$

Since μ is σ -finite on \mathcal{M} , and $\bar{\mathcal{P}}$ is a σ -ring, it now follows that $\mathcal{M} \subset \bar{\mathcal{P}}$.

In particular, Lebesgue measure on \mathcal{L}^k is the unique extension of the concept of length from the semi-ring \mathcal{P}^k to the σ -ring \mathcal{L}^k which is the completion of \mathcal{P}^k .

Exercises 4.2

1. Suppose μ is a measure on a σ -ring \mathcal{Q} and $\bar{\mu}$ on $\bar{\mathcal{Q}}$ is its completion. Show that if $A, B \in \mathcal{Q}$ with $A \subset E \subset B$, $\mu(B - A) = 0$ then $E \in \bar{\mathcal{Q}}$, and $\bar{\mu}(E) = \mu(A) = \mu(B)$.

2. Given a σ -finite measure μ on a ring \mathcal{R} the extension given by theorem 4.2 yields a complete measure on the class \mathcal{M} of μ^* -measurable sets which is the completion of \mathcal{S} the generated σ -ring. The following example shows that this is not true if the hypothesis of σ -finiteness is omitted: Let Ω be non-countable, \mathcal{S} the ring (also a σ -ring) of all sets which are countable or have countable complements, $\mu(E) = \text{number of points in } E$ for $E \in \mathcal{S}$. Then \mathcal{S} is complete with respect to μ , but applying theorem 4.2 yields a complete measure on the class of all subsets (as every subset is measurable).

4.3 Approximation theorems

We have seen how the definition of a measure can be extended from a ring \mathcal{R} to the generated σ -ring \mathcal{S} , and its completion $\bar{\mathcal{P}}$. It is convenient to think of the sets of \mathcal{R} as having a simple structure, so that it becomes interesting to see that the sets of $\bar{\mathcal{P}}$ can always be approximated in measure with arbitrary accuracy by sets in the original ring \mathcal{R} .

Theorem 4.4. Suppose \mathcal{R} is a ring for which Ω is σ - \mathcal{R} , and the σ -finite measure $\mu: \mathcal{R} \rightarrow \mathbf{R}^+$ has been extended (uniquely) to the completion $\bar{\mathcal{P}}$ of the σ -ring \mathcal{S} generated by \mathcal{R} . Then for any $\epsilon > 0$, any set $E \in \bar{\mathcal{P}}$ with $\mu(E) < \infty$, there is a set $F \in \mathcal{R}$ such that

$$\mu(E \Delta F) < \epsilon.$$

Proof. First, find a set $E_1 \in \mathcal{S}$ such that

$$\mu(E \triangle E_1) = 0.$$

Then $\mu(E_1) = \mu(E) < \infty$, so that by the construction of theorem 4.2, we have

$$\mu(E_1) = \mu^*(E_1) = \inf_{\substack{E_1 \subset \cup T_i \\ T_i \in \mathcal{R}}} \sum \mu(T_i),$$

so that we can choose a sequence of disjoint sets $\{T_i\}$ of \mathcal{R} such that

$$E_1 \subset \bigcup_{i=1}^{\infty} T_i \quad \text{and} \quad \mu^*(E_1) + \frac{1}{2}\epsilon > \sum_{i=1}^{\infty} \mu(T_i).$$

Now choose a finite integer n such that

$$\sum_{i=n+1}^{\infty} \mu(T_i) < \frac{1}{2}\epsilon,$$

and put

$$F = \bigcup_{i=1}^n T_i \in \mathcal{R}.$$

Then $E_1 - F \subset \bigcup_{i=n+1}^{\infty} T_i$, so that $\mu(E_1 - F) < \frac{1}{2}\epsilon$;

and $F - E_1 \subset \bigcup_{i=1}^{\infty} T_i - E_1$ so that $\mu(F - E_1) < \frac{1}{2}\epsilon$.

Hence $\mu(E \triangle F) = \mu(E_1 \triangle F) < \epsilon$. \blacksquare

Remark. The condition $\mu(E) < \infty$ cannot be omitted from the above theorem, since it is possible for a finite measure μ on \mathcal{R} to have an extension to \mathcal{S} which is σ -finite but not finite (for example, Lebesgue measure).

It is also worth noticing that the sets E of $\bar{\mathcal{S}}$ can be approximated exactly in measure by sets in \mathcal{S} , by theorem 4.3. We noticed earlier that the outer measure μ^* generated by the process of theorem 4.2 is always regular. This means that an arbitrary set $E \subset \Omega$ is always contained in a set $F \in \mathcal{S}$ for which $\mu^*(E) = \mu(F)$, so that every set can be approximated from the outside by a set of \mathcal{S} of the same measure. If E is not μ^* -measurable (i.e. not in $\bar{\mathcal{S}}$) then two-sided approximation is not possible.

Up to the present we have only considered general approximation theorems valid in any abstract space. If the measure is defined in a topological space, then it is of interest to obtain approximation theorems which connect the measure properties to the topology of the space. We do not, however, discuss this problem in general: instead we consider Euclidean space with the usual topology, and Lebesgue measure.

Regular measure

Suppose \mathcal{S} is a σ -ring of subsets of a topological space Ω which includes the open and the closed subsets of Ω , and $\mu: \mathcal{S} \rightarrow \mathbf{R}^+$ is a measure. Then the measure μ is said to be regular if, for each $\epsilon > 0$,

- (i) given $E \in \mathcal{S}$, there is an open $G \supset E$ with $\mu(G - E) < \epsilon$;
- (ii) given $E \in \mathcal{S}$, there is a closed $F \subset E$ with $\mu(E - F) < \epsilon$.

Since the class \mathcal{B} of Borel sets in Ω is the σ -ring generated by the open sets, the condition that \mathcal{S} includes the open sets implies $\mathcal{S} \supset \mathcal{B}$. If μ is regular on \mathcal{S} , then $\mathcal{B} \supset \mathcal{S}$, where \mathcal{B} denotes the completion of \mathcal{S} with respect to μ ; for if δ_n is a sequence of positive numbers decreasing to zero one can find for any E in \mathcal{S} an open set G_n and a closed set F_n such that

$$\mu(G_n - F_n) < \delta_n \quad \text{and} \quad G_n \supset E \supset F_n,$$

and

$$G = \bigcap_{n=1}^{\infty} G_n, \quad F = \bigcup_{n=1}^{\infty} F_n$$

will then be Borel sets with $G \supset E \supset F$ and $\mu(G - F) = 0$.

Metric outer measure

An outer measure μ^* defined on a metric space Ω and such that μ^* is additive on separated sets, i.e.

$$d(E, F) > 0 \Rightarrow \mu^*(E \cup F) = \mu^*(E) + \mu^*(F),$$

is said to be a metric outer measure. It can be proved that, for any metric outer measure, the class \mathcal{M} of measurable sets contains the open sets (and therefore contains \mathcal{B}), and that, if μ^* is also σ -finite, the restriction of μ^* to \mathcal{M} is regular. Since Lebesgue measure is generated by a metric outer measure, this general theory would allow us to deduce that Lebesgue measure is regular. However, we prefer instead to prove the result only for the special case of Lebesgue measure.

Theorem 4.5. *Lebesgue k -dimensional measure, defined on the class \mathcal{L}^k of Lebesgue measurable sets in \mathbf{R}^k , is a regular measure.*

Proof. We give the details of the proof for $k = 1$; only obvious alterations are needed for general k . Suppose $E \in \mathcal{L} = \mathcal{L}^1$; then $E \cap [n, n+1] = E_n \in \mathcal{L}$ for every integer n , and $\mu(E_n) \leq 1 < \infty$. By the construction of theorem 4.2, there is a countable covering $\{C_{ni}\}$ of E_n by $\frac{1}{2}$ -open intervals of \mathcal{P} such that

$$\mu(E_n) + \frac{1}{4} \frac{\epsilon}{2^{|n|}} > \sum_{i=1}^{\infty} \mu(C_{ni}).$$

Enlarge each of these intervals C_{ni} to an open interval G_{ni} such that

$$\mu(G_{ni} - C_{ni}) < \frac{1}{4} \frac{\epsilon}{2^{|n|+i}}.$$

Then $Q_n = \bigcup_{i=1}^{\infty} G_{ni}$ is an open set which contains E_n and satisfies

$$\mu(Q_n - E_n) < \frac{1}{2} \frac{\epsilon}{2^{|n|}}.$$

If we now put $Q = \bigcup_{n=-\infty}^{\infty} Q_n$, then Q is open, $Q \subset E$, and $\mu(Q - E) < \epsilon$.

This proves condition (i) for regularity.

For any $E \in \mathcal{L}$, $\Omega - E \in \mathcal{L}$, and we can apply the above argument to obtain an open $R \supset \Omega - E$ such that $\mu(R - (\Omega - E)) < \epsilon$. Then $F = \Omega - R$ is closed, $F \subset E$ and $\mu(E - F) = \mu(R \cap E) < \epsilon$, so that the second condition for regularity is also satisfied. **■**

Corollary. *Given any set $E \in \mathcal{L}^k$, there is a \mathcal{G}_δ -set Q and an \mathcal{F}_σ -set R such that*

$$Q \supset E \supset R \quad \text{and} \quad \mu(Q - R) = 0.$$

Proof. Note that \mathcal{G}_δ and \mathcal{F}_σ sets were defined in §2.5. For each integer n , take an open set $G_n \supset E$ and a closed set $F_n \subset E$ such that

$$\mu(G_n - E) < \frac{1}{n}, \quad \mu(E - F_n) < \frac{1}{n}.$$

The sets

$$Q = \bigcap_{n=1}^{\infty} G_n \quad \text{and} \quad R = \bigcup_{n=1}^{\infty} F_n$$

then satisfy the conditions of the corollary. **■**

This corollary strengthens the result that any set in \mathcal{L}^k can be approximated exactly in measure by a set in \mathcal{B}^k —which follows from the fact that \mathcal{L}^k is the completion of \mathcal{B}^k with respect to Lebesgue measure.

Exercises 4.3

1. Suppose \mathcal{Q} is the σ -ring generated by a ring \mathcal{R} and μ, ν are two σ -finite measures on \mathcal{R} . Show that if $E \in \mathcal{Q}$ is such that both $\mu(E)$, $\nu(E)$ are finite then, for any $\epsilon > 0$, there is a set $E_0 \in \mathcal{R}$ for which

$$\mu(E \Delta E_0) < \epsilon, \quad \nu(E \Delta E_0) < \epsilon.$$

2. Suppose Ω is a metric space and μ^* is an outer measure on Ω such that every Borel set is μ^* -measurable. Show that μ^* is a metric outer measure, i.e. that for $E_1, E_2 \subset \Omega$,

$$d(E_1, E_2) > 0 \Rightarrow \mu^*(E_1 \cup E_2) = \mu^*(E_1) + \mu^*(E_2).$$

Hint. Take an open set $G \supset E_1$, $G \cap E_2 = \emptyset$ and use $E_1 \cup E_2$ as a test set for the measurability of G .

3. Suppose μ^* is a metric outer measure on a metric space Ω . Show that if E is a subset of an open subset G and $E_n = E \cap \{x: d(x, \Omega - G) \geq 1/n\}$ then $\lim_{n \rightarrow \infty} \mu^*(E_n) = \mu^*(E)$.

Hint. $\{E_n\}$ is a monotonic increasing sequence of sets whose limit is E . Put $E_0 = \emptyset$, $D_n = E_{n+1} - E_n$ and notice that if neither D_{n+1} nor E_n is empty then $d(D_{n+2}, D_n) > 0$ so that

$$\mu^*(E_{2n+1}) \geq \sum_{i=1}^n \mu^*(D_{2i}), \quad \mu^*(E_{2n}) \geq \sum_{i=1}^n \mu^*(D_{2i-1}).$$

If either of these series diverges, then $\mu^*(E_n) \rightarrow \infty = \mu^*(E)$. If both converge, use

$$\mu^*(E) \leq \mu^*(E_{2n}) + \sum_{i=n}^{\infty} \mu^*(D_{2i}) + \sum_{i=n}^{\infty} \mu^*(D_{2i+1}).$$

4. If μ^* is a metric outer measure, show that all open sets (and therefore all Borel sets) are μ^* -measurable.

Hint. If G is open, A any subset, use notation of (3) applied to $E = A \cap G$. Then $d(E_n, A \cap -G) > 0$ so

$$\mu^*(A) \geq \mu^*\{E_n \cup (A \cap -G)\} = \mu^*(E_n) + \mu^*(A \cap -G).$$

4.4* Geometrical properties of Lebesgue measure

We have now defined Lebesgue measure in Euclidean space and considered some of its measure-theoretic properties. However, the justification for studying Lebesgue measure is that it makes precise our intuitive notion of length, area, volume in Euclidean space and generalises these notions to sets where our intuition breaks down. In the present section we want to show that Lebesgue measure has got the properties which geometrical intuition would lead us to expect.

It is convenient to adopt the notation $|E|$ for the Lebesgue measure of any set $E \in \mathcal{L}^k$, so that for sets $E \in \mathcal{L}^1$, $|E|$ is a generalisation of length; for sets $E \in \mathcal{L}^2$, $|E|$ is a generalisation of area; for sets $E \in \mathcal{L}^k$ ($k \geq 3$), $|E|$ is a generalisation of volume.

Since the set consisting of a single point x can be enclosed in an interval of \mathcal{P} of arbitrarily small length, it follows that

$$|\{x\}| = 0 \quad \text{for } x \in \mathbf{R}^k.$$

In particular, in \mathbf{R}^1 ,

$$|[a, b]| = |(a, b)| = |(a, b]| = |[a, b) = b - a$$

so that the Lebesgue measure of any interval on the line is its length. Any countable set in \mathbf{R}^k is the union of its single points, and is therefore

of zero measure. In particular the set of points in \mathbf{R}^k with rational coordinates forms a set of zero measure (even though this set is dense in the whole space).

In \mathbf{R}^k ($k \geq 2$), any segment of length l of a straight line can be covered by $[nl] + 1$ cubes of \mathcal{P}^k of side $1/n$ so that the Lebesgue measure of such a segment must be less than $([x])$ ($[x]$ denotes the largest integer not greater than x)

$$\left(\frac{1}{n}\right)^k \{[nl] + 1\} = O\left(\frac{1}{n^{k-1}}\right) \quad \text{as } n \rightarrow \infty,$$

and so $|L| = 0$ for any segment L of finite length. Any infinite straight line in \mathbf{R}^k , $k \geq 2$, is the countable union of segments of finite length so that $|L| = 0$ for any straight line L in \mathbf{R}^k ($k \geq 2$). It follows that, if we are calculating the measure of any geometrical figure in the plane which is bounded by a countable collection of straight lines, then the area will be the same whether all, some or none of the boundary lines are included in the set.

The above argument shows that there are sets E in \mathbf{R}^k ($k \geq 2$) which are not countable, but such that $|E| = 0$. The question arises whether or not such sets exist in \mathbf{R}^1 . This is easily answered by the Cantor set

$$C = \bigcap_{n=0}^{\infty} F_n,$$

defined in §2.7 where $F_0 = [0, 1]$ and F_n is obtained from F_{n-1} by replacing each closed interval of F_{n-1} by two closed intervals obtained by removing an open interval of one third its length from the centre. We proved that C was perfect and therefore non-countable. But

$$|F_n| = \frac{2}{3} |F_{n-1}| = \left(\frac{2}{3}\right)^n |F_0| = \left(\frac{2}{3}\right)^n,$$

so that

$$|C| = \lim_{n \rightarrow \infty} |F_n| = 0.$$

It is worth remarking that it is also possible for perfect nowhere dense sets in \mathbf{R} to have positive measure—see exercises 4.4 (2, 3).

We now consider what happens to the Lebesgue measure of sets under elementary transformations of the space.

(i) *Translation*

Suppose $x \in \mathbf{R}^k$ and $E \subset \mathbf{R}^k$. Put

$$E(x) = \{z: z = x + y, \quad y \in E\}.$$

For the intervals $I \in \mathcal{P}^k$, it is immediate that

$$|I(x)| = |I|$$

so that the outer measure μ^* is invariant under translations, and Lebesgue measure must therefore also be invariant provided measurability is preserved. Suppose $E \in \mathcal{L}^k$, and A is a test set for $E(\mathbf{x})$. Then since E is measurable, using $A(-\mathbf{x})$ as a test set,

$$\mu^*(A(-\mathbf{x})) = \mu^*(A(-\mathbf{x}) \cap E) + \mu^*(A(-\mathbf{x}) - E)$$

so that $\mu^*(A) = \mu^*(A \cap E(\mathbf{x})) + \mu^*(A - E(\mathbf{x}))$
and $E(\mathbf{x})$ must also be measurable.

(ii) *Reflexion in a plane perpendicular to an axis*

(For $k = 1$ this means reflexion in a point, for $k = 2$ this means reflexion in a line parallel to an axis.) It is clear that μ^* is invariant under such a reflexion because the reflexion of the covering sets of \mathcal{P}^k again gives $\frac{1}{2}$ -open intervals of the same measure. A similar argument to that used in (i) shows that measurability is preserved, so that Lebesgue measure is invariant under such reflexions.

(iii) *Uniform magnification*

For $p > 0$, the transformation of \mathbf{R}^k obtained by putting $\mathbf{y} = p\mathbf{x}$ for all $\mathbf{x} \in \mathbf{R}^k$ will be called a magnification by the factor p , and pE denotes the result of applying this magnification to the set E . If $I \in \mathcal{P}^k$, then it is clear that

$$pI \in \mathcal{P}^k \quad \text{and} \quad |pI| = p^k |I|.$$

Hence, if μ^* denotes the outer measure generated by Lebesgue measure on \mathcal{P}^k ,

$$\mu^*(pE) = p^k \mu^*(E)$$

for all sets E . A similar argument to that used in (i) shows that measurability is preserved by magnification, so that if E is Lebesgue measurable, so is pE and $|pE| = p^k |E|$.

(iv) *Rotation about the origin*

Lebesgue measure is invariant in this case also, but rather more work is needed to prove it. The key idea needed for the proof is that an open sphere centre O is invariant under rotation about O . Suppose I is a fixed interval of \mathcal{P}^k

$$I = \{\mathbf{x} : a_i < x_i \leq b_i, \quad i = 1, 2, \dots, k\}.$$

Then for any $\mathbf{x} \in \mathbf{R}^k$ ($p > 0$), $(pI)(\mathbf{x})$ is an interval of \mathbf{R}^k similar, and similarly situated to I . If χ denotes the transformation of \mathbf{R}^k consisting of a fixed rotation about O , then

$$\chi(pI)(\mathbf{x}) = (p\chi I)(\chi\mathbf{x}).$$

By (i) and (ii)

$$|\chi(pI)(x)| = p^k |\chi I|, \quad |(pI)(x)| = p^k |I|,$$

so that

$$|\chi(pI)(x)| = \frac{|\chi I|}{|I|} |(pI)(x)|$$

for all $p > 0$, $x \in \mathbf{R}^k$. This means that, for a given χ and I , the effect on the measure is the same for all intervals of the form $(pI)(x)$.

Now any open set G can be expressed as a countable union of disjoint sets of the form $(pI)(x)$. In particular the unit open sphere S centre the origin, can be expressed this way

$$S = \bigcup_{i=1}^{\infty} (p_i I)(x_i),$$

and

$$|S| = \sum_{i=1}^{\infty} |(p_i I)(x_i)|.$$

But $\chi S = S$, so that

$$\sum_{i=1}^{\infty} |(p_i I)(x_i)| = |S| = |\chi S| = \sum_{i=1}^{\infty} |\chi(p_i I)(x_i)| = \frac{|\chi I|}{|I|} \sum_{i=1}^{\infty} |(p_i I)(x_i)|,$$

which implies that $|\chi I| = |I|$. This argument is valid for any interval $I \in \mathcal{P}^k$.

We can now use arguments similar to those in (i) to show that, for any set $E \subset \mathbf{R}^k$

$$\mu^*(\chi E) = \mu^*(E)$$

and measurability is preserved under χ . Thus if $E \in \mathcal{L}^k$, χE is also in \mathcal{L}^k and

$$|\chi E| = |E|.$$

Note finally that reflexion in an arbitrary plane can be obtained by successively applying the operations (iv), (ii), (i), (iv). We have thus proved

Theorem 4.6. *The class \mathcal{L}^k of Lebesgue measurable subsets of \mathbf{R}^k , and Lebesgue measure on \mathcal{L}^k are invariant under translations, reflexions and rotations. If E and F are two subsets of \mathbf{R}^k which are congruent in the sense of Euclid and E is measurable, then so is F and*

$$|E| = |F|.$$

For $p > 0$, if pE denotes the set of vectors x of the form py , $y \in E$, then $E \in \mathcal{L}^k \Rightarrow pE \in \mathcal{L}^k$, and $|pE| = p^k |E|$.

If k, l, r are positive integers and $k + l = r$, then the Euclidean space \mathbf{R}^r can be thought of as a Cartesian product $\mathbf{R}^k \times \mathbf{R}^l$. We have defined Lebesgue measure independently in each dimension, but the

measure of the primary sets \mathcal{P}^r could have been obtained as a product of the measures of corresponding sets in $\mathcal{P}^k, \mathcal{P}^l$. It is therefore not surprising that this is true of a wider class of sets.

Theorem 4.7. *If $E \in \mathcal{L}^k, F \in \mathcal{L}^l$ then the Cartesian product $E \times F \in \mathcal{L}^{k+l}$ and*

$$|E \times F| = |E| \cdot |F|.$$

Proof. We use μ^* to denote the outer measure generated by Lebesgue measure in the space where the set lies. Suppose first that E, F are bounded so that there are finite open intervals J, K such that $E \subset J, F \subset K$. We can then cover E and F by countable collections of open intervals such that

$$\begin{aligned} E &\subset \bigcup_{i=1}^{\infty} Q_i \subset J, & F &\subset \bigcup_{j=1}^{\infty} R_j \subset K, \\ \sum_{i=1}^{\infty} |Q_i| &< |E| + \epsilon, & \sum_{j=1}^{\infty} |R_j| &< |F| + \epsilon. \end{aligned}$$

Then $E \times F \subset \bigcup_{i,j} Q_i \times R_j$, so that

$$\begin{aligned} \mu^*(E \times F) &\leq \sum_{i,j} |Q_i \times R_j| = \sum_{i,j} |Q_i| |R_j| \\ &= \sum_{i=1}^{\infty} |Q_i| \sum_{j=1}^{\infty} |R_j| < (|E| + \epsilon)(|F| + \epsilon). \end{aligned}$$

Since ϵ is arbitrary, it follows that

$$\mu^*(E \times F) \leq |E| \cdot |F|. \quad (4.4.1)$$

But

$$J \times K = E \times F \cup (J - E) \times F \cup E \times (K - F) \cup (J - E) \times (K - F),$$

and the subadditivity of μ^* gives, with (4.4.1),

$$\mu^*(J \times K) \leq |E| \cdot |F| + |J - E| \cdot |F| + |E| \cdot |K - F| + |J - E| \cdot |K - F|.$$

But $J \times K$ is an open rectangle and therefore in \mathcal{L}^{k+l} , and

$$\mu^*(J \times K) = |J| \cdot |K| = (|E| + |J - E|)(|F| + |K - F|).$$

It follows that all the inequalities of type (4.4.1) must be equalities. In particular

$$\mu^*(E \times F) = |E| \cdot |F|. \quad (4.4.2.)$$

By the corollary to theorem 4.5, we can find sequences $\{A_n\}, \{B_n\}$ of disjoint closed sets such that

$$\begin{aligned} A &= \bigcup_{n=1}^{\infty} A_n \subset E, & B &= \bigcup_{m=1}^{\infty} B_m \subset F, \\ |E - A| &= 0, & |F - B| &= 0. \end{aligned}$$

Since $A \times B$ is an \mathcal{F}_σ -set in \mathbf{R}^{k+l} it is measurable and

$$\mu^*(A \times B) = |A \times B| = |A| \cdot |B| = |E| \cdot |F|.$$

But $A \times B \subset E \times F$, and Lebesgue measure is complete so that we must have $E \times F$ measurable and

$$|E \times F| = \mu^*(E \times F) = |E| \cdot |F|.$$

In order to remove the restriction of boundedness, apply the above to $E \cap S_n, F \cap S'_n$, where S_n, S'_n are spheres of radius n centre the origin in k -space, l -space respectively. This shows that, for each n ,

$$(E \cap S_n) \times (F \cap S'_n) \in \mathcal{L}^{k+l},$$

$$|(E \cap S_n) \times (F \cap S'_n)| = |E \cap S_n| |F \cap S'_n|$$

and the result follows from the continuity of measures on letting $n \rightarrow \infty$. \blacksquare

Non-measurable sets

We have now seen that Lebesgue measure can be defined on \mathcal{L}^k , a large class of subsets of \mathbf{R}^k , in such a way as to preserve the intuitive geometrical ideas of volume. We also remarked earlier that it is impossible to define such a measure on all subsets of \mathbf{R}^k , so we now demonstrate the existence of at least one subset which is not in \mathcal{L}^k . Again we carry out the construction for $k = 1$. Consider subsets $E \subset (0, 1]$ and for $x \in (0, 1]$ let $E(x)$ be the set of real numbers z such that

$$z = x + y, \quad y \in E \quad \text{and} \quad x + y \leq 1,$$

or

$$z = x + y - 1, \quad y \in E \quad \text{and} \quad x + y > 1;$$

that is, $E(x)$ is the result of translating E a distance x and then taking the non-integer part. From property (i), it follows immediately that

$$E \in \mathcal{L} \Rightarrow E(x) \in \mathcal{L}, \quad |E| = |E(x)|.$$

Now let Z be the set of rationals in $(0, 1]$. Two sets $Z(x_1), Z(x_2)$ will be disjoint if $(x_1 - x_2)$ is irrational and identical if $(x_1 - x_2)$ is rational. Let \mathcal{C} be the class of disjoint sets of the form $Z(x)$. By the axiom of choice (see §1.6) there is a set T containing precisely one point from each of the sets in \mathcal{C} . If Z is the set (r_1, r_2, \dots) , we put

$$Q_n = T(r_n) \quad (n = 1, 2, \dots).$$

Then

$$\bigcup_{n=1}^{\infty} Q_n = (0, 1],$$

since every point $x \in (0, 1]$ is in $Z(x_1)$ for some x_1 and if $q \in Z(x_1) \cap T$, we have $q - x_1 = r_n$ so that $x \in Q_n$. Also the sets Q_n are disjoint as T contains only one point from each set in \mathcal{C} and therefore cannot contain two points differing by a rational. If $T \in \mathcal{L}$, then $Q_n \in \mathcal{L}$ ($n = 1, 2, \dots$) and

$$|T| = |Q_n| \quad (n = 1, 2, \dots).$$

But then

$$1 = |(0, 1]| = \sum_{n=1}^{\infty} |Q_n|$$

and this equation cannot possibly be satisfied either by $|Q_n| = 0$ or $|Q_n| = c > 0$ for all n . The only possibility is that the set T is not measurable.

It is worth remarking that there are many more Lebesgue sets than there are Borel sets. The number of sets in \mathcal{L}^k is not more than the number of subsets of \mathbf{R}^k , i.e. not more than 2^c . However it is at least 2^c for it contains all subsets of the Cantor set (perfect with c points in it), so that the cardinality of \mathcal{L}^k must be 2^c . However the cardinality of the class \mathcal{B}^k of Borel subsets of \mathbf{R}^k is c and $c < 2^c$ (see §1.3) so that there must be some sets which are in \mathcal{L}^k but not in \mathcal{B}^k ; this means that the class \mathcal{B}^k is not complete with respect to Lebesgue measure. In order actually to exhibit a set in \mathcal{L}^k but not in \mathcal{B}^k one has to work a bit harder so we do not include such an example.

Exercises 4.4

1. Show that the set of points in $[0, 1]$ whose binary expansion has zero in all the even places is a Lebesgue measurable set of zero measure. Is it a Borel set?

2. By changing the lengths of the extracted intervals in the construction of the Cantor set, show how to obtain a nowhere dense perfect set of measure $\frac{1}{2}$.

3. Generalise (2) to show that for any $\epsilon > 0$ there is a nowhere dense, perfect subset of $[0, 1]$ with measure greater than $1 - \epsilon$.

4. Consider a union of sets of (3) to obtain a subset of $[0, 1]$ of full measure which is of the first category, and another subset of $[0, 1]$ of zero measure which is of the second category.

5. Show that any bounded set in Euclidean space \mathbf{R}^k has finite Lebesgue outer measure. Is the converse of this statement true?

6. Suppose X is the circumference of a unit circle in \mathbf{R}^2 . Show that there is a unique measure μ defined on Borel subsets of X such that $\mu(X) = 1$ and μ is invariant under all rotations of X into itself.

7. By considering suitable approximating polygons (finite unions of rectangles will do), show that the area of the plane region bounded by $x = 1$, $y = 0$, $y = x^3$ is $\frac{1}{4}$. Generalise to the case $y = x^k$, where $k > 0$ but need not be an integer.

8. Show that a subset E of a bounded interval $I \subset \mathbf{R}^k$ is measurable if, for any $\epsilon > 0$ there are elementary figures $Q_1, Q_2 \in \mathcal{E}^k$ such that $Q_1 \supset E$, $Q_2 \supset I - E$ and

$$|Q_1| + |Q_2| < |I| + \epsilon.$$

9. Suppose X is the unit square $\{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$. If $E \subset [0, 1]$ put $\hat{E} = \{(x, y): x \in E, 0 \leq y \leq 1\}$ and let \mathcal{S} be the class of sets \hat{E} such that E is \mathcal{L}^1 -measurable. Put $\mu(\hat{E}) = |E|$, and show that the subset $M = \{(x, y): 0 \leq x \leq 1, y = \frac{1}{2}\}$ is not measurable with respect to the outer measure μ^* generated by μ on the class of all subsets of X . Show that

$$\mu^*(M) = 1, \quad \mu^*(X - M) = 1.$$

4.5 Lebesgue-Stieltjes measure

There are other measures in \mathbf{R}^k which are of importance in probability theory. Suppose $F: \mathbf{R} \rightarrow \mathbf{R}$ is a monotone increasing real valued function of a real variable which is everywhere continuous on the right. Such a function is called a *Stieltjes measure function*. Put

$$\mu_F(a, b] = F(b) - F(a)$$

for each $(a, b] \in \mathcal{P}$. Then μ_F is non-negative and (finitely) additive on \mathcal{P} —the proof used for the length function in §3.4 can be easily adapted to show this (the length function corresponds to $F(x) = x$). By applying theorem 3.4 we can extend μ_F uniquely to an additive set function on \mathcal{E} , the ring of elementary figures. As in §3.4 we again have at least two methods of showing that μ_F is a measure on \mathcal{E} . By theorem 3.2(iii) if μ_F is not a measure, then there is a monotone decreasing sequence $\{E_n\}$ of sets of \mathcal{E} such that $\lim E_n = \emptyset$, but $\lim \mu_F(E_n) = \delta > 0$. The argument used in the Lebesgue case for $k = 1$ can be modified by using the fact that, for any $\epsilon > 0$, if

$$\mu_F(a, b] > 0,$$

we can always find a $y > 0$ such that

$$(a + y, b] \subset [a + y, b] \subset (a, b]$$

and

$$\mu_F(a + y, b] > \mu_F(a, b] - \epsilon,$$

since F is continuous on the right at a . This leads us to a contradiction which establishes that μ_F is a measure on \mathcal{E} .

For $k \geq 2$, we must start with a function $F: \mathbf{R}^k \rightarrow \mathbf{R}$ which is con-

tinuous on the right in each variable separately and such that, for $I \in \mathcal{P}^k$,

$$\mu_F(I) = \sum_{i=1}^{2k} y_i F(V_i) \geq 0, \quad (4.5.1)$$

where V_i are the 2^k vertices of the set $I \in \mathcal{P}^k$ and $y_i = +1$ for the vertex in which each co-ordinate is largest and $y_i = (-1)^r$ if the vertex V_i is such that r of its coordinates are at the lower bound (and $(k-r)$ at the upper bound). Any such function F is called a *k-dimensional Stieltjes measure function*. With a little care it is not difficult to show that, under these conditions, μ_F is a non-negative additive set function on \mathcal{P}^k and that it therefore has a unique extension to \mathcal{E}^k . Either of the arguments given in §3.4 can now be modified to show that μ_F is a measure on \mathcal{E}^k .

We can now apply theorem 4.2 to this measure μ_F to extend it to the σ -ring \mathcal{B}^k of Borel sets in \mathbf{R}^k . As in the case of Lebesgue measure, this extension automatically defines μ_F on the completion \mathcal{L}_F^k of \mathcal{B}^k with respect to μ_F . The class \mathcal{L}_F^k is called the class of sets which are *Lebesgue-Stieltjes measurable* for the function F . The class clearly depends on the function F —for in the particular case $F \equiv c$, \mathcal{L}_F^k contains all subsets of \mathbf{R}^k as $\mu_F(\mathbf{R}^k) = 0$ and μ_F is complete; while if $F(x_1, x_2, \dots, x_k) = x_1 x_2 \dots x_k$, then μ_F is the length function and \mathcal{L}_F^k is the Lebesgue class \mathcal{L}^k .

Each of these measures $\mu_F: \mathcal{L}_F^k \rightarrow \mathbf{R}^+$ is regular. The proof given in theorem 4.5 can easily be modified to show this (we again do the case $k = 1$) by using the fact that, for any $\epsilon > 0$, if $(a, b] \in \mathcal{P}$, there is a $y > 0$ such that

$$(a, b + y] \supset (a, b + y) \supset (a, b]$$

and

$$\mu_F(a, b + y) \leq \mu_F(a, b + y] < \mu_F(a, b] + \epsilon,$$

to obtain economical coverings by open intervals.

Probability measure

Given a σ -field \mathcal{F} of subsets of Ω , any measure $\mathbf{P}: \mathcal{F} \rightarrow \mathbf{R}^+$ such that $\mathbf{P}(\Omega) = 1$ is called a *probability measure* on \mathcal{F} . If in addition \mathcal{F} is complete with respect to \mathbf{P} we will say that the triple $(\Omega, \mathcal{F}, \mathbf{P})$ form a *probability space*.

Distribution function

A function $F: \mathbf{R} \rightarrow \mathbf{R}$ is called a *distribution function* if

- (i) F is monotonic increasing, continuous on the right;
- (ii) $F(x) \rightarrow 0$ as $x \rightarrow -\infty$, $F(x) \rightarrow 1$ as $x \rightarrow +\infty$.

A function $F: \mathbf{R}^k \rightarrow \mathbf{R}$ is called a (k -dimensional) distribution function if

- (i) F is continuous on the right in each variable;
- (ii) $\mu_F(I) \geq 0$ for all $I \in \mathcal{P}^k$, where μ_F is defined by (4.5.1),
- (iii) $F(x_1, x_2, \dots, x_k) \rightarrow 0$ as any one of $x_1, x_2, \dots, x_k \rightarrow -\infty$,

$$F(x_1, x_2, \dots, x_k) \rightarrow 1 \quad \text{as } x_1, x_2, \dots, x_k \text{ all } \rightarrow +\infty.$$

It is immediate from our definitions that any distribution function F can be used to define a Lebesgue-Stieltjes measure μ_F on the σ -field \mathcal{L}_F^k . Further $\mu_F(\mathbf{R}^k) = 1$ and μ_F is complete, so that every distribution function determines a probability measure and $(\mathbf{R}^k, \mathcal{L}_F^k, \mu_F)$, is a probability space. There is a sense in which these are the only interesting probability measures on \mathbf{R}^k .

Theorem 4.8. *Suppose \mathcal{S} is a σ -field of sets in \mathbf{R} , \mathcal{S} contains the open sets and $\mu: \mathcal{S} \rightarrow \mathbf{R}^+$ is a complete measure which is finite on bounded sets in \mathcal{S} . Then there is a Stieltjes measure function $F: \mathbf{R} \rightarrow \mathbf{R}$ such that $\mathcal{S} \supset \mathcal{L}_F$ and μ coincides with μ_F on \mathcal{L}_F . If $(\mathbf{R}, \mathcal{S}, \mu)$ is a probability space, then F can be chosen to be a distribution function.*

Proof. Since \mathcal{S} contains the open sets and is a σ -field, it must contain \mathcal{B} , the Borel sets and in particular $\mathcal{S} \supset \mathcal{P}$, the class of half-open intervals. Define F by

$$F(x) = \begin{cases} \mu(0, x] & \text{for } x \geq 0, \\ -\mu(x, 0] & \text{for } x < 0. \end{cases}$$

Then $F: \mathbf{R} \rightarrow \mathbf{R}$ is clearly defined and is monotonic increasing for all real x (note that $F(0) = 0$). By theorem 3.2 (i), if $\{x_n\}$ is any monotonic sequence decreasing to x , $\lim_{n \rightarrow \infty} F(x_n) = F(x)$; since

$$\text{if } x \geq 0, \quad \lim_{n \rightarrow \infty} (0, x_n] = (0, x],$$

$$\text{if } x < 0, \quad \lim_{n \rightarrow \infty} (x_n, 0] = (x, 0].$$

Thus F is continuous on the right, and must therefore be a Stieltjes measure function.

$$\text{Now if } a \geq 0, \quad \mu(a, b] = \mu(0, b] - \mu(0, a] = F(b) - F(a);$$

$$\text{if } a < 0 \leq b, \quad \mu(a, b] = \mu(a, 0] + \mu(0, b] = F(b) - F(a);$$

$$\text{if } b < 0, \quad \mu(a, b] = \mu(a, 0] - \mu(b, 0] = F(b) - F(a);$$

so that μ coincides with μ_F on \mathcal{P} . By uniqueness of the extension of a measure to the generated σ -field and its completion, we have $\mu = \mu_F$ on \mathcal{L}_F and $\mathcal{S} \supset \mathcal{L}_F$.

If μ is a probability measure on \mathcal{S} , we must have

$$\lim_{x \rightarrow +\infty} F(x) - \lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow \infty} \mu(-n, n] = 1,$$

so that

$$F_1(x) = F(x) - \lim_{x \rightarrow -\infty} F(x)$$

will be a distribution function generating the same Stieltjes measure as F . \blacksquare

Remark. The case where μ is a probability measure could have been done directly by defining

$$F_1(x) = \mu(-\infty, x].$$

It is clear that this case extends immediately to \mathbf{R}^k since if we put

$$F(x_1, x_2, \dots, x_k) = \mu\{(\xi_1, \dots, \xi_k) : \xi_i \leq x_i, i = 1, \dots, k\}$$

it is easy to check that F is a k -dimensional distribution.

Discrete probability

There is a special case of a probability measure in which all the probability is concentrated on a countable set $E_0 \subset \Omega$. This can be defined by specialising example (6) of §3.1. If $\{x_n\}$ is any sequence in Ω , and $\{p_n\}$ is a sequence of positive real numbers with $\sum_{n=1}^{\infty} p_n = 1$, then it is clear that

$$P(E) = \sum_{x_n \in E} p_n$$

defines a probability measure on the class of all subsets of Ω . When $\Omega = \mathbf{R}$, this measure can be obtained from the distribution function

$$F(x) = \sum_{x_n \leq x} p_n$$

so that, in \mathbf{R} (or in \mathbf{R}^k for that matter) a discrete probability measure can be expressed as the Lebesgue–Stieltjes measure of a suitable distribution function.

Exercises 4.5

1. To see that condition (4.5.1) for k -dimensional Stieltjes measure functions is not implied by the condition that F be monotonic increasing in each variable separately consider

$$F(x_1, x_2) = \begin{cases} \max(0, x_1 + x_2 + 1) & \text{for } x_1 + x_2 < 0, \\ 1 & \text{for } x_1 + x_2 \geq 0. \end{cases}$$

Does this condition (4.5.1) imply that F is monotonic in each variable?

2. If $F: \mathbf{R} \rightarrow \mathbf{R}$ is a Stieltjes measure function, show that

$$\mu_F(a, b) = F(b-0) - F(a), \quad \mu_F[a, b] = F(b) - F(a-0)$$

and determine μ_F for intervals of the form

$$[a, b), \quad (-\infty, a), \quad (a, \infty).$$

3. If F is a Stieltjes measure function in \mathbf{R} which generates the Stieltjes measure μ_F , show that $F(x)$ is continuous if and only if $\mu_F\{x\} = 0$ for all single point sets $\{x\}$. What is the corresponding continuity condition in \mathbf{R}^k ?

4. Consider Lebesgue measure on \mathcal{L}^1 -subsets of $[0, 1]$ and let E_0 be a subset of $[0, 1]$ which is non-measurable, such that the Lebesgue outer measure of E_0 and $([0, 1] - E_0)$ are both 1. Let \mathcal{Q} be the smallest σ -field of subsets of $[0, 1]$ containing E_0 and \mathcal{L}^1 . Show that \mathcal{Q} consists of sets of the form

$$E = A \cap E_0 + B \cap ([0, 1] - E_0)$$

for $A, B \in \mathcal{L}^1$ and that $\mu(E) = |A \cap [0, 1]|$ defines a probability measure on the σ -field \mathcal{Q} . By applying theorem 4.8 to this probability measure show that, in general it is not possible to deduce in theorem 4.8 that $\mathcal{S} = \mathcal{L}_F$.

5. Suppose

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 & \text{for } x \geq 0. \end{cases}$$

Show that

$$\mu_F(-1, 0) < F(0) - F(-1).$$

6. Give an example of a right-continuous monotone F such that

$$\mu_F(a, b) < F(b) - F(a) < \mu_F[a, b].$$

7. Show that, if F, G are distribution functions in \mathbf{R}^k , then $aF + bG$ is a distribution function for any $a \geq 0, b \geq 0, a + b = 1$.

8. In \mathbf{R}^2 ,

$$F(x_1, x_2) = \begin{cases} 1 & \text{for } x_1 \geq 0, \quad x_2 \geq 0, \\ 0 & \text{for all other points.} \end{cases}$$

Show that this F is a distribution function describing a unit mass at 0.

9. State and prove an n -dimensional form of theorem 4.8.

10. We can obtain completely additive set functions in \mathbf{R}^1 which are not necessarily non-negative by the following method. Suppose $F: \mathbf{R} \rightarrow \mathbf{R}$ is continuous on the right everywhere and of bounded variation in each finite interval and $F(b) - F(a)$ is bounded below for all $a < b$ and define

$$\tau_F(a, b] = F(b) - F(a).$$

Show that τ_F is additive on \mathcal{P} and can be extended to \mathcal{E} . By an extension of theorem 4.2, τ_F can then be extended to a σ -additive set function on \mathcal{B} . Now apply theorem 3.3 to express τ_F as the difference of two measures. Finally, the argument of theorem 4.8 shows that τ_F is the difference of two Stieltjes measures.

DEFINITIONS AND PROPERTIES OF THE INTEGRAL

5.1 What is an integral?

Historically the concept of integration was first considered for real functions of a real variable where either the notion of 'the process inverse to differentiation' or the notion of 'area under a curve' was the starting point. In the first case a real number was obtained as the difference of two values of the 'indefinite' integral, while the second case corresponds immediately to the 'definite' integral. The so-called 'fundamental theorem of the integral calculus' provided the link between the two ideas. Our discussion of the operation of integration will start from the notion of a definite integral, though in the first instance the 'interval' over which the function is integrated will be the whole space. Thus, for 'suitable' functions $f: \Omega \rightarrow \mathbf{R}^*$ we want to define the integral $\mathcal{J}(f)$ as a real number. The 'suitable' functions will be called integrable and $\mathcal{J}(f)$ will be called the integral of f .

Before defining such an operator \mathcal{J} , we examine the sort of properties \mathcal{J} should have before we would be justified in calling it an 'integral'. Suppose then that \mathcal{A} is a class of functions $f: \Omega \rightarrow \mathbf{R}^*$, and $\mathcal{J}: \mathcal{A} \rightarrow \mathbf{R}$ defines a real number for every $f \in \mathcal{A}$. Then we want \mathcal{J} to satisfy:

- (i) $f \in \mathcal{A}, f(x) \geq 0$ all $x \in \Omega \Rightarrow \mathcal{J}(f) \geq 0$, that is \mathcal{J} preserves positivity;
- (ii) $f, g \in \mathcal{A}, \alpha, \beta \in \mathbf{R} \Rightarrow \alpha f + \beta g \in \mathcal{A}$ and

$$\mathcal{J}(\alpha f + \beta g) = \alpha \mathcal{J}(f) + \beta \mathcal{J}(g),$$

that is \mathcal{J} is linear on \mathcal{A} ;

(iii) \mathcal{J} is continuous on \mathcal{A} in some sense, at least we would want to have $\mathcal{J}(f_n) \rightarrow 0$ as $n \rightarrow \infty$ for any sequence $\{f_n\}$ of functions in \mathcal{A} which is monotone decreasing with $f_n(x) \rightarrow 0$ for all x in Ω .

These conditions are satisfied by the elementary integration process, but the Riemann integral does not satisfy the following strengthened form of (iii):

- (iii)* If $\{f_n\}$ is an increasing sequence of functions in \mathcal{A} , and

$$f_n(x) \rightarrow f(x) \quad \text{for all } x \in \Omega,$$

then $f \in \mathcal{A}$ and $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$ as $n \rightarrow \infty$.

This is the most serious limitation of the Riemann integral for, with this definition of integration, it is necessary in (iii)* to postulate $f_n(x) \rightarrow f(x)$ uniformly in x before one can conclude that $f \in \mathcal{A}$ and $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$. Now conditions about the continuity of \mathcal{J} are really essential if the operation is to be a useful tool in analysis—there would not be much of analysis left if one could not carry out at least sequential limiting operations. One of our main objectives, therefore, is to define an operator \mathcal{J} which satisfies (iii)*.

One method of studying integration theory (essentially due to P. J. Daniell) is to start with a restricted class \mathcal{A}_0 of functions with a simple structure, define $\mathcal{J}:\mathcal{A}_0 \rightarrow \mathbf{R}$ to satisfy (i), (ii) and (iii) and then extend \mathcal{A}_0 and the functional \mathcal{J} step by step until $\mathcal{J}:\mathcal{A} \rightarrow \mathbf{R}$ is defined on a sufficiently large class while (i), (ii) and (iii)* are satisfied. Using this approach one can deduce a measure on a suitable σ -ring of subsets of Ω by putting

$$\mu(E) = \mathcal{J}(\chi_E)$$

for those sets E for which $\chi_E \in \mathcal{A}$. Condition (i) then implies that μ is non-negative, condition (ii) that it is additive and condition (iii) that it is σ -additive provided the domain of definition is a ring. We will give details of this approach in §9.4, but for the present we will regard the measure as the primary concept and define the integral in terms of a given measure. We will, however, obtain an operator $\mathcal{J}:\mathcal{A} \rightarrow \mathbf{R}$ which has the above properties and moreover in defining \mathcal{J} we will continually have these desired properties in mind. Thus out of many possible ways of obtaining the integral starting from a measure, we choose the method of definition by limits of monotone sequences of 'simple' functions.

5.2 Simple functions; measurable functions

We now assume given $(\Omega, \mathcal{F}, \mu)$ where Ω is a space, \mathcal{F} a σ -field of subsets of Ω and μ a measure on \mathcal{F} . All the concepts we now define are relative to $(\Omega, \mathcal{F}, \mu)$. It is worth remarking that our definitions can be modified to apply to the case where \mathcal{F} is a σ -ring rather than a σ -field, but this results in additional complications in proofs. The additional labour involved does not seem justified for the small gain in generality.

Our object is to define an operation, called integration, having the properties discussed in §5.1 on a suitable class of functions $f:\Omega \rightarrow \mathbf{R}^*$. Ultimately we want this domain of definition for the integral to be as large as possible. In the present section we obtain the properties of certain classes of functions which will be important later.

Dissection

If $\Omega = \bigcup_{i=1}^n E_i$ and the sets E_i are disjoint, then E_1, E_2, \dots, E_n are said to form a (finite) dissection of Ω . They are said to form an \mathcal{F} -dissection if, in addition, $E_i \in \mathcal{F}$ ($i = 1, 2, \dots, n$).

Simple function

A function $f: \Omega \rightarrow \mathbf{R}$ is called \mathcal{F} -simple if it can be expressed as

$$f(x) = \sum_{i=1}^n c_i \chi_{E_i}(x),$$

where E_1, E_2, \dots, E_n form an \mathcal{F} -dissection of Ω and

$$c_i \in \mathbf{R} \quad (i = 1, 2, \dots, n).$$

Thus an \mathcal{F} -simple function is one which takes a constant value c_i on the set E_i where the sets E_i are disjoint sets of \mathcal{F} . The additional condition implied by our definition that $\Omega = \bigcup_{i=1}^n E_i$ is not important (and is omitted by many authors), since if

$$E_{n+1} = \Omega - \bigcup_{i=1}^n E_i \neq \emptyset$$

we can always put $c_{n+1} = 0$ and write

$$f = \sum_{i=1}^{n+1} c_i \chi_{E_i}$$

to see that the function is \mathcal{F} -simple. If there is only one σ -field \mathcal{F} under consideration we will talk of *simple* functions rather than \mathcal{F} -simple functions.

Lemma. *The sum, difference and product of two simple functions is a simple function.*

Proof. Suppose we have the representations

$$f = \sum_{i=1}^n c_i \chi_{E_i}, \quad g = \sum_{j=1}^m d_j \chi_{A_j};$$

then the sets $H_{ij} = E_i \cap A_j$ ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) are in \mathcal{F} and form a dissection of Ω . Further

$$f(x) = c_i \quad \text{and} \quad g(x) = d_j \quad \text{for} \quad x \in H_{ij}, \quad \chi_{H_{ij}} = \chi_{E_i} \cdot \chi_{A_j}$$

so that $(f \pm g)(x) = c_i \pm d_j, \quad (fg)(x) = c_i d_j \quad \text{for} \quad x \in H_{ij}$

$$\text{and} \quad f \pm g = \sum_{i=1}^n \sum_{j=1}^m (c_i \pm d_j) \chi_{H_{ij}}, \quad fg = \sum_{i=1}^n \sum_{j=1}^m c_i d_j \chi_{H_{ij}}. \quad \square$$

Note that the constant functions

$$f(x) = c \quad \text{all } x \in \Omega$$

are simple, so that by this lemma it also follows that cf is simple if f is and the class of simple functions forms a linear space over the reals. One should regard simple functions as a generalisation of 'step' functions, but it is clear that they form a very restricted class since the image of Ω under a simple function is a finite subset of \mathbf{R} .

In defining measurability we will want to consider functions $f: \Omega \rightarrow \mathbf{R}^*$ with extended real values. It is possible to define a topology in \mathbf{R}^* and to define the class of Borel sets in \mathbf{R}^* in terms of this topology. However, we adopt the simpler procedure of defining the class \mathcal{B}^* of Borel sets in \mathbf{R}^* directly. We say that a set $B \subset \mathbf{R}^*$ is a Borel set in \mathbf{R}^* if it is the union of a set in \mathcal{B}^1 (the class of Borel sets in \mathbf{R}) with any subset of $\mathbf{R}^* - \mathbf{R} = \{-\infty, +\infty\}$.

Measurable function

A function $f: \Omega \rightarrow \mathbf{R}^*$ is said to be \mathcal{F} -measurable if and only if

$$f^{-1}(B) \in \mathcal{F}$$

for every $B \in \mathcal{B}^*$. If there is only one σ -field \mathcal{F} under discussion we may say that f is a measurable function.

From the definition it appears at first sight that one has to work hard to check that a given function is \mathcal{F} -measurable. However, in practice it is sufficient to check that $f^{-1}(E) \in \mathcal{F}$ for a suitable class of subsets which generates the σ -field \mathcal{B}^* . The most important such class is given by the next theorem.

Theorem 5.1. *In order that $f: \Omega \rightarrow \mathbf{R}^*$ be \mathcal{F} -measurable each of the following conditions is necessary and sufficient:*

- (i) $\{x: f(x) \leq c\} \in \mathcal{F}$ for all $c \in \mathbf{R}$;
- (ii) $\{x: f(x) > c\} \in \mathcal{F}$ for all $c \in \mathbf{R}$;
- (iii) $\{x: f(x) \geq c\} \in \mathcal{F}$ for all $c \in \mathbf{R}$;
- (iv) $\{x: f(x) < c\} \in \mathcal{F}$ for all $c \in \mathbf{R}$.

Proof (i). Since $[-\infty, c] \in \mathcal{B}^*$, it is clear that the given condition is necessary. If we suppose that the condition is satisfied, and put

$$E_c = \{x: f(x) \leq c\} = f^{-1}[-\infty, c],$$

then $E_c \in \mathcal{F}$, for all $c \in \mathbf{R}$. But the sets $I_c = [-\infty, c]$, $c \in \mathbf{R}$ generate the σ -field \mathcal{B}^* , so that, for each $B \in \mathcal{B}^*$ (see exercise 1.5.(10)), the set $f^{-1}(B)$

is in the σ -field of subsets of Ω generated by the sets E_c , $c \in R$. Since \mathcal{F} is a σ -field we have

$$f^{-1}(B) \in \mathcal{F}$$

for all $B \in \mathcal{B}^*$.

(ii), (ii) and (iv). A similar proof can be constructed for each case. Alternatively, it is easy to prove directly that each of (i), (ii), (iii), (iv) is equivalent to each of the others. \blacksquare

Corollary. Any \mathcal{F} -simple function is \mathcal{F} -measurable.

Proof. If

$$f = \sum_{i=1}^n c_i \chi_{E_i}, \quad \text{then} \quad E_c = \{x: f(x) \leq c\}$$

is the finite union of those sets E_i ($\in \mathcal{F}$) for which $c_i \leq c$, and is therefore in \mathcal{F} . By condition (i) of the theorem, this implies that f is measurable. \blacksquare

The next theorem examines further the relationship between simple functions and measurable functions. It is both important and somewhat surprising.

Theorem 5.2. Any non-negative measurable function $f: \Omega \rightarrow \mathbb{R}^+$ is the limit of a monotone increasing sequence of non-negative simple functions.

Proof. For each positive integer s , let

$$Q_{p,s} = \left\{ x: \frac{p-1}{2^s} \leq f(x) < \frac{p}{2^s} \right\} \quad (p = 1, 2, \dots, 2^{2s});$$

$$Q_{0,s} = \Omega - \bigcup_{p=1}^{2^{2s}} Q_{p,s} = \{x: f(x) \geq 2^s\}.$$

Then, since f is \mathcal{F} -measurable, $Q_{p,s} \in \mathcal{F}$ and the sets

$$Q_{p,s} \quad (p = 0, 1, \dots, 2^{2s})$$

form an \mathcal{F} -dissection of Ω . The function

$$\begin{aligned} f_s(x) &= \frac{p-1}{2^s} \quad \text{for } x \in Q_{p,s} \quad (p = 1, 2, \dots, 2^{2s}); \\ &= 2^s \quad \text{for } x \in Q_{0,s} \end{aligned}$$

is a simple function and it is immediate that

$$0 \leq f_s \leq f.$$

If $x \in Q_{p,s}$, then either $x \in Q_{2p-1,s+1}$ or $x \in Q_{2p,s+1}$ so that either

$$f_s(x) = f_{s+1}(x) \quad \text{or} \quad f_s(x) + \frac{1}{2^{s+1}} = f_{s+1}(x).$$

Further, if $x \in Q_{0,s}$, then $f_s(x) = 2^s \leq f(x)$ so that either $x \in Q_{0,s+1}$, or $x \in Q_{p,s+1}$ for some $p \geq 2^{2s+1} + 1$; and in either case $f_{s+1}(x) \geq f_s(x)$. Thus for each integer s

$$f_{s+1}(x) \geq f_s(x) \quad \text{for all } x \in \Omega;$$

and the sequence $\{f_s\}$ of simple functions is monotone increasing.

If x is such that $f(x)$ is finite, then, if $2^s > f(x)$ we have

$$0 \leq f(x) - f_s(x) < 2^{-s}$$

so that in this case $f_s(x) \rightarrow f(x)$ as $s \rightarrow \infty$. On the other hand, if

$$f(x) = +\infty, \quad \text{then } f_s(x) = 2^s,$$

so that again $f_s(x) \rightarrow f(x)$ as $s \rightarrow \infty$. \blacksquare

For any function $f: \Omega \rightarrow \mathbf{R}^*$ we define the positive and negative parts f_+, f_- of f by

$$f_+(x) = \max[0, f(x)],$$

$$f_-(x) = -\min[0, f(x)].$$

Then clearly for all x , $f(x) = f_+(x) - f_-(x)$,

$$|f(x)| = f_+(x) + f_-(x),$$

and each of the functions f_+, f_- is non-negative. It follows immediately from theorem 5.1 that, for any measurable f , f_+ and f_- are both measurable. An application of theorem 5.2 now shows that any measurable function can be expressed as a limit of simple functions. Our next step is to show that the class of measurable functions $f: \Omega \rightarrow \mathbf{R}^*$ is closed both for finite algebraic operations and for countable limiting operations. A minor difficulty arises in that \mathbf{R}^* is not an algebraic field so that, for example, $(f+g)(x)$ is not defined at points where $f(x) = +\infty$, $g(x) = -\infty$. In the following theorem therefore, we assume that the functions are such that the algebraic operations are possible.

Theorem 5.3. *If f and g are measurable functions: $\Omega \rightarrow \mathbf{R}^*$ and $k \in \mathbf{R}$, then each of the functions:*

$$f+k, \quad kf, \quad f+g, \quad f^2, \quad fg, \quad 1/f \quad (\text{where } (1/f)(x) = +\infty \text{ if } f(x) = 0), \\ \max(f, g), \quad \min(f, g), \quad f_+, \quad f_-, \quad |f|;$$

which is defined, is measurable.

Proof. The measurability of the first two functions $f+k$, kf follows immediately from any part of theorem 5.1. Consider now the function $(f+g)$. Let $\{r_i\}$ be a sequence containing all the rationals in \mathbf{R} . Then, since $\{r_i\}$ is dense in \mathbf{R} , for any $c \in \mathbf{R}$,

$$\{x: f(x) + g(x) > c\} = \bigcup_{i=1}^{\infty} \{x: f(x) > r_i\} \cap \{x: g(x) > c - r_i\}.$$

By theorem 5.1 each of the sets on the right-hand side is in \mathcal{F} so that, because \mathcal{F} is a σ -field, the set on the left is also in \mathcal{F} , and $(f+g)$ must be measurable.

Now

$$\{x: (f(x))^2 \leq c\} = \begin{cases} \emptyset & \text{if } c < 0, \\ \{x: f(x) = 0\} & \text{if } c = 0, \\ \{x: -c \leq f(x) \leq c\} & \text{if } c > 0, \end{cases}$$

and each of these sets is in \mathcal{F} , so that f^2 is measurable. Further

$$\{x: (1/f)(x) < c\} = \begin{cases} \{x: 1/c < f(x) < 0\} & \text{if } c < 0, \\ \{x: -\infty < f(x) < 0\} & \text{if } c = 0, \\ \{x: -\infty \leq f(x) < 0\} \cup \{x: f(x) > 1/c\} & \text{if } c > 0, \end{cases}$$

and each of these sets is in \mathcal{F} , so that $1/f$ is measurable.

We have already seen that f_+ and f_- are measurable, so that $|f| = f_+ + f_-$ is also measurable. The measurability of the remaining functions now follow from the identities:

$$\begin{aligned} fg &= \frac{1}{2}[(f+g)^2 - f^2 - g^2], \\ \max(f, g) &= \frac{1}{2}[f+g - |f-g|], \\ \min(f, g) &= f+g - \max(f, g). \quad \blacksquare \end{aligned}$$

It is clear that the above theorem extends immediately to functions obtained by carrying out a finite number of algebraic operations on any finite collection of measurable functions.

Theorem 5.4. Suppose $\{f_n\}, n = 1, 2, \dots$ is a sequence of measurable functions: $\Omega \rightarrow \mathbb{R}^*$; then

- (i) the functions $\sup_n f_n$ and $\inf_n f_n$ are measurable;
- (ii) the functions $\limsup_{n \rightarrow \infty} f_n$, $\liminf_{n \rightarrow \infty} f_n$ are measurable;
- (iii) if the sequence $\{f_n\}$ converges and in particular if it is monotone, $\lim_{n \rightarrow \infty} f_n$ is measurable.

Proof. (i) $\{x: \sup_n f_n > c\} = \bigcup_{n=1}^{\infty} \{x: f_n(x) > c\}.$

Since \mathcal{F} is a σ -field, an application of theorem 5.1 now shows that $\sup_n f_n$ is measurable. The case of $\inf_n f_n$ can be proved similarly or it may be deduced from

$$\inf_n f_n = -\sup_n (-f_n).$$

Suppose now that $\{f_n\}$ is monotone increasing; then

$$\lim_{n \rightarrow \infty} f_n = \sup_n f_n$$

and is therefore measurable. Similarly, if $\{f_n\}$ is decreasing,

$$\lim_{n \rightarrow \infty} f_n = \inf_{n \rightarrow \infty} f_n.$$

$$(ii) \text{ If } g_n = \sup_{r \geq n} f_r, \quad h_n = \inf_{r \geq n} f_r,$$

then $\{g_n\}, \{h_n\}$ are monotone sequences, and

$$\lim_{n \rightarrow \infty} \sup f_n = \lim_{n \rightarrow \infty} g_n, \quad \lim_{n \rightarrow \infty} \inf f_n = \lim_{n \rightarrow \infty} h_n$$

so that both are measurable.

(iii) If $\{f_n\}$ converges its limit will be measurable because it is the common value of the measurable functions $\limsup f_n, \liminf f_n$. **]**

It should be remarked that the class of measurable functions is not closed for non-countable operations of the above type. Thus, if A is non-countable and $f_\alpha: \Omega \rightarrow \mathbf{R}^*$ is measurable for each $\alpha \in A$, there is no reason for

$$f(x) = \sup_{\alpha \in A} f_\alpha(x)$$

to be measurable. For example, let A be a subset of $[0, 1]$ which is not \mathcal{L} -measurable (see §4.4), and put

$$\begin{aligned} f_\alpha(x) &= 1 & \text{if } x = \alpha; \\ &= 0 & \text{if } x \neq \alpha. \end{aligned}$$

Then for each $\alpha \in A$, f_α is \mathcal{L} -measurable (it is actually \mathcal{L} -simple) but

$$\chi_A(x) = \sup_{\alpha \in A} f_\alpha$$

is certainly not \mathcal{L} -measurable. In practice when one needs to consider non-countable suprema (as in the theory of stochastic processes with continuous time parameter) one tries to replace the index set A by a countable subset giving the same supremum for the family (at least except for a special subset of Ω of zero measure). If this procedure is impossible for any reason, then there are very serious difficulties in using non-countable suprema.

In the special case where Ω is a topological space and \mathcal{B} is the σ -field of Borel sets in Ω , there is a special name for \mathcal{B} -measurable functions.

Borel measurability

If \mathcal{B} is the class of Borel sets in Ω and $f: \Omega \rightarrow \mathbf{R}^*$ is \mathcal{B} -measurable, then we say that f is a *Borel measurable* function on Ω .

Lemma. Any continuous function $f: \Omega \rightarrow \mathbf{R}$ on a topological space Ω is Borel measurable.

Proof. Since, for continuous f , the inverse image of an open set in \mathbf{R}^* is open in Ω it follows that $\{x: f(x) < c\}$ is open for all $c \in \mathbf{R}$ and is therefore in \mathcal{B} . **]**

If \mathcal{F} , \mathcal{L} are any two σ -fields of subsets of Ω such that $\mathcal{F} \supset \mathcal{L}$, it is immediate that any function $f: \Omega \rightarrow \mathbf{R}^*$ which is \mathcal{L} -measurable is also \mathcal{F} -measurable. In particular if $\mathcal{F} \supset \mathcal{B}$, then any continuous function on a topological space Ω is \mathcal{F} -measurable. If $\Omega = \mathbf{R}^k$ (Euclidean k -space) then we know that the class \mathcal{L}^k of Lebesgue measurable sets, and the class \mathcal{L}_F^k of sets which are measurable with respect to the Lebesgue–Stieltjes measure defined by F each contain \mathcal{B}^k , the Borel sets in \mathbf{R}^k . Hence, all continuous functions on \mathbf{R}^k are Borel measurable and therefore \mathcal{L}_F^k -measurable for any F (in particular they are \mathcal{L}^k -measurable which we call *Lebesgue measurable*). Functions which normally occur in real analysis are usually obtainable from continuous functions and simple functions by the operations of the following types:

- (i) finite algebraic operations;
- (ii) countable limiting operations;
- (iii) composition.

We have already seen that operations of types (i) and (ii) preserve measurability so that we should consider whether composition operations can be carried out within the class of measurable functions.

Lemma. Suppose $f: \mathbf{R}^* \rightarrow \mathbf{R}^*$ is Borel measurable and $g: \Omega \rightarrow \mathbf{R}^*$ is \mathcal{F} -measurable, then the composite function $f \circ g: \Omega \rightarrow \mathbf{R}^*$ is \mathcal{F} -measurable.

Proof. If A is any Borel set in \mathbf{R}^* , then since f is Borel measurable, the set $f^{-1}(A)$ is also a Borel set in \mathbf{R}^* . Now

$$\{x: f(g(x)) \in A\} = \{x: g(x) \in B\} \in \mathcal{F}$$

since $B = f^{-1}(A) \in \mathcal{B}^*$. **]**

Remark. In the above lemma, it is not sufficient to assume that $f: \mathbf{R}^* \rightarrow \mathbf{R}^*$ is Lebesgue measurable —see exercise 5.2 (9).

This means that, for most of the functions which normally occur in analysis, it is immediately obvious that they are \mathcal{L}_F^k -measurable for every F , and in particular that they are Lebesgue measurable.

Almost everywhere (a.e.)

It is convenient to have a way of describing the behaviour of a function $f: \Omega \rightarrow \mathbf{R}^*$ outside an (unspecified) set of zero measure. If P is some property describing the behaviour of $f(x)$ at a particular point x , then we say that $f(x)$ has a property P almost everywhere

with respect to μ , if there is some set $E \in \mathcal{F}$ with $\mu(E) = 0$ such that $f(x)$ has property P for all $x \in \Omega - E$. We then write

$f(x)$ has property P a.e. (μ);

and, if there is no ambiguity about the measure being considered, (μ) can be omitted.

Lemma. *If \mathcal{F} is complete with respect to μ , and $f = g$ a.e., then f is measurable if and only if g is measurable.*

Proof. For any $c \in \mathbf{R}$ the set

$$\{x: f(x) < c\} \Delta \{x: g(x) < c\} \subset \{x: f(x) \neq g(x)\}$$

so that $\{x: f(x) < c\}$ differs from $\{x: g(x) < c\}$ by a subset of a set of zero measure. If \mathcal{F} is complete with respect to μ , all such sets are in \mathcal{F} so that

$$\{x: f(x) < c\} \in \mathcal{F} \quad \text{if and only if} \quad \{x: g(x) < c\} \in \mathcal{F}. \quad \blacksquare$$

Exercises 5.2

1. In theorem 5.1, show that the condition $\{x: f(x) \leq c\} \in \mathcal{F}$ for all rational c is sufficient to imply that $f: \Omega \rightarrow \mathbf{R}^*$ is \mathcal{F} -measurable.

2. Suppose $\{f_n\}$ is a sequence of functions: $\Omega \rightarrow \mathbf{R}^*$ each of which is finite a.e. Show that, for almost all x in Ω , $f_n(x)$ is finite for all n .

3. Suppose G is an open set in \mathbf{R} and $\{f_n\}$ is a convergent sequence of functions: $\Omega \rightarrow \mathbf{R}$. Show that

$$\{x: \lim_{n \rightarrow \infty} f_n(x) \in G\} = \bigcup_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ x: d(f_n(x), \mathbf{R} - G) > \frac{1}{m} \right\},$$

where $d(y, E)$ denotes the distance from y to E (defined in §2.1).

4. Show that, in theorem 5.2, the condition $f \geq 0$ can be deleted provided we do not require monotonicity for the sequence $\{f_n\}$ of simple functions converging to f . Show that if f is unbounded above and below it is impossible to arrange for the sequence $\{f_n\}$ to be monotone.

5. An elementary function is one which assumes a countable set of values each on a measurable subset of Ω . Show that, if $f: \Omega \rightarrow \mathbf{R}$ is measurable then it is the *uniform* limit of a monotone sequence of elementary functions, but that if f is not bounded it is not the uniform limit of simple functions.

6. If \mathcal{F} is a finite field of subsets of Ω , show that $f: \Omega \rightarrow \mathbf{R}$ is \mathcal{F} -measurable if and only if it is \mathcal{F} -simple.

7. If Ω is a topological space, give examples to show that, for $f: \Omega \rightarrow \mathbf{R}$, the condition

‘ f is continuous a.e. in Ω ’

neither implies nor is implied by the condition

‘there is a continuous $g: \Omega \rightarrow \mathbf{R}$ for which $f = g$ a.e.’

8. Suppose Ω is a topological space, $\mathcal{F} \supset \mathcal{B}$ and $(\Omega, \mathcal{F}, \mu)$ is such that \mathcal{F} is complete with respect to μ . Show that any function f which is continuous a.e. is \mathcal{F} -measurable. Give an example of a measurable function which cannot be made continuous by altering its values on any set of zero measure.

9. If \mathcal{L} is the class of Lebesgue measurable sets in \mathbf{R} , give an example of an \mathcal{L} -measurable function $f: \mathbf{R} \rightarrow \mathbf{R}$ and an \mathcal{L} -measurable set $E \subset \mathbf{R}$ for which $f^{-1}(E)$ is not \mathcal{L} -measurable.

Hint. Use a suitable subset of the Cantor set (see §2.7).

5.3 Definition of the integral

Our method is to define the operation of integration first for non-negative simple functions, and then extend the definition step-by-step showing at each stage that the desirable properties discussed in §5.1 are obtained. If we think of measure as a mass distribution in Ω , and integration as a means of averaging a given function f with respect to this mass distribution it is clear that there is only one reasonable definition for the integral of

(1) *A non-negative simple function*

If

$$f(x) = \sum_{i=1}^n c_i \chi_{E_i}(x), \quad (5.3.1)$$

where $c_i \geq 0$ ($i = 1, 2, \dots, n$) we define

$$\int f d\mu = \sum_{i=1}^n c_i \mu(E_i).$$

This sum is always defined since each of the terms is non-negative. It is called the integral of f with respect to μ . (Note that if $c_i = 0$, $\mu(E_i) = +\infty$ our convention is that $c_i \mu(E_i) = 0$.) Since the representation of a simple function in the form (5.3.1.) is not unique we must first see that our definition of the integral does not depend on the particular representation used. Suppose

$$f = \sum_{i=1}^n c_i \chi_{E_i} = \sum_{j=1}^m d_j \chi_{F_j},$$

then since both systems of sets are dissections of Ω

$$\mu(E_i) = \sum_{j=1}^m \mu(E_i \cap F_j) \quad \text{and} \quad \mu(F_j) = \sum_{i=1}^n \mu(E_i \cap F_j). \quad (5.3.2)$$

Also if $E_i \cap F_j$ is not empty, it will contain a point x and $f(x) = c_i = d_j$.

Thus

$$\begin{aligned} \sum_{i=1}^n c_i \mu(E_i) &= \sum_{i=1}^n \sum_{j=1}^m c_i \mu(E_i \cap F_j) = \sum_{i=1}^n \sum_{j=1}^m d_j \mu(E_i \cap F_j) \\ &= \sum_{j=1}^m d_j \mu(F_j). \end{aligned}$$

Now consider two non-negative simple functions

$$f = \sum_{i=1}^n c_i \chi_{E_i}, \quad g = \sum_{j=1}^m d_j \chi_{F_j}$$

and use the representations

$$f = \sum_{i=1}^n \sum_{j=1}^m c_i \chi_{E_i \cap F_j}, \quad g = \sum_{j=1}^m \sum_{i=1}^n d_j \chi_{E_i \cap F_j}$$

in terms of the dissection $E_i \cap F_j$. Then the simple function $(f+g)$ has the representation

$$f+g = \sum_{i=1}^n \sum_{j=1}^m (c_i + d_j) \chi_{E_i \cap F_j},$$

$$\begin{aligned} \text{and} \quad \int (f+g) d\mu &= \sum_{i=1}^n \sum_{j=1}^m (c_i + d_j) \mu(E_i \cap F_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m c_i \mu(E_i \cap F_j) + \sum_{i=1}^n \sum_{j=1}^m d_j \mu(E_i \cap F_j) \\ &= \sum_{i=1}^n c_i \mu(E_i) + \sum_{j=1}^m d_j \mu(F_j), \quad \text{using (5.3.2)} \\ &= \int f d\mu + \int g d\mu. \end{aligned}$$

It is now immediate from the definition that if $\alpha \geq 0$, $\beta \geq 0$ and f, g are non-negative simple functions then

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$$

so that our operator is linear on the class of non-negative simple functions. It is also clear that it is order preserving; that is, if f, g are simple functions and $f \geq g$ then $\int f d\mu \geq \int g d\mu$.

These properties allow us to extend our definition to:

(2) *Non-negative measurable functions*

Given a measurable $f: \Omega \rightarrow \mathbf{R}^+$, by theorem 5.2 there is a monotone increasing sequence f_n of simple functions such that $f_n \rightarrow f$. Since

$\int f_n d\mu$ is defined for all n , and is monotone increasing it has a limit in \mathbb{R}^+ (which may be $+\infty$). We define

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu. \quad (5.3.3)$$

Since there are many possible monotone sequences of simple functions which converge to a given non-negative measurable f , we must show that the integral $\int f d\mu$ defined in this way is independent of the particular sequence used.

Suppose $\{f_n\}$ is an increasing sequence of non-negative simple functions and $f = \lim_{n \rightarrow \infty} f_n \geq g$, where g is non-negative simple. The first (and main) step in showing that our definition (5.3.3) is proper is to show that, in these circumstances

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int g d\mu. \quad (5.3.4)$$

Put

$$g = \sum_{i=1}^k c_i \chi_{E_i},$$

then if $\int g d\mu = +\infty$, there must be an integer i , $1 \leq i \leq k$ such that $c_i > 0$, $\mu(E_i) = +\infty$. Then for any fixed ϵ such that $0 < \epsilon < c_i$, the sequence of sets $\{A_n \cap E_i\}$ ($n = 1, 2, \dots$) is monotone increasing to E_i where

$$A_n = \{x: f_n + \epsilon > g\}. \quad (5.3.5)$$

Hence $\mu(A_n \cap E_i) \rightarrow +\infty$ as $n \rightarrow \infty$, by theorem 3.2. But

$$\int f_n d\mu \geq \int f_n \chi_{A_n \cap E_i} d\mu \geq (c_i - \epsilon) \mu(A_n \cap E_i) \rightarrow +\infty \quad \text{as } n \rightarrow \infty.$$

Thus (5.3.4) is established, if $\int g d\mu = +\infty$. Now assume that $\int g d\mu$ is finite and put

$$A = \{x: g(x) > 0\} = \bigcup_{c_i > 0} E_i.$$

Since g is simple, $c = \min_{c_i > 0} c_i$ is positive, and $\mu(A) < \infty$. We now suppose $\epsilon > 0$ and again define A_n by (5.3.5). Then

$$\begin{aligned} \int f_n d\mu &\geq \int f_n \chi_{A_n \cap A} d\mu \geq \int (g - \epsilon) \chi_{A_n \cap A} d\mu \\ &= \int g \chi_{A_n \cap A} d\mu - \epsilon \mu(A_n \cap A) \geq \int g \chi_{A_n} d\mu - \epsilon \mu(A). \end{aligned}$$

Since $\mu(A_n \cap E_i) \rightarrow \mu(E_i)$ for each i , we can evaluate the integrals as finite sums and find an integer $n_0 = n_0(\epsilon)$ such that

$$\int f_n d\mu \geq \int g \chi_A d\mu - \epsilon - \epsilon \mu(A) \quad \text{for } n \geq n_0.$$

so that we have established (5.3.4) also in the case $\int g d\mu < \infty$.

We can now suppose given two sequences of simple functions

$$0 \leq f_1 \leq f_2 \leq \dots \leq f_n \leq \dots \rightarrow f,$$

$$0 \leq g_1 \leq g_2 \leq \dots \leq g_m \leq \dots \rightarrow f$$

each monotone increasing and convergent to f . Then for each fixed m we have, by (5.3.4), since

$$f = \lim_{n \rightarrow \infty} f_n \geq g_m,$$

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int g_m d\mu.$$

If we now let $m \rightarrow \infty$ $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \lim_{m \rightarrow \infty} \int g_m d\mu$.

Since the situation is symmetrical, the opposite inequality is similarly proved and we must have

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu.$$

Thus the operation of integration is properly defined for non-negative measurable functions. Because of the corresponding result for non-negative simple functions, it now follows that, if f, g are two non-negative measurable functions and $\alpha \geq 0, \beta \geq 0$ then

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

By our definition, for $f \geq 0$ and measurable, $\int f d\mu$ may be finite or $+\infty$. A non-negative measurable function f is said to be *integrable* with respect to the measure μ if $\int f d\mu$ is finite.

There are clearly two possible reasons for such an f to fail to be integrable. Either there is a simple function $g \leq f$ for which $\int g d\mu = \infty$, which would imply the existence of a $c > 0$ for which $\mu\{x: f(x) > c\} = +\infty$, or alternatively it is possible that $\int g d\mu$ is finite for all simple functions $g \leq f$ (which implies $\mu\{x: f(x) > c\} < \infty$, all $c > 0$) but, for any sequence g_n of simple functions converging to f , $\int g_n d\mu \rightarrow +\infty$ as $n \rightarrow \infty$.

We can now define the integral for:

(3) Integrable measurable functions

We know that if $f: \Omega \rightarrow \mathbf{R}^*$ is measurable, then so are

$$f_+, f_- \quad \text{and} \quad f = f_+ - f_-.$$

If both f_+ and f_- are integrable, then we say that f is integrable and define

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu.$$

Thus our operation of integration is now well-defined on the class \mathcal{A} of integrable functions. We will show in the next section that all the desirable properties discussed in §5.1 are satisfied by this operation.

Finally, we define:

(4) *Integral of a function f over a set A*

This can be considered only for sets A in \mathcal{F} . Put

$$\int_A f d\mu = \int f \chi_A d\mu$$

provided $\int f \chi_A d\mu$ is defined. Thus $\int_A f d\mu$ will be defined if either

(i) $f \chi_A$ is non-negative and measurable, or (ii) $f \chi_A$ is measurable and integrable. We say that f is integrable over A (with respect to μ) if the function $f \chi_A$ is integrable. It is clear that

$$\int_{\Omega} f d\mu = \int f d\mu$$

and we will usually continue to omit the set Ω when we are integrating over the whole space.

Note that, if $E \in \mathcal{F}$ and $\mu(E) = 0$, then any function $f: \Omega \rightarrow \mathbb{R}^*$ is integrable over E with

$$\int_E f d\mu = 0.$$

Exercises 5.3

1. Show that a simple function

$$f(x) = \sum_{i=1}^n c_i \chi_{E_i}(x)$$

is integrable if and only if $c_i = 0$ for each integer i such that $\mu(E_i) = +\infty$.

2. Let Ω be a finite set, $\mu(E)$ the number of points in E . Show that all functions on Ω are simple functions and that the theory of integration reduces to the theory of finite sums.

3. If $f: \Omega \rightarrow \mathbb{R}^*$ is integrable (μ) show that, for any $\epsilon > 0$

$$\mu\{x: |f(x)| \geq \epsilon\} < \infty.$$

4. Suppose μ_1 and μ_2 are two measures defined on \mathcal{F} and $\nu = \mu_1 + \mu_2$. Show that if f is integrable with respect to μ_1 and μ_2 over a set E , then it is integrable with respect to ν and

$$\int_E f d\nu = \int_E f d\mu_1 + \int_E f d\mu_2.$$

5. Suppose $f: \Omega \rightarrow \mathbf{R}^+$ is a non-negative measurable function. Show that

$$\int_E f d\mu = \sup \left[\sum_{k=1}^n \mu(E_k) \inf \{f(x) : x \in E_k\} \right],$$

where the supremum is taken over the collection of all finite classes of disjoint measurable sets with

$$E = \bigcup_{k=1}^n E_k.$$

(This is a possible way of defining $\int_E f d\mu$ which leads to the same class of integrable functions).

6. Suppose $\mu(E) < \infty$ and $f: E \rightarrow \mathbf{R}$ is a measurable finite-valued function defined on $E \in \mathcal{F}$. Put

$$S_n(E) = \sum_{k=-\infty}^{\infty} \frac{k}{2^n} \mu \left\{ x \in E, \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\}.$$

Show that this series is absolutely convergent for all n if it is absolutely convergent for any one $n \in \mathbf{Z}$. Show that f is integrable on E if and only if the series converges absolutely for all n and in this case

$$\int_E f d\mu = \lim_{n \rightarrow \infty} S_n(E).$$

Show that this is not valid if $\mu(E) = +\infty$.

(This is another possible way of defining $\int_E f d\mu$.)

5.4 Properties of the integral

We have now defined the operation of integration with respect to a measure μ on a class of integrable functions. The first objective must be to show that our operation has the properties outlined in §5.1. These are of two types: those involving only a finite number of functions, and operations involving a countable class of functions. We will obtain various closure properties of the class \mathcal{A} while we are examining the integration operation.

Theorem 5.5. Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space, A, B are disjoint sets in \mathcal{F} and $f: \Omega \rightarrow \mathbf{R}^*$, $g: \Omega \rightarrow \mathbf{R}^*$ are two functions integrable

(over Ω) with respect to μ . Then f is integrable over A , $f+g$ and $|f|$ are integrable (over Ω) and

- (i) $\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu$;
- (ii) f is finite a.e.;
- (iii) $\int (f+g) d\mu = \int f d\mu + \int g d\mu$;
- (iv) $|\int f d\mu| \leq \int |f| d\mu$;
- (v) for any $c \in \mathbf{R}$, cf is integrable and $\int cf d\mu = c \int f d\mu$;
- (vi) $f \geq 0 \Rightarrow \int f d\mu \geq 0$; $f \geq g \Rightarrow \int f d\mu \geq \int g d\mu$;
- (vii) if $f \geq 0$ and $\int f d\mu = 0$, then $f = 0$ a.e.;
- (viii) $f = g$ a.e. $\Rightarrow \int f d\mu = \int g d\mu$;
- (ix) if $h: \Omega \rightarrow \mathbf{R}^*$ is \mathcal{F} -measurable and $|h| \leq f$, then h is integrable.

Corollary. Any function $f: \Omega \rightarrow \mathbf{R}^*$ which is bounded, \mathcal{F} -measurable, and zero outside a set E in \mathcal{F} of finite μ -measure is integrable (over Ω) with respect to μ .

Proof. If $f: \Omega \rightarrow \mathbf{R}^+$ is non-negative measurable and integrable (over Ω) and $0 \leq g \leq f$ with $g: \Omega \rightarrow \mathbf{R}^+$ measurable, it follows immediately from the definition of the integral of non-negative measurable functions that g is integrable. Since for any $A \in \mathcal{F}$, χ_A is measurable, $0 \leq f_+ \chi_A \leq f_+$ and $0 \leq f_- \chi_A \leq f_-$,

and a function f which is integrable over Ω is also integrable over any measurable set A .

(i) If A, B are disjoint,

$$\chi_{A \cup B} = \chi_A + \chi_B,$$

so that

$$f_+ \chi_{A \cup B} = f_+ \chi_A + f_+ \chi_B,$$

$$f_- \chi_{A \cup B} = f_- \chi_A + f_- \chi_B;$$

and since property (i) is already known for non-negative measurable functions we must have

$$\begin{aligned} \int_{A \cup B} f d\mu &= \int f_+ \chi_{A \cup B} d\mu - \int f_- \chi_{A \cup B} d\mu \\ &= \int f_+ \chi_A d\mu - \int f_- \chi_A d\mu + \int f_+ \chi_B d\mu - \int f_- \chi_B d\mu; \end{aligned}$$

and

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu, \quad (5.4.1)$$

since all the terms are finite.

(ii) If f is not finite a.e., then at least one of the sets

$$A_1 = \{x: f(x) = +\infty\}, \quad A_2 = \{x: f(x) = -\infty\}$$

has positive measure. Suppose $\mu(A_1) > 0$. Then it follows immediately from the definition that $\int f_+ d\mu = +\infty$ which means that f is not integrable (over Ω).

(iii) This has already been proved for non-negative functions f, g . If f_1, f_2 are non-negative and $f = f_1 - f_2$, then $f_1 + f_- = f_2 + f_+$ and applying (iii) for non-negative functions gives

$$\int f_1 d\mu + \int f d\mu = \int f_2 d\mu + \int f_+ d\mu$$

so that

$$\int f d\mu = \int f_1 d\mu - \int f_2 d\mu.$$

Now the general result follows since, for finite f, g

$$f + g = (f_+ + g_+) - (f_- + g_-),$$

so that

$$\begin{aligned} \int (f + g) d\mu &= \int (f_+ + g_+) d\mu - \int (f_- + g_-) d\mu \\ &= \int f_+ d\mu - \int f_- d\mu + \int g_+ d\mu - \int g_- d\mu \\ &= \int f d\mu + \int g d\mu. \end{aligned}$$

Finally apply (iii) to the function $|f| = f_+ + f_-$ to deduce that $|f|$ is integrable and

$$\int |f| d\mu = \int f_+ d\mu + \int f_- d\mu.$$

(iv) This now follows immediately as

$$\left| \int f d\mu \right| = \left| \int f_+ d\mu - \int f_- d\mu \right| \leq \int f_+ d\mu + \int f_- d\mu.$$

(v) If $c = 0$, $cf = 0$ and $\int cf d\mu = 0 = c \int f d\mu$. If $c > 0$ then

$$(cf)_+ = cf_+, \quad (cf)_- = cf_-$$

and the result follows since it has already been proved for non-negative functions (p. 113). Similarly, if $c < 0$

$$(cf)_+ = -cf_-, \quad (cf)_- = -cf_+,$$

$$\int (cf)_+ d\mu - \int (cf)_- d\mu = (-c) \int f_- d\mu + c \int f_+ d\mu = c \int f d\mu.$$

(vi) The first statement follows from the definition.

If $f \geq g$, then $f = g + (f - g)$, and $(f - g) \geq 0$. By (iii), we now have

$$\int f d\mu = \int g d\mu + \int (f - g) d\mu \geq \int g d\mu.$$

(vii) If $\{x: f(x) > 0\}$ has positive measure, then by theorem 3.2 there is an integer n such that, if

$$A = \{x: f(x) > 1/n\},$$

$\mu(A) > 0$. But $n^{-1}\chi_A \leq f\chi_A \leq f$, so that

$$\int f d\mu \geq \frac{1}{n} \int \chi_A d\mu = \frac{1}{n} \mu(A) > 0.$$

Hence, if $f \geq 0$ and $\int f d\mu = 0$, we must have $\mu\{x: f(x) > 0\} = 0$.

(viii) If $f = g$ a.e., then $f_+ = g_+$, $f_- = g_-$ a.e. In the construction of theorem 5.2, the sets $Q_{p,s}$ for the two functions f_+ and g_+ will all have the same measure. Hence, there are simple functions $f_n \rightarrow f_+$, $g_n \rightarrow g_+$ such that

$$\int f_n d\mu = \int g_n d\mu \quad (n = 1, 2, \dots),$$

and it follows that $\int f_+ d\mu = \int g_+ d\mu$. Similarly $\int f_- d\mu = \int g_- d\mu$.

(ix) If $|h| \leq f$ then $0 \leq h_+ \leq f$, $0 \leq h_- \leq f$. From (vi) it now follows that each of $\int h_+ d\mu$, $\int h_- d\mu$ is finite, and h is therefore integrable.

Proof of corollary. If $|f| \leq K$, then the simple function $K\chi_E$ is integrable and the integrability of f now follows from (ix). \blacksquare

Remark. If \mathcal{F} is complete with respect to μ , then (viii) can clearly be strengthened as follows:

If $f: \Omega \rightarrow \mathbb{R}^*$ is integrable, and $g: \Omega \rightarrow \mathbb{R}^*$ is such that $f = g$ a.e., then g is integrable and $\int f d\mu = \int g d\mu$. There is also a converse to this remark: if f and g are integrable functions such that

$$\int_E f d\mu = \int_E g d\mu \quad \text{for all } E \in \mathcal{F},$$

then $f = g$ a.e. For, suppose not, so that $\mu\{x: f(x) \neq g(x)\} > 0$. Then

at least one of $\{x: f(x) > g(x)\}$, $\{x: f(x) < g(x)\}$ has positive measure. By theorem 3.2 there must be an integer n such that, for

$$E_n = \left\{x: f(x) \geq g(x) + \frac{1}{n}\right\}, \quad \mu(E_n) > 0.$$

But then
$$\int_{E_n} f d\mu - \int_{E_n} g d\mu > \frac{1}{n} \mu(E_n) > 0$$

which is a contradiction establishing the required result.

We can now consider theorems about the continuity of the integration operator.

Theorem 5.6. Suppose $\{f_n\}$ is a monotone increasing sequence of non-negative measurable functions: $\Omega \rightarrow \mathbb{R}^+$ and $f_n(x) \rightarrow f(x)$ for all $x \in \Omega$: then

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu,$$

in the sense that, if f is integrable, the integrals $\int f_n d\mu$ converge to $\int f d\mu$; while if f is not integrable either f_n is integrable for all n and $\int f_n d\mu \rightarrow +\infty$ as $n \rightarrow \infty$, or there is an integer N such that f_N is not integrable so that $\int f_n d\mu = +\infty$ for $n \geq N$.

Proof. For each $n = 1, 2, \dots$ choose an increasing sequence

$$\{f_{n,k}\} (k = 1, 2, \dots)$$

of non-negative simple functions converging to f_n , and put

$$g_k = \max_{n \leq k} [f_{n,k}].$$

Then $\{g_k\}$ is a non-decreasing sequence of non-negative simple functions and

$$g = \lim_{k \rightarrow \infty} g_k$$

is non-negative measurable. But

$$f_{n,k} \leq g_k \leq f_k \leq f \quad \text{for } n \leq k, \quad (5.4.2)$$

so that

$$f_n \leq g \leq f;$$

and, if we now let $n \rightarrow \infty$, we see that $f = g$. Using the order property (vi) of theorem 5.5 and (5.4.2) gives

$$\int f_{n,k} d\mu \leq \int g_k d\mu \leq \int f_k d\mu \quad \text{for } n \leq k.$$

For fixed n , let $k \rightarrow \infty$; from the definition of the integral,

$$\int f_n d\mu \leq \int g d\mu \leq \lim_{k \rightarrow \infty} \int f_k d\mu.$$

If we now let $n \rightarrow \infty$, we obtain

$$\lim_{n \rightarrow \infty} \int f_n d\mu \leq \int g d\mu \leq \lim_{n \rightarrow \infty} \int f_k d\mu.$$

Since the two extremes of the inequality are the same, we must have

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int g d\mu = \int f d\mu. \blacksquare$$

Corollary (absolute continuity). *If f is integrable (over Ω) then, for $A \in \mathcal{F}$,*

$$\int_A f d\mu \rightarrow 0 \quad \text{as} \quad \mu(A) \rightarrow 0.$$

Proof. Put

$$\begin{aligned} f_n &= f & \text{if } |f| \leq n \\ &= n & \text{if } |f| \geq n. \end{aligned}$$

Then $|f_n|$ is monotonic increasing to $|f|$ as $n \rightarrow \infty$. Since, by theorem 5.5, $|f|$ is integrable, we have

$$\int |f_n| d\mu \rightarrow \int |f| d\mu \quad \text{as } n \rightarrow \infty.$$

Given $\epsilon > 0$, choose N such that

$$\int |f| d\mu < \int |f_n| d\mu + \frac{1}{2}\epsilon \quad \text{for } n \geq N.$$

Then if $A \in \mathcal{F}$ is such that $\mu(A) < \epsilon/2N$, we have, by theorem 5.5,

$$\begin{aligned} \left| \int_A f d\mu \right| &\leq \int_A |f| d\mu = \int_A |f_N| d\mu \\ &+ \int_A |f| - |f_N| d\mu < \frac{1}{2}\epsilon + \int_\Omega (|f| - |f_N|) d\mu < \epsilon. \blacksquare \end{aligned}$$

Remark. The notion of absolute continuity for a set function $\nu: \mathcal{F} \rightarrow \mathbf{R}$ will be considered more fully in §6.4.

Theorem 5.7 (Fatou). *If $\{f_n\}$ is a sequence of measurable functions which is bounded below by an integrable function, then*

$$\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu$$

Remark. The operation \liminf picks out the small values of a sequence. This theorem says that if this is done point-wise and the result is then integrated the answer will be not greater than the result of first integrating and then applying the operation.

Proof. Since $\{f_n\}$ is bounded below by an integrable function g we may assume without loss of generality that $f_n \geq 0$ for all n . For $h_n = f_n - g \geq 0$ a.e.† and

$$\int h_n d\mu = \int f_n d\mu - \int g d\mu, \quad \liminf h_n = \liminf f_n - g \text{ a.e.}$$

Put $g_n = \inf_{k \geq n} f_k$, then g_n is an increasing sequence of measurable functions and

$$\lim_{n \rightarrow \infty} g_n = \liminf_{n \rightarrow \infty} f_n.$$

Since $f_n \geq g_n$, for all n

$$\liminf_{n \rightarrow \infty} \int f_n d\mu \geq \lim_{n \rightarrow \infty} \int g_n d\mu = \int \lim_{n \rightarrow \infty} g_n d\mu = \int \liminf_{n \rightarrow \infty} f_n d\mu,$$

by theorem 5.6.]

Corollary. *If $\{f_n\}$ is a sequence of measurable functions which is bounded above by an integrable function, then*

$$\int \limsup_{n \rightarrow \infty} f_n d\mu \geq \limsup_{n \rightarrow \infty} \int f_n d\mu.$$

Proof. This can be proved directly by a method similar to that of theorem 5.7, or it can be deduced from that theorem by putting $g_n = -f_n$ ($n = 1, 2, \dots$).]

Theorem 5.8 (Lebesgue). (i) *If $g: \Omega \rightarrow \mathbf{R}^+$ is integrable, $\{f_n\}$ is a sequence of measurable functions $\Omega \rightarrow \mathbf{R}^*$ such that $|f_n| \leq g$ ($n = 1, 2, \dots$) and $f_n \rightarrow f$ as $n \rightarrow \infty$, then f is integrable and*

$$\int f_n d\mu \rightarrow \int f d\mu \quad \text{as } n \rightarrow \infty.$$

(ii) *Suppose $g: \Omega \rightarrow \mathbf{R}^+$ is integrable, $-\infty \leq a < b \leq +\infty$, and for each $t \in (a, b)$, f_t is a measurable function Ω to \mathbf{R}^* . Then if $|f_t| \leq g$ for all $t \in (a, b)$ and $f_t \rightarrow f$ as $t \rightarrow a+$ or $t \rightarrow b-$, then f is integrable and*

$$\int f_t d\mu \rightarrow \int f d\mu.$$

Proof. (i) We first prove the special case of the theorem where

† Since g is integrable the set $\{x: |g(x)| = +\infty\}$ has zero measure, so that the operation $f_n(x) - g(x)$ can be carried out at least outside the set $\{x: |g(x)| = +\infty\}$. We put in a.e. to cover the possible exceptional set of zero measure where $(f_n - g)$ is not defined. By theorem 5.5 (viii) such exceptional sets do not effect the value of the integrals, and we could arbitrarily define $(f_n - g)$ to be zero at the points where $f_n = g = \pm\infty$.

$f_n \geq 0$ and $f_n \rightarrow 0$ as $n \rightarrow \infty$. In this case we can apply theorem 5.7 and corollary to give

$$\begin{aligned} \limsup \int f_n d\mu &\leq \int \limsup f_n d\mu = \int 0 d\mu = 0 \\ &= \int \liminf f_n d\mu \leq \liminf \int f_n d\mu \leq \limsup \int f_n d\mu. \end{aligned}$$

Hence all the inequalities must be equalities, $\lim \int f_n d\mu$ exists, and has the value zero.

In the general case, put $h_n = |f_n - f|$; then $0 \leq h_n \leq 2g$, $2g$ is integrable and h_n is measurable with $h_n \rightarrow 0$ as $n \rightarrow \infty$. But then

$$\left| \int f_n d\mu - \int f d\mu \right| \leq \int |f_n - f| d\mu \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and f is integrable by theorem 5.5(ix).

(ii) Suppose, for example, that $f_t \rightarrow f$ as $t \rightarrow a+$, then we can apply the sequence form of the theorem to $f_n = f_{t_n}$, where $\{t_n\}$ is any sequence in (a, b) converging to a . Since $f = \lim f_n$ we must have

$$\int f_n d\mu \rightarrow \int f d\mu.$$

But the right-hand side is now independent of the particular sequence $\{t_n\}$ chosen so that $\int f_t d\mu$ must approach the limit $\int f d\mu$ as $t \rightarrow a$ through values in (a, b) . \blacksquare

Exercises 5.4

1. Suppose $f: \Omega \rightarrow \mathbb{R}$ is measurable, $A \in \mathcal{F}$, $\mu(A) < \infty$ and

$$f(x) = 0 \quad \text{for } x \in \Omega - A,$$

$$m \leq f(x) \leq M \quad \text{for } x \in A, \quad \text{where } m, M \in \mathbb{R}.$$

Show that f is integrable and

$$m\mu(A) \leq \int f d\mu \leq M\mu(A).$$

2. Prove that, if f and g are integrable functions,

$$\min \left[\int f d\mu, \int g d\mu \right] \geq \int \min(f, g) d\mu.$$

If the two sides of this inequality are equal, what deduction can be made about the relation between f and g ?

3. Prove that, for any $\epsilon > 0$, if f is integrable over E there is a subset $E_0 \subset E$ such that $\mu(E_0) < \infty$, and

$$\left| \int_E f d\mu - \int_{E_0} f d\mu \right| < \epsilon.$$

4. Show that $f: \Omega \rightarrow \mathbf{R}^*$ is integrable if and only if for any $\epsilon < 0$, there exist integrable functions g and h with $g \geq f \geq h$ and $\int (g - h) d\mu < \epsilon$.

5. If $E = \bigcup_{r=1}^{\infty} E_r$ is a countable union of disjoint sets of \mathcal{F} , and f is integrable over E , then

$$\int_E f d\mu = \sum_{r=1}^{\infty} \int_{E_r} f d\mu$$

and the series converges absolutely.

6. Suppose \mathbf{Z} is the set of positive integers, \mathcal{F} is the class of all subsets of \mathbf{Z} and $\mu(E)$ denotes the number of points in E . Show that any $f: \mathbf{Z} \rightarrow \mathbf{R}^*$ is \mathcal{F} -measurable and that f is integrable if and only if $\sum_{n=1}^{\infty} f(n)$ converges absolutely. Deduce that the sum of an absolutely convergent series is unaffected by any rearrangement of the terms.

7. Suppose $\{f_n\}$ is a sequence of integrable functions and

$$\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty.$$

Show that the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely a.e. to an integrable function f and that

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

8. Suppose $\{E_n\}$ is a sequence of sets in \mathcal{F} , m is a fixed positive integer, and G is the set of points which are in E_n for at least m integers n . Then G is measurable and

$$m\mu(G) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

9. Show that a measurable function f is integrable over a measurable set E if and only if

$$\sum \mu[E \cap \{x: |f(x)| \geq n\}]$$

converges.

10. Suppose f is measurable, g is integrable and $\alpha, \beta \in \mathbf{R}$ with $\alpha \leq f(x) \leq \beta$ a.e. Then there is a real γ such that $\alpha \leq \gamma \leq \beta$ and

$$\int f|g| d\mu = \gamma \int |g| d\mu.$$

Show by an example that we cannot replace $|g|$ by g in this equation.

11. Suppose μ is Lebesgue measure in \mathbf{R} and put

$$\begin{aligned} f_n(x) &= -n^2 \quad \text{for } x \in (0, 1/n), \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Then $\liminf f_n = \lim f_n = 0$ for all x , but

$$\int f_n d\mu = -n.$$

This shows that theorem 5.7 is not valid without the restriction that $\{f_n\}$ be bounded below by an integrable g .

12. State and prove a version of Fatou's lemma (theorem 5.7) for a family f_t , $t \in (a, b)$, of non-negative measurable functions.

13. Is it true that, for measurable $f, g: \Omega \rightarrow \mathbf{R}^*$,

$$f^2 \text{ and } g^2 \text{ integrable} \Rightarrow fg \text{ integrable?}$$

Show that, if

$$\left[\int fg d\mu \right]^2 = \int f^2 d\mu \int g^2 d\mu,$$

then f and g are essentially proportional : that is, there is a real α such that $f = \alpha g$ a.e., or $g = 0$ a.e.

5.5 Lebesgue integral; Lebesgue–Stieltjes integral

We have defined the operation of integration on an abstract measure space $(\Omega, \mathcal{F}, \mu)$. Historically this method of integration was first defined on $(\mathbf{R}, \mathcal{L}, \mu)$, where μ denotes Lebesgue measure on the σ -field \mathcal{L} of Lebesgue measurable sets. We have made the definition in the general case since no more work is involved, but we must now specialise it to obtain the Lebesgue integral.

If E is a Lebesgue measurable set in \mathbf{R} , μ denotes Lebesgue measure in \mathbf{R} , f is \mathcal{L} -measurable, then it is usual to use the notation

$$\int_E f(x) dx \quad \text{for} \quad \int_E f d\mu.$$

In particular, if E is an interval with end-points a, b we use the notation

$$\int_a^b f(x) dx \quad \text{for} \quad \int_E f dx,$$

where $E = [a, b]$ or (a, b) or $[a, b)$ or $(a, b]$. Note that, since the Lebesgue measure of a single point is zero, it makes no difference whether the interval is open or closed.

In the above notation a may be $-\infty$ and b may be $+\infty$ so that

$$\int_{-\infty}^{\infty} f(x) dx \quad \text{means} \quad \int_{\mathbf{R}} f d\mu = \int f d\mu.$$

It is worth remarking that the integral over an infinite interval is defined directly (an infinite interval is a measurable set) and not as the limit of integrals over finite intervals.

In \mathbf{R}^k similar notations are used $\int_E \dots \int f(x) dx$ means a 'multiple integral' of f over the set $E \in \mathcal{L}^k$ in Euclidean k -space with respect to Lebesgue measure.

If instead of using Lebesgue measure we use a Lebesgue-Stieltjes measure (defined in §4.5) given a point function F in \mathbf{R}^k , this is equivalent to working in the measure space $(\mathbf{R}^k, \mathcal{L}_F^k, \mu_F)$. We use the notation

$$\int_E f(x) dF(x) \quad \text{for} \quad \int_E f d\mu_F.$$

In this case we do not, in general, obtain the same result when we integrate over $E_1 = [a, b]$ and $E_2 = (a, b)$ so we will not use the notation

$$\int_a^b f(x) dF(x)$$

unless we know that F is continuous for all x . (This condition is sufficient to imply that the μ_F measure of single point sets is zero, so that the integrals over E_1 and E_2 are the same).

Because \mathcal{L}^k is complete with respect to Lebesgue measure (\mathcal{L}_F^k is complete with respect to μ_F) we see that if $f: \mathbf{R}^k \rightarrow \mathbf{R}^*$ is integrable and $f = g$ a.e., then $g: \mathbf{R}^k \rightarrow \mathbf{R}^*$ is also integrable.

The theorems of §5.4 were proved for any measure space $(\Omega, \mathcal{F}, \mu)$ so they are true in particular for Lebesgue measure in \mathbf{R}^k . Thus the Lebesgue integral is an order preserving linear operation on the class \mathcal{A} of Lebesgue integrable functions. It is also a continuous operator in the following senses.

Theorem 5.6 A. *If $\{f_n\}$ ($n = 1, 2, \dots$) is a monotone increasing sequence of non-negative Lebesgue measurable functions on $\mathbf{R}^k \rightarrow \mathbf{R}^+$ and*

$$f = \lim f_n,$$

then
$$\int_{-\infty}^{\infty} f(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx.$$

Corollary. *If $\{f_n\}$ is any sequence of non-negative Lebesgue measurable functions on $\mathbf{R}^k \rightarrow \mathbf{R}^+$ and $f = \sum_{n=1}^{\infty} f_n$, then*

$$\int_{-\infty}^{\infty} f(x) dx = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} f_n(x) dx.$$

Theorem 5.8 A. *If g is Lebesgue integrable and $\{f_n\}$ is a sequence of Lebesgue measurable functions $\mathbf{R}^k \rightarrow \mathbf{R}$ such that $f_n \rightarrow f$ a.e. as $n \rightarrow \infty$*

and $|f_n| \leq g$ a.e. for each n ; then the functions f_n, f are Lebesgue integrable and

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx = \int_{-\infty}^{\infty} f(x) dx.$$

Corollary. If $E \in \mathcal{L}^k$ and $|E|$ is finite, then for any sequence $\{f_n\}$ of \mathcal{L}^k -measurable functions $\mathbf{R}^k \rightarrow \mathbf{R}$ such that $|f_n(x)| \leq a < \infty$ for all n , all $x \in E$, $f_n \rightarrow f$ a.e. in E we have

$$\int_E f(x) dx = \lim_{n \rightarrow \infty} \int_E f_n(x) dx.$$

It is clear that theorem 5.8A can also be translated to give a corresponding result for series. It is also worth remarking that the theorems corresponding to theorems 5.6A, 5.8A for the Riemann integral can only be proved by using some additional assumption that ensures that f is integrable: for example, it is sufficient to assume that $f_n \rightarrow f$ uniformly.

Exercises 5.5

1. From first principles calculate the Lebesgue integrals

$$(i) \int_0^1 x^p dx \quad (p > -1); \quad (ii) \int_1^\infty x^q dx \quad (q < -1);$$

(iii) $\int_S f d\mu$, where μ is Lebesgue measure in \mathbf{R}^2 , $f(x, y) = xy$ and S is the unit square $0 \leq x \leq 1, 0 \leq y \leq 1$.

2. Suppose $f: \mathbf{R} \rightarrow \mathbf{R}^*$ is Lebesgue integrable and

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Show that F is a uniformly continuous function.

3. Show that if $\{f_n\}$ is a sequence of integrable functions $E \rightarrow \mathbf{R}^*$ such that

$$\sum \int_E |f_n(x)| dx < \infty,$$

then $f_n(x) \rightarrow 0$ for almost all $x \in E$.

4. Show that if $|f_n(x)| \leq 1/n^2$ for all integers $n, x \in E$, and each f_n is measurable and g is integrable over E , then

$$\int_E \sum_{n=1}^{\infty} f_n(x) g(x) dx = \sum_{n=1}^{\infty} \int_E f_n(x) g(x) dx.$$

5. Carathéodory defines the Lebesgue integral of a non-negative measurable function in \mathbf{R} as the Lebesgue measure of the ordinate set in \mathbf{R}^2

$$\int_E f(x) dx = |\{(x, y) : x \in E, \quad 0 \leq y \leq f(x)\}|.$$

Show that this definition is equivalent to the one we have given.

6. Suppose $\{x_i\}$ is a sequence of points in \mathbf{R} and $p_i \geq 0$, $\sum_{i=1}^{\infty} p_i < \infty$. $F(x)$ is defined by

$$F(x) = \sum_{x_i \leq x} p_i$$

and μ_F denotes the Lebesgue-Stieltjes measure with respect to F . Show that all functions $f: \mathbf{R} \rightarrow \mathbf{R}^*$ are measurable, and that f is integrable if and only if $\sum_{i=1}^{\infty} p_i f(x_i)$ converges absolutely.

7. Show that the function $f(x) = 1/x^2$ is integrable with respect to Lebesgue measure over $[1, \infty)$, but not with respect to the Lebesgue-Stieltjes measure generated by $F(x) = x^3$.

8. Show that the function $f(x) = x^2$ is integrable with respect to the Lebesgue-Stieltjes measure generated by

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - \frac{1}{(x+1)^4}, & x > 0. \end{cases}$$

9. Show that, for non-negative measurable functions $f: \mathbf{R} \rightarrow \mathbf{R}^+$, the Cauchy definition of the integral over an infinite interval

$$\int_0^{\infty} f d\mu = \lim_{t \rightarrow \infty} \int_0^t f(x) dx$$

is equivalent to the Lebesgue definition.

By considering the function $f(x) = \sin x/x$, show that this equivalence does not extend to all measurable f .

10. Show that if $f: [a, b] \rightarrow \mathbf{R}$ is continuous and $t \in (a, b)$

$$\lim_{v \rightarrow t} \frac{1}{y-t} \left[\int_a^y f(x) dx - \int_a^t f(x) dx \right] = f(t);$$

thus the Lebesgue indefinite integral can be differentiated at points where the integrand is continuous.

5.6* Conditions for integrability

The strength of the integration operator we have defined is that it works on a very wide class of functions. Provided the σ -field \mathcal{F} is large, the restriction that f has to be \mathcal{F} -measurable is not a serious one,

for we have seen that in a topological space Ω , if \mathcal{F} contains the open sets, then any function which can be obtained from continuous functions or simple functions by countable operations will be \mathcal{F} -measurable. The only additional restriction for integrability of f is on the size (that is, the measure) of the sets where $|f|$ is large. It should be emphasised that our operation could be called 'absolute integration' for f is integrable if and only if $|f|$ is and we do not allow the large negative values of f to 'cancel out' the large positive values to give a finite integral unless each of f_+ and f_- is separately integrable (see exercise, 5.5 (9)).

If we restrict our consideration now to the Lebesgue integral on \mathbb{R} , these general comments still apply, but here it is worth comparing the Lebesgue integral with the Riemann integral over finite intervals. Since we want to compare integration operators, for the present section (only) we will use

$$\mathcal{L} \int_a^b f(x) dx \quad \text{to denote the Lebesgue integral,}$$

$$\mathcal{R} \int_a^b f(x) dx \quad \text{to denote the Riemann integral.}$$

It is easy to give examples of functions which are \mathcal{L} -integrable but not \mathcal{R} -integrable. There are two kinds of bad behaviour which can prevent a function from being \mathcal{R} -integrable. These are illustrated by:

(1) bounded functions which are badly discontinuous but still \mathcal{L} -measurable. For example

$$f(x) = \begin{cases} 1 & \text{when } x \text{ is rational,} \\ 0 & \text{when } x \text{ is irrational,} \end{cases}$$

is discontinuous everywhere. For any $a < b$, it is clear that

$$\mathcal{R} \int_a^b f(x) dx$$

cannot exist. However, the set of rational points is countable, and therefore \mathcal{L} -measurable with zero measure, so that $f(x)$ is an \mathcal{L} -simple function and

$$\mathcal{L} \int_a^b f(x) dx = 0.$$

(2) Functions which are unbounded in (a, b) cannot be \mathcal{R} -integrable even if they are continuous everywhere. For example, $f(x) = x^{-\alpha}$ ($0 < \alpha < 1$) is not \mathcal{R} -integrable over $(0, 1)$, although an elementary

calculation shows that it is \mathcal{L} -integrable. If the points of unboundedness of f (as in the above case) are finite in number, it is sometimes possible to use the 'Cauchy-Riemann' process to define the integral. Thus

$$\lim_{\epsilon \rightarrow 0+} \mathcal{R} \int_{\epsilon}^1 f(x) dx$$

is defined in the above case and could be used as a definition of $\int_0^1 f(x) dx$. Provided the Cauchy-Riemann integral of $|f(x)|$ exists, it is not difficult to show that, if the Cauchy-Riemann process for $f(x)$ works, then f is \mathcal{L} -integrable to the same value. This is not true without the condition that the process works for $|f(x)|$, since the \mathcal{L} -integral is an absolute integral.

We know (corollary to theorem 5.5) that any function $f: [a, b] \rightarrow \mathbf{R}$ which is \mathcal{L} -measurable and bounded is \mathcal{L} -integrable. For the existence of the \mathcal{R} -integral it is necessary for f to be bounded, but the condition of measurability does not give sufficient smoothness. In fact the natural way of characterising functions which are \mathcal{R} -integrable over a finite interval is in terms of the measure of the set of points where the function is discontinuous.

Theorem 5.9. *A bounded function $f: [a, b] \rightarrow \mathbf{R}$ is Riemann integrable if and only if the set E of points in $[a, b]$ at which f is discontinuous satisfies $|E| = 0$. Any $f: [a, b] \rightarrow \mathbf{R}$ which is Riemann integrable is Lebesgue integrable to the same value.*

Proof. We use the following definition for the Riemann integral of f over $[a, b]$ (this is not the usual one but can easily be seen to be equivalent by using the basic theory of the \mathcal{R} -integral). For any positive integer n , divide $I_0 = (a, b]$ into 2^n equal half-open intervals

$$I_{n,i} = (a_{n,i-1}, a_{n,i}] \quad (i = 1, 2, \dots, 2^n);$$

put

$$m_{n,i} = \inf \{f(x) : a_{n,i-1} < x \leq a_{n,i}\},$$

$$M_{n,i} = \sup \{f(x) : a_{n,i-1} < x \leq a_{n,i}\},$$

$$g_n(x) = \begin{cases} m_{n,i} & \text{for } x \in I_{n,i}, \\ 0 & \text{for } x \notin I_0; \end{cases}$$

$$h_n(x) = \begin{cases} M_{n,i} & \text{for } x \in I_{n,i}, \\ 0 & \text{for } x \notin I_0. \end{cases}$$

Then for each integer n , $x \in I_0$

$$g_n(x) \leq f(x) \leq h_n(x);$$

$\{g_n\}$ is a monotone increasing sequence of simple functions, and $\{h_n\}$ is a monotone decreasing sequence of simple functions. If we put

$$g = \lim_{n \rightarrow \infty} g_n, \quad h = \lim_{n \rightarrow \infty} h_n,$$

then $g \leq f \leq h$. Further, by definition,

$$\begin{aligned} \mathcal{L} \int_a^b g(x) dx &= \lim_{n \rightarrow \infty} \mathcal{L} \int_a^b g_n(x) dx \\ &= \lim_{n \rightarrow \infty} \frac{b-a}{2^n} \sum_{i=1}^{2^n} m_{n,i} = \lim_{n \rightarrow \infty} s_n, \quad \text{say;} \\ \mathcal{L} \int_a^b h(x) dx &= \lim_{n \rightarrow \infty} \mathcal{L} \int_a^b h_n(x) dx \\ &= \lim_{n \rightarrow \infty} \frac{b-a}{2^n} \sum_{i=1}^{2^n} M_{n,i} = \lim_{n \rightarrow \infty} S_n, \quad \text{say.} \end{aligned}$$

We say that f is \mathcal{R} -integrable over $[a, b]$ if, only if

$$\lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} S_n \quad \text{and} \quad \mathcal{R} \int_a^b f(x) dx$$

is then the common value of the limit.

Now notice that if f is continuous at $x \in (a, b)$ then $g(x) = h(x)$. Conversely if $g(x) = h(x)$ and x is not a dyadic point (that is, $x \notin D$, where D is the countable set of end-points of intervals $I_{n,i}$), then f is continuous at x .

If $\mathcal{R} \int_a^b f(x) dx$ exists, since $g \leq f \leq h$,

$$\mathcal{L} \int_a^b g(x) dx = \mathcal{R} \int_a^b f(x) dx = \mathcal{L} \int_a^b h(x) dx$$

so that, by theorem 5.5 (vii) $g = h$ a.e. Since the set E of points where f is discontinuous is contained in $D \cup \{x: g(x) \neq h(x)\}$ it follows that $|E| = 0$. Further, since Lebesgue measure is complete, f is \mathcal{L} -measurable and, by theorem 5.5 (viii),

$$\mathcal{L} \int_a^b f(x) dx = \mathcal{L} \int_a^b g(x) dx = \mathcal{R} \int_a^b f(x) dx.$$

Conversely if the set E satisfies $|E| = 0$, this implies $g(x) = h(x)$ a.e., which gives, by theorem 5.5 (viii)

$$\mathcal{L} \int_a^b g(x) dx = \mathcal{L} \int_a^b h(x) dx$$

so that f is \mathcal{R} -integrable. **]**

Theorem 5.9 shows that \mathcal{R} -integrable functions have to be continuous at most points. We have many examples of \mathcal{L} -integrable functions which are continuous nowhere. However, there is a sense in which even \mathcal{L} -integrable functions have to be approximable by continuous functions—in fact by functions which are arbitrarily smooth, that is, functions that can be differentiated arbitrarily often.

Theorem 5.10. *Given any \mathcal{L} -integrable function $f: \mathbf{R} \rightarrow \mathbf{R}^*$ and any $\epsilon > 0$ there is a finite interval (a, b) , and a bounded function $g: \mathbf{R} \rightarrow \mathbf{R}$ such that $g(x)$ vanishes outside (a, b) , is infinitely differentiable for all real x and*

$$\mathcal{L} \int_{-\infty}^{\infty} |f(x) - g(x)| dx < \epsilon.$$

Proof. We carry out the approximation in 4 stages.

(i) First, find a finite interval $[a, b]$ and a bounded measurable function f_1 which vanishes outside $[a, b]$ and is such that

$$\mathcal{L} \int_{-\infty}^{\infty} |f(x) - f_1(x)| dx < \frac{1}{2}\epsilon.$$

This can be done by considering the sequence of functions

$$g_n(x) = \begin{cases} f(x) & \text{if } x \in [-n, n] \text{ and } |f(x)| \leq n, \\ n & \text{if } x \in [-n, n] \text{ and } f(x) > n, \\ -n & \text{if } x \in [-n, n] \text{ and } f(x) < -n, \\ 0 & \text{if } x \notin [-n, n]. \end{cases}$$

Then $g_n(x) \rightarrow f(x)$ for all x and $|g_n| \leq |f|$. By theorem 5.8 it follows that

$$\int_{-\infty}^{\infty} |f(x) - g_n(x)| dx \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

so that we can fix a sufficiently large N and put $f_1(x) = g_N(x)$.

(ii) The next step is to approximate f_1 by an \mathcal{L} -simple function f_2 which vanishes outside $[a, b]$ and satisfies

$$\int_{-\infty}^{\infty} |f_1(x) - f_2(x)| dx < \frac{1}{4}\epsilon.$$

This is clearly possible since we defined the integral as a limit of the integrals of simple functions.

(iii) Now a simple function is a finite sum of multiples of indicator functions. If each indicator function can be approximated by the indicator function of a finite number of disjoint intervals, then it will follow that f_2 can be approximated by f_3 , a step function of the form

$$f_3(x) = \sum_{i=1}^n q_i \chi_{J_i}(x),$$

where each J_i is a finite interval and

$$\int_{-\infty}^{\infty} |f_2(x) - f_3(x)| dx < \frac{1}{4}\epsilon.$$

To see that this is possible start with a bounded \mathcal{L} -measurable set E and $\eta > 0$. Find an open set $G \supset E$ such that $|G - E| < \frac{1}{2}\eta$ and from the countable union of disjoint open intervals making up G pick a finite number to form G_0 such that $|G - G_0| < \frac{1}{2}\eta$. It will then follow that $|E \Delta G_0| < \eta$ so that

$$\int_{-\infty}^{\infty} |\chi_E(x) - \chi_{G_0}(x)| dx < \eta.$$

(iv) In order to obtain the required infinitely differentiable function g for which

$$\int_{-\infty}^{\infty} |f_3(x) - g(x)| dx < \frac{1}{4}\epsilon$$

it is now sufficient to find a function for one of the components χ_{J_i} of f_3 .

Suppose $J = (a, b)$ and $0 < 2\eta < b - a$. Put

$$\phi_{a,\eta}(x) = \begin{cases} \exp[(x-a)^2 - \eta^2]^{-1} & \text{for } |x-a| < \eta, \\ 0 & \text{for } |x-a| \geq \eta. \end{cases}$$

$$\text{If } c_\eta^{-1} = \int_{-\infty}^{\infty} \phi_{a,\eta}(x) dx, \quad \text{let } h(x) = c_\eta \int_{-\infty}^x \{\phi_{a,\eta}(t) - \phi_{b,\eta}(t)\} dt.$$

It is easy to check that h is infinitely differentiable and

$$\int |\chi_J(x) - h(x)| dx < 4\eta,$$

since $0 \leq h(x) \leq 1$ for all x and $\{x: \chi_J(x) \neq h(x)\}$ is contained in the two intervals $(a - \eta, a + \eta)$ and $(b - \eta, b + \eta)$. \blacksquare

Remark 1. We stated our approximation theorem in \mathbf{R}^1 . It is also true in \mathbf{R}^k for every k , and in this case we can require the approximating function to have partial derivatives of all orders everywhere. Our proof requires only minor modifications to give the corresponding theorem in \mathbf{R}^k .

Remark 2. If Ω is a topological space, and \mathcal{F} includes the Baire sets in Ω then theorem 5.10 can be generalized to $(\Omega, \mathcal{F}, \mu)$ to show that any integrable function can be approximated by a continuous function.

(The Baire sets are the sets in the σ -ring generated by sets

$$\{x: f(x) > 0\} \quad \text{where } f: \Omega \rightarrow \mathbf{R}$$

is continuous and vanishes outside a compact set).

Exercises 5.6

1. In theorem 5.10 it was shown that any integrable function f could be approximated by a step function g in the sense that

$$\int |f(x) - g(x)| dx < \epsilon.$$

Show that in general it is not possible to arrange at the same time that $g \leq f$.

2. Show if $f: \mathbf{R} \rightarrow \mathbf{R}^*$ is integrable, then

$$\int |f(x+h) - f(x)| dx \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

Hint. Approximate to f by a function which is uniformly continuous and zero outside a bounded interval.

3. If $f_n(x) = e^{-nx} - 2e^{-2nx}$ show that f_n is integrable over $[0, +\infty)$ but that

$$\int_0^\infty \left(\sum_{n=1}^\infty f_n(x) \right) dx \neq \sum_{n=1}^\infty \int_0^\infty f_n(x) dx.$$

4. Put
$$g(x) = \begin{cases} x^2 \sin 1/x^3 & (x \neq 0), \\ 0 & (x = 0), \end{cases}$$

and
$$g'(x) = f(x) \quad \text{for all } x \in \mathbf{R}.$$

Show that $f(x)$ is finite for all x , but unbounded near $x = 0$. Show that $f(x)$ is not \mathcal{R} -integrable over $(0, 1)$, but that it is Cauchy–Riemann integrable (evaluate its integral). Is $f(x)$ \mathcal{L} -integrable over $(0, 1)$?

RELATED SPACES AND MEASURES

6.1 Classes of subsets in a product space

In the last few chapters we have defined all our concepts in a single abstract space Ω and usually we have at any time considered only one measure defined on a fixed class of subsets of Ω . In applications one often requires to consider more than one measure, and the relationship between the spaces and measures involved become important. We first consider measures defined on the Cartesian product of two measure spaces. Before considering the definition of such measures we must examine, in the present section, the structure of the relevant classes of subsets.

In §1.1 we defined the Cartesian product $X \times Y$ of two spaces X, Y to be the set of all ordered pairs (x, y) with $x \in X, y \in Y$.

Rectangle

Any set in $X \times Y$ of the form $E \times F$ with $E \subset X, F \subset Y$ is called a rectangle (set).

Product of classes

If \mathcal{C}, \mathcal{D} denote classes of subsets in X, Y respectively, then $\mathcal{C} \times \mathcal{D}$ denotes the class of all rectangles $E \times F$ with $E \in \mathcal{C}, F \in \mathcal{D}$.

Product ring, field, σ -field

If \mathcal{z} -class again denotes any one of ring, field, σ -ring, σ -field and \mathcal{C}, \mathcal{D} are \mathcal{z} -classes in X, Y , respectively, then the product \mathcal{z} -class is the \mathcal{z} -class in $X \times Y$ generated by $\mathcal{C} \times \mathcal{D}$.

Lemma. *If \mathcal{C}, \mathcal{D} are semi-rings in X, Y respectively, then $\mathcal{C} \times \mathcal{D}$ is a semi-ring in $X \times Y$.*

Proof. It is immediate that $\mathcal{C} \times \mathcal{D}$ is closed for finite intersections, so that we have only to prove that

$$E_1 \times F_1 - E_2 \times F_2$$

can be expressed as a union of disjoint sets of $\mathcal{C} \times \mathcal{D}$ for any

$$E_1, E_2 \in \mathcal{C}; \quad F_1, F_2 \in \mathcal{D}.$$

Since \mathcal{C}, \mathcal{D} are semi-rings we have

$$E_1 - E_2 = \bigcup_{i=3}^n E_i, \quad F_1 - F_2 = \bigcup_{j=3}^m F_j,$$

where the sets E_i ($i = 3, 4, \dots, n$) are disjoint sets of \mathcal{C} , F_j ($j = 3, 4, \dots, m$) are disjoint sets of \mathcal{D} . Now

$$\begin{aligned} E_1 \times F_1 - E_2 \times F_2 &= (E_1 \cap E_2) \times (F_1 - F_2) \cup (E_1 - E_2) \times (F_1 \cap F_2) \\ &\quad \cup (E_1 - E_2) \times (F_1 - F_2) \\ &= \bigcup_{j=3}^m (E_1 \cap E_2) \times F_j \cup \bigcup_{i=3}^n E_i \times (F_1 \cap F_2) \cup \bigcup_{i=3}^n \bigcup_{j=3}^m E_i \times F_j, \end{aligned}$$

and these are all disjoint sets in $\mathcal{C} \times \mathcal{D}$. **]**

It is important to notice that this lemma does not extend to any of the z -classes as $\mathcal{C} \times \mathcal{D}$ is not closed under the operation of union. In particular, if \mathcal{C}, \mathcal{D} are σ -fields then $\mathcal{C} \times \mathcal{D}$ will not be a σ -field. We will use the notation $\mathcal{C} * \mathcal{D}$ for the σ -field in $X \times Y$ generated by $\mathcal{C} \times \mathcal{D}$. We also need some operations which are effective in the opposite direction, from the product space to the components.

Section

Given any set $E \subset X \times Y$ and any point $x \in X$, the subset

$$E_x = \{y: (x, y) \in E\}$$

of Y is called the section of E at x . Similarly, for $y \in Y$, the subset $E^y = \{x: (x, y) \in E\}$ of X is the section of E at y .

Projection

Given any set $E \subset X \times Y$ the sets $\{x: \text{there exists } y \text{ with } (x, y) \in E\}$, $\{y: \text{there exists } x \text{ with } (x, y) \in E\}$ are called the projections of E into the spaces X, Y , respectively.

Although the product σ -field $\mathcal{C} * \mathcal{D}$ of two σ -fields \mathcal{C}, \mathcal{D} contains more than the rectangle sets $E \times F$, $E \in \mathcal{C}$, $F \in \mathcal{D}$, one can deduce an important restriction on the sections of its sets.

Theorem 6.1. *If \mathcal{F}, \mathcal{G} are σ -fields in X, Y , respectively, and \mathcal{H} is the product σ -field in $X \times Y$, then all sets E in \mathcal{H} have the property that $E_x \in \mathcal{G}$ for all $x \in X$ and $E^y \in \mathcal{F}$ for all $y \in Y$.*

Proof. Let \mathcal{C} be the class of subsets E of $X \times Y$ with the property that every section of E is in the appropriate σ -field. Since rectangle sets certainly have this property, it is immediate that $\mathcal{C} \supset \mathcal{F} \times \mathcal{G}$. Moreover, it is not difficult to verify that \mathcal{C} satisfies all the axioms of a σ -field. Hence, $\mathcal{C} \supset \mathcal{H}$ the σ -field generated by $\mathcal{F} \times \mathcal{G}$. **]**

Sections of functions

Given any function $f: X \times Y \rightarrow \mathbf{R}^*$; for each fixed $x \in X$,

$$f_x(y) = f(x, y) \quad \text{for } y \in Y$$

defines a function $Y \rightarrow \mathbf{R}^*$ and for each fixed $y \in Y$, $f_y(x) = f(x, y)$ for $x \in X$ defines a function on X to \mathbf{R}^* . These functions f_x, f_y are called the *sections* of f at x, y , respectively.

Corollary. *Under the conditions of theorem 6.1, given any \mathcal{H} -measurable function $f: X \times Y \rightarrow \mathbf{R}^*$, each of the sections $f_x(y)$ is \mathcal{G} -measurable and each of the sections $f_y(x)$ is \mathcal{F} -measurable.*

Proof. Suppose x_0 is a fixed point in X and M is a Borel set in \mathbf{R}^* ; then

$$\{y: f_{x_0}(y) \in M\} = \{y: f(x_0, y) \in M\} = \{(x, y): f(x, y) \in M\}_{x_0}$$

so that the test set is the section at x_0 of a set in \mathcal{H} .]

The results of theorem 6.1 and corollary can be extended in an obvious way to finite Cartesian products $X_1 \times X_2 \times \dots \times X_n$; there is no difficulty in making the required modifications to the definitions and proofs. It is not quite so immediate that they can also be extended to arbitrary Cartesian products $\prod_{i \in I} X_i$.

Let us recall that a point in $\prod_{i \in I} X_i$ can be thought of as a function $f: I \rightarrow \bigcup_{i \in I} X_i$ such that $f(i) \in X_i$ for each $i \in I$. Suppose then that we have a collection $\{X_i, i \in I\}$ of spaces and σ -fields \mathcal{F}_i of subsets of X_i .

Cylinder set

If i_1, i_2, \dots, i_n is any finite subset of I and $E_{i_k} \in \mathcal{F}_{i_k}$, $k = 1, \dots, n$; the set of points $f \in \prod X_i$ such that

$$f(i_k) \in E_{i_k} \quad (k = 1, 2, \dots, n),$$

is said to be a (finite dimensional) cylinder set in $\prod X_i$. When we say that f is in a cylinder set C , the values of f are restricted only on a finite set of indices. The class of all such cylinder sets in $\prod X_i$ will be denoted $\mathcal{C}(\mathcal{F}_i, i \in I)$.

Lemma. *The class $\mathcal{C}(\mathcal{F}_i, i \in I)$ of cylinder sets is a semi-ring of subsets of $\prod_{i \in I} X_i$.*

Proof. We can think of \mathcal{F}_i as a semi-ring in X_i which contains the whole space X_i . Then if two sets

$$A = \{f: f(i) \in E_i, i \in J\},$$

$$B = \{f: f(i) \in F_i, i \in K\},$$

are in $\mathcal{C}(\mathcal{F}_i, i \in I)$, J and K must be finite subsets of I , and each of the sets E_i, F_i must be in the relevant σ -field \mathcal{F}_i . The set $J \cup K = L$ is also a finite subset of I and, if we put

$$E_i = X_i \quad \text{for } i \in K - J, \quad F_i = X_i \quad \text{for } i \in J - K,$$

then $A = \{f: f(i) \in E_i, i \in L\}, \quad B = \{f: f(i) \in F_i, i \in L\},$

are now cylinder sets in which the same finite subset L of indices are restricted. Since we know that any finite Cartesian product of semi-rings is a semi-ring, we can deduce that $A - B$ is a finite disjoint union of sets of this type and $A \cap B$ is a set of this type. Hence $\mathcal{C}(\mathcal{F}_i, i \in I)$ is a semi-ring. **]**

Note. The case $I = \mathbb{Z}$ is important. The cylinder sets in $\prod_{i=1}^{\infty} X_i$ then reduce to sets of the form $E_1 \times E_2 \times \dots \times E_n \times \prod_{i=n+1}^{\infty} X_i$ with

$$E_i \in \mathcal{F}_i \quad (i = 1, 2, \dots, n).$$

The results corresponding to theorem 6.1 for $\prod_{i=1}^{\infty} X_i$ are formulated as examples for the reader to prove.

Exercises 6.1

1. If \mathcal{R} is a ring of subsets of X , \mathcal{S} is a ring of subsets of Y , show that the product ring consists of those sets in $X \times Y$ which are finite unions of disjoint rectangles in $\mathcal{R} \times \mathcal{S}$.

2. If $A_1, A_2 \subset X, B_1, B_2 \subset Y$ and $A_1 \times B_1 = A_2 \times B_2$ is not null, prove that $A_1 = A_2, B_1 = B_2$.

3. Suppose $E = A \times B, E_1 = A_1 \times B_1$ and $E_2 = A_2 \times B_2$ are all non-empty rectangles in $X \times Y$. Show that E is a disjoint union of E_1 and E_2 if and only if *either* A is a disjoint union of A_1, A_2 and $B = B_1 = B_2$, *or* B is a disjoint union of B_1, B_2 and $A = A_1 = A_2$.

4. If \mathcal{S}, \mathcal{T} are σ -rings in X, Y , respectively, then the product σ -ring in $X \times Y$ is a σ -field if and only if both \mathcal{S} and \mathcal{T} are σ -fields.

5. Show that the intersection of a class of rectangles is a rectangle.

6. Suppose $X = Y$ is any uncountable set and $\mathcal{S} = \mathcal{T}$ is the class of subsets which are either countable or have countable complement. Determine the product σ -field of \mathcal{S} and \mathcal{T} . If $D = \{(x, y): x = y\}$ is the diagonal in $X \times Y$ show that every section of D is in \mathcal{S} or \mathcal{T} but D is not in the product σ -field. This shows that theorem 6.1 has no converse.

7. Suppose \mathcal{F}, \mathcal{G} are σ -fields in X, Y ; then a rectangle set $E \times F$ is in the product σ -field if and only if $E \in \mathcal{F}, F \in \mathcal{G}$.

8. Suppose \mathcal{H} is the product σ -field of two σ -fields \mathcal{F} , \mathcal{G} . Show that any function on $X \times Y \rightarrow \mathbf{R}$ which is \mathcal{H} -simple has all its sections \mathcal{F} -simple or \mathcal{G} -simple.

9. Suppose \mathcal{F} is the product σ -field of two σ -fields $\mathcal{F}_1, \mathcal{F}_2$. Show that the projection of a set in \mathcal{F} on an axis need not be in $\mathcal{F}_1, \mathcal{F}_2$, respectively.

10. Suppose \mathcal{F}_i is a σ -field in $X_i (i = 1, 2, \dots)$ and the σ -field generated by cylinder sets $\mathcal{C}(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots)$ in $\prod_{i=n}^{\infty} X_i$ is denoted by \mathcal{S}_n . Then given any set E in $\prod_{i=1}^{\infty} X_i$, the (finite dimensional) section of E at x_1, x_2, \dots, x_k is the set $\left(\text{in } \prod_{i=k+1}^{\infty} X_i \right)$ of points $(x_{k+1}, x_{k+2}, \dots)$ such that $(x_1, x_2, \dots) \in E$. Then if $E \in \mathcal{S}_1$ the product σ -field in $\prod_{i=1}^{\infty} X_i$, all its k -dimensional sections belong to \mathcal{S}_{k+1} .

6.2 Product measures

We now assume that $(X_1, \mathcal{F}_1, \mu_1)$ and $(X_2, \mathcal{F}_2, \mu_2)$ are measure spaces and μ_1, μ_2 are σ -finite measures. The product σ -field \mathcal{H} in $X_1 \times X_2$ was defined as the smallest σ -field containing the class $\mathcal{F}_1 \times \mathcal{F}_2$ which is known to be a semi-ring since each of $\mathcal{F}_1, \mathcal{F}_2$ are semi-rings. In Chapters 3 and 4 we developed a general method of extending a measure from a semi-ring to the generated σ -ring. Since the semi-ring $\mathcal{F}_1 \times \mathcal{F}_2$ contains the whole space $X_1 \times X_2$ this generated σ -ring must be a σ -field and is therefore $\mathcal{F}_1 * \mathcal{F}_2$, the product σ -field. Thus if we use theorems 3.5 and 4.2 we can extend any σ -finite measure on $\mathcal{F}_1 \times \mathcal{F}_2$ to a σ -finite measure on $\mathcal{F}_1 * \mathcal{F}_2$ in a unique way.

Suppose $E_1 \times E_2$ is any rectangle set in $\mathcal{F}_1 \times \mathcal{F}_2$ and put

$$\mu(E_1 \times E_2) = \mu_1(E_1) \mu_2(E_2),$$

with the usual convention that $0 \cdot \infty = \infty \cdot 0 = 0$. Then μ is a non-negative set function on $\mathcal{F}_1 \times \mathcal{F}_2$ which is easily seen to be σ -finite. Our first objective is to show that μ is a measure on the semi-ring $\mathcal{F}_1 \times \mathcal{F}_2$. First, suppose that

$$E \times F = \bigcup_{i=1}^n (E_i \times F_i)$$

with the sets $E_i \times F_i$ disjoint. Define the functions $f_i: X_1 \rightarrow \mathbf{R}^+$ by $f_i(x) = \mu_2(F_i) \chi_{E_i}(x)$ ($i = 1, 2, \dots, n$). Then f_i is a non-negative function or possibly a function which takes the value $+\infty$ on a measurable set E_i and zero outside it: in any case

$$\int f_i d\mu_1 = \mu_1(E_i) \mu_2(F_i) \quad (i = 1, 2, \dots, n).$$

Similarly, if $f(x) = \mu_2(F) \chi_E(x)$ we have

$$\int f d\mu_1 = \mu_1(E) \mu_2(F).$$

Now for each fixed x in X_1 we have

$$(E \times F)_x = \bigcup_{i=1}^n (E_i \times F_i)_x$$

with the sets $(E_i \times F_i)_x$ disjoint. Since μ_2 is (finitely) additive it follows that

$$f(x) = \sum_{i=1}^n f_i(x).$$

If we now use (finite) additivity for integrals of non-negative simple functions we have

$$\mu_1(E) \mu_2(F) = \int f d\mu_1 = \int \sum_{i=1}^n f_i d\mu_1 = \sum_{i=1}^n \int f_i d\mu_1 = \sum_{i=1}^n \mu_1(E_i) \mu_2(F_i).$$

This shows that the set function μ we have defined is finitely additive on $\mathcal{F}_1 \times \mathcal{F}_2$. The same argument extends without difficulty to countable unions of disjoint rectangles

$$\bigcup_{i=1}^{\infty} (E_i \times F_i) = E \times F$$

because all the functions $f_i(x)$ are non-negative measurable, so that the monotone convergence theorem 5.6 justifies the inversion of integration and summation. Thus μ is a measure on the semi-ring $\mathcal{F}_1 \times \mathcal{F}_2$. It can be extended uniquely by theorem 3.5 to the generated ring, and then, by theorem 4.2, to the generated σ -ring which is the product σ -field $\mathcal{F}_1 * \mathcal{F}_2$. The result is called the *product measure* on $\mathcal{F}_1 * \mathcal{F}_2$. We have thus proved

Theorem 6.2. *Given two measure spaces $(X_1, \mathcal{F}_1, \mu_1)$, $(X_2, \mathcal{F}_2, \mu_2)$ such that μ_1, μ_2 are σ -finite, there is a unique measure μ defined on the product σ -field $\mathcal{F}_1 * \mathcal{F}_2$ in $X_1 \times X_2$ such that*

$$\mu(E_1 \times E_2) = \mu_1(E_1) \mu_2(E_2) \quad \text{for } E_1 \in \mathcal{F}_1, \quad E_2 \in \mathcal{F}_2.$$

The above theorem clearly extends immediately to any finite Cartesian product of σ -finite measure spaces. Difficulties arise with the Cartesian product of an enumerable collection of measure spaces unless we arrange that the infinite products of real numbers occurring converge. The easiest way to ensure this is to restrict the discussion to countable products of measure spaces $(X_i, \mathcal{F}_i, \mu_i)$ with $\mu_i(X_i) = 1$.

It is possible to define product measures on arbitrary product spaces $\prod_{i \in I} X_i$ such that $\mu_i(X_i) = 1$ by exactly the method used below.

We carry out the construction only for enumerable products as, in applications, it is not usually appropriate to consider the product measure for non-countable products. In §6.6 we will give a general construction for a measure in $\prod_{i \in I} X_i$, an arbitrary product space—

this construction could clearly be specialized to give the results of the remainder of this section, but it is simpler to deal with the case of product measures in countable product spaces first. We will set up our measure on the product σ -field by a slightly different procedure.

Let $\mathcal{C} = \mathcal{C}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ be the semi-ring of cylinder sets in $\prod_{i=1}^{\infty} X_i$. We define μ on \mathcal{C} by $\mu(E) = \mu_1(E_1)\mu_2(E_2)\dots\mu_n(E_n)$, if

$$E = E_1 \times E_2 \times \dots \times E_n \times \prod_{i=n+1}^{\infty} X_i; \quad E_i \in \mathcal{F}_i, \quad (i = 1, 2, \dots, n).$$

It is clear that $0 \leq \mu(E) \leq 1$ for all E in \mathcal{C} . To see that μ is finitely additive on \mathcal{C} it is sufficient to see that, in any finite collection of cylinder sets, only a finite number of coordinates are involved so that, if $C = \bigcup_{j=1}^m C_j$ is a dissection of $C \in \mathcal{C}$ into disjoint sets of \mathcal{C} , there is an integer N such that C and C_j ($j = 1, \dots, m$) can all be expressed in the form

$$E_1 \times E_2 \times \dots \times E_N \times \prod_{i=N+1}^{\infty} X_i.$$

We can then apply theorem 6.2 to the finite products to see that

$$\mu(C) = \sum_{j=1}^m \mu(C_j).$$

By theorem 3.4, μ has a unique additive extension to the ring \mathcal{R} of finite unions of cylinder sets. In order to apply theorem 4.2 we must show that μ is a measure on \mathcal{R} . This can be done by using the continuity theorem 3.2. It is sufficient to show that any monotone decreasing sequence $\{A_n\}$ of sets in \mathcal{R} such that

$$\mu(A_n) \rightarrow \epsilon > 0$$

has a non-void intersection.

Let $Y_n = \prod_{i=n+1}^{\infty} X_i$. Then by the above procedure we can define product set functions $\nu^{(n)}$ on the class $\mathcal{R}^{(n)}$ of finite unions of cylinder

sets $\mathcal{C}(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots)$. It is clear that, for each integer n , we can obtain μ on \mathcal{C} by taking the product of the measures μ_i ($i = 1, 2, \dots, n$) and $\nu^{(n)}$. Let

$$A_n(x_1) = \{y: y \in Y_1, (x_1, y) \in A_n\}$$

be the section of A_n at $x_1 \in X_1$. It is clear that, for each $x_1 \in X_1$, $A_n(x_1) \in \mathcal{B}^{(1)}$ and if

$$B_{n,1} = \{x_1: \nu^{(1)}(A_n(x_1)) > \tfrac{1}{2}\epsilon\}$$

then $B_{n,1}$ is a finite union of sets in \mathcal{F}_1 and is therefore in \mathcal{F}_1 : further we must have

$$\mu_1(B_{n,1}) + \tfrac{1}{2}\epsilon(1 - \mu_1(B_{n,1})) \geq \mu(A_n) \geq \epsilon,$$

by considering $A_n \cap (B_{n,1} \times Y_1)$ and $A_n \cap (X_1 - B_{n,1}) \times Y_1$. It follows that

$$\mu_1(B_{n,1}) > \tfrac{1}{2}\epsilon \quad (n = 1, 2, \dots).$$

But $\{A_n\}$ is monotone decreasing so $\{B_{n,1}\}$ must also decrease with n and

$$\mu_1\left\{\bigcap_{n=1}^{\infty} B_{n,1}\right\} \geq \tfrac{1}{2}\epsilon.$$

Since μ_1 is a measure on \mathcal{F}_1 , it follows that there must be at least one point $x_1 \in X_1$ for which

$$\nu^{(1)}(A_n(x_1)) > \tfrac{1}{2}\epsilon \quad \text{for all } n.$$

We now suppose x_1 is fixed as such a point in $\bigcap B_{n,1}$ and repeat the argument to the sequence of sets $\{A_n(x_1)\}$ in the space Y_1 . This gives a point $x_2 \in X_2$ such that

$$\nu^{(2)}(A_n(x_1, x_2)) > \epsilon/2^2 \quad \text{for all } n.$$

By an induction argument we obtain a point (x_1, x_2, \dots) in $\prod_{i=1}^{\infty} X_i$ such that, for any k, n

$$A_n(x_1, x_2, \dots, x_k) \neq \emptyset.$$

But each set A_n has only a finite number of coordinates restricted so the point (x_1, x_2, \dots) must be in A_n for all n . This completes the proof that μ is continuous from above at \emptyset .

Since μ is now seen to be a finite measure on the ring \mathcal{R} it has a unique extension to the generated σ -ring which is also the product σ -field in $\prod_{i=1}^{\infty} X_i$. This extension is called the *product measure*. Thus we have proved

Theorem 6.3. *If $(X_i, \mathcal{F}_i, \mu_i)$ are measure spaces with*

$$\mu_i(X_i) = 1 \quad (i = 1, 2, \dots);$$

then there is a unique measure μ defined on the product σ -field \mathcal{F} of subsets of $X = \prod_{i=1}^{\infty} X_i$ which is generated by the cylinder sets of the form

$$E_1 \times E_2 \times \dots \times E_n \times \prod_{i=n+1}^{\infty} X_i \quad (E_i \in \mathcal{F}_i, i = 1, 2, \dots),$$

such that

$$\mu\left(E_1 \times \dots \times E_n \times \prod_{i=n+1}^{\infty} X_i\right) = \mu_1(E_1) \mu_2(E_2) \dots \mu_n(E_n).$$

Exercises 6.2

1. Given 3 σ -finite measure spaces $(X_1, \mathcal{F}_1, \mu_1)$, $(X_2, \mathcal{F}_2, \mu_2)$, $(X_3, \mathcal{F}_3, \mu_3)$ let τ be the product measure of μ_1, μ_2 in $X_1 \times X_2$ and ν the product measure of μ_2, μ_3 in $X_2 \times X_3$. Show that, in the space $X_1 \times X_2 \times X_3$ the product measure of τ and μ_3 is the same as the product measure of μ_1 and ν .

2. Suppose $(X_i, \mathcal{F}_i, \mu_i) (i = 1, 2, \dots)$ is a sequence of measure spaces with $\mu_i(X_i) = 1$. Let μ be the product measure of theorem 6.3 on $\prod_{i=1}^{\infty} X_i$ and suppose τ_n is the corresponding product measure of $\prod_{i=n+1}^{\infty} X_i$. Show that μ is the same as the product measure of $\mu_1, \mu_2, \dots, \mu_n, \tau_n$ on the finite Cartesian product

$$X_1 \times X_2 \times \dots \times X_n \times \left(\prod_{i=n+1}^{\infty} X_i \right).$$

3. The product measure of two σ -finite complete measures need not be complete. As an example take $X_1 = X_2 =$ unit interval with Lebesgue measure. Suppose M is a non-measurable set in X_1 , and consider the set $M \times \{y\}$; use exercise 6.1 (7).

4. Suppose $\prod_{i=1}^{\infty} X_i$ is a product space with $\mu_i(X_i) = 1$. Let $E_i \in \mathcal{F}_i$, $(i = 1, 2, \dots)$. Then the set $\prod_{i=1}^{\infty} E_i$ is in the product σ -field and $\mu(E) = \prod_{i=1}^{\infty} \mu(E_i)$.

5. If a cylinder set $E_1 \times E_2 \times \dots \times E_n \times \prod_{i=n+1}^{\infty} X_i$ is in the product σ -field \mathcal{F} generated by $\mathcal{C}(\mathcal{F}_1, \mathcal{F}_2, \dots)$, then it is in $\mathcal{C}(\mathcal{F}_1, \mathcal{F}_2, \dots)$; in fact $E_i \in \mathcal{F}_i$ $(i = 1, 2, \dots, n)$.

6.3 Fubini's theorem

Given two measure spaces (X, \mathcal{F}, μ) , (Y, \mathcal{G}, ν) we have now seen how to define a product measure on the product σ -field in $X \times Y$. Given a function $f: X \times Y \rightarrow \mathbf{R}^*$ there are sections $f_x: Y \rightarrow \mathbf{R}^*$ defined for every $x \in X$. Our objective in the present section is to compare the integral of $f(x, y)$ with respect to the product measure with the iterated integral obtained by first integrating $f_x(y)$ with respect to ν for each fixed x , and then integrating the resulting function of x with respect to the measure μ . Because of our method of defining the integral the general result will follow easily from the special case of simple functions. The essential step towards this case is given by the next theorem.

Theorem 6.4. *Given (X, \mathcal{F}, μ) , (Y, \mathcal{G}, ν) two σ -finite measure spaces, let λ be the product measure defined on the product σ -field $\mathcal{F} * \mathcal{G}$. Then for all $A \in \mathcal{F} * \mathcal{G}$, $\nu(A_x)$ is \mathcal{F} -measurable and $\mu(A^y)$ is \mathcal{G} -measurable; and*

$$\lambda(A) = \int \mu(A^y) d\nu = \int \nu(A_x) d\mu.$$

Proof. Suppose first that $\mu(X)$, $\nu(Y)$ are both finite. Let \mathcal{M} be the class of subsets of $X \times Y$ for which the conclusions of the theorem are valid. Then $\mathcal{M} \supset \mathcal{F} \times \mathcal{G}$ since if $A = E_1 \times E_2$, $E_1 \in \mathcal{F}$, $E_2 \in \mathcal{G}$

$\nu(A_x)$ is \mathcal{F} -simple as a function of x ,

$\mu(A^y)$ is \mathcal{G} -simple as a function of y ,

and both these functions integrate to $\lambda(A)$ by the definition of λ on $\mathcal{F} \times \mathcal{G}$. It follows that \mathcal{M} contains the ring \mathcal{R} of finite unions of rectangle sets of $\mathcal{F} \times \mathcal{G}$. Since the limit of a monotone sequence of measurable functions is measurable, and theorem 5.6 applies to the integrals, it follows immediately that \mathcal{M} is a monotone class. Hence, by theorem 1.5, \mathcal{M} is a σ -ring. But clearly \mathcal{M} contains $X \times Y$ so that \mathcal{M} is a σ -field and $\mathcal{M} \supset \mathcal{F} * \mathcal{G}$. The restriction $\mu(X) < \infty$, $\nu(Y) < \infty$ can now be removed by the usual device of taking measurable sequences $\{A_n\}$ increasing to X and $\{B_n\}$ increasing to Y for which $\mu(A_n) < \infty$, $\nu(B_n) < \infty$ for all n , and considering the set $A \cap (A_n \times B_n)$ which increases to A as $n \rightarrow \infty$. **]**

Corollary. *Under the conditions of theorem 6.4, if $A \in \mathcal{F} * \mathcal{G}$, $\lambda(A) = 0$ if and only if $\nu(A_x) = 0$ for almost all x , and if and only if $\mu(A^y) = 0$ for almost all y .*

This follows from the theorem using the fact that a non-negative measurable function can integrate to zero only if it is zero almost everywhere. **]**

Theorem 6.5. *Given all the conditions of theorem 6.4, we write \mathcal{H} for the product σ -field $\mathcal{F} * \mathcal{G}$.*

(i) *If $h: X \times Y \rightarrow \mathbf{R}^+$ is any non-negative \mathcal{H} -measurable function then*

$$\int h \, d\lambda = \int \left(\int h_x \, d\nu \right) d\mu = \int \left(\int h_y \, d\mu \right) d\nu.$$

(ii) *If $h: X \times Y \rightarrow \mathbf{R}^*$ is \mathcal{H} -measurable and λ -integrable, then $h_x: Y \rightarrow \mathbf{R}^*$ is ν -integrable for almost all x and $h_y: X \rightarrow \mathbf{R}^*$ is μ -integrable for almost all y . Further*

$$\int h \, d\lambda = \int f \, d\mu = \int g \, d\nu,$$

where $f(x) = \int h_x \, d\nu$ when h_x is ν -integrable,

$$g(y) = \int h_y \, d\mu \text{ when } h_y \text{ is } \mu\text{-integrable}$$

and f, g are defined to be zero on the remaining null sets.

(iii) *If $f: X \times Y \rightarrow \mathbf{R}^*$ is \mathcal{H} -measurable and $\int (\int |f_x| \, d\nu) \, d\mu$ is finite, then*

$$\int f \, d\lambda = \int \left(\int f_y \, d\mu \right) d\nu = \int \left(\int f_x \, d\nu \right) d\mu.$$

Proof (i). If h is the indicator function of a set in \mathcal{H} the result follows by theorem 6.4. Because of the linearity of the integration process it now follows for non-negative \mathcal{H} -simple functions (note that sections of an \mathcal{H} -simple function will be simple by theorem 6.1). If we now take a sequence $\{h^{(n)}\}$ of non-negative simple functions increasing to h , we will have the sections $\{h_x^{(n)}\}, \{h_y^{(n)}\}$ increasing to h_x, h_y respectively. Hence, as $n \rightarrow \infty$,

$$\int h^{(n)} \, d\lambda \rightarrow \int h \, d\lambda,$$

$$\int h_x^{(n)} \, d\nu \rightarrow \int h_x \, d\nu \text{ for all } x, \quad \int h_y^{(n)} \, d\mu \rightarrow \int h_y \, d\mu \text{ for all } y,$$

and an application of the monotone convergence theorem (5.6) now suffices to complete the proof.

(ii) Since h is integrable, the positive and negative parts h^+, h^- are integrable. Apply (i) to each of these functions. Then

$$f^+(x) = \int h_x^+ \, d\nu$$

will always be defined, though it may take the value $+\infty$. Since $\int f^+(x) d\mu$ exists, we must have f^+ finite except for a set of zero μ -measure. Similarly, f^- is finite almost everywhere. If we put

$$f(x) = f^+(x) - f^-(x)$$

when both f^+, f^- are finite and $f(x) = 0$ otherwise, we see that

$$\begin{aligned} \int h d\lambda &= \int h^+ d\lambda - \int h^- d\lambda \\ &= \int f^+ d\mu - \int f^- d\mu = \int f d\mu. \end{aligned}$$

(iii) Again split f into positive and negative parts. Since

$$0 \leq f^+ \leq |f|,$$

we can apply (i) to each of the positive and negative parts to deduce that $\int f^+ d\lambda$ and $\int f^- d\lambda$ are both finite. The result now follows by (ii). **]**

We should remark that theorem 6.5 is one of the most useful tools in the theory of integration as we have developed it. This result again exhibits the power and neatness of the absolute integral.

We have been careful to define the product measure λ on the smallest σ -field \mathcal{H} which contains $\mathcal{F} \times \mathcal{G}$. Some authors define product measure to be the completion of this λ obtained by the process of theorem 4.3. If one uses this definition then some of our statements have to be modified to exclude possible subsets of zero measure, though the essential content of the results remain valid. In particular, given a function $f(x, y)$ which is measurable with respect to the completed σ -field \mathcal{H} , one can only say that the section f_x is \mathcal{G} -measurable for almost all x . However, provided \mathcal{F} and \mathcal{G} are complete with respect to their respective measures, theorem 6.5 remains valid as stated.

We can use our definition of product measure to give an alternative definition of the integral of a non-negative measurable function.

Theorem 6.6. *Suppose $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space, $(\mathbf{R}, \mathcal{L}, \nu)$ denotes the real line with Lebesgue measure on it and τ is the product measure $\mu \times \nu$ defined on the product σ -field \mathcal{H} in $\Omega \times \mathbf{R}$. Then if $E \in \mathcal{F}$ and $f: E \rightarrow \mathbf{R}^+$ is non-negative, f is \mathcal{F} -measurable over E if and only if $Q(E, f) \in \mathcal{H}$, and in this case,*

$$\int_E f d\mu = \tau(Q(E, f));$$

where $Q(E, f)$ is the ordinate set defined by

$$\{(x, y): x \in E, y \in \mathbf{R}, 0 \leq y < f(x)\}.$$

Proof. Suppose first that $Q(E, f) \in \mathcal{H}$. Then by theorem 6.1 all its sections are in \mathcal{F} . But the section of $Q(E, f)$ at $y = a$ is the set

$$\{x: f(x) > a\},$$

so that by definition, f is \mathcal{F} -measurable. Conversely, if f is \mathcal{F} -measurable then there is a sequence $\{f_n\}$ of \mathcal{F} -simple non-negative functions which increases to f . Now for any \mathcal{F} -simple function $Q(E, f_n)$ is a finite union of measurable rectangles and is therefore in \mathcal{H} . Also $Q(E, f_n)$ increases monotonely to $Q(E, f)$ so we must have $Q(E, f) \in \mathcal{H}$. Further if

$$f_n = \sum_{i=1}^{r_n} C_{n,i} \chi_{E_{n,i}},$$

with $E_{n,i}$ a disjoint partition of E ,

$$\int_E f_n d\mu = \sum C_{n,i} \mu(E_{n,i}) = \sum \tau(E_{n,i} \times [0, C_{n,i})) = \tau(Q(E, f_n)).$$

If we now let $n \rightarrow \infty$ we obtain the desired result. \blacksquare

Corollary. If $f: \mathbf{R} \rightarrow \mathbf{R}^+$ is \mathcal{L} -measurable, then the ordinate set $\{(x, y): a \leq x \leq b, 0 \leq y < f(x)\}$ is \mathcal{L}^2 -measurable and has planar

Lebesgue measure $\int_a^b f(x) dx$.

In many elementary accounts of integration the notion of 'area under the curve' is intuitively important. This last corollary makes this notion rigorous for the Lebesgue integral of non-negative functions.

It is possible to consider Euclidean k -dimensional space \mathbf{R}^k as the Cartesian product of k distinct spaces \mathbf{R} . Since we have a natural measure $(\mathbf{R}, \mathcal{L}, \nu)$ on each of these spaces we could form the product measure defined on \mathcal{F}^k the product σ -field in \mathbf{R}^k by the process of theorem 6.2. How does this measure compare with Lebesgue measure in \mathbf{R}^k ? Since all the extension processes used are unique, and the two measures clearly coincide on $\mathcal{P}^k = \mathcal{P} \times \mathcal{P} \times \dots \times \mathcal{P}$, the half-open rectangles in \mathbf{R}^k , it is clear that the two measures coincide whenever both are defined. However, \mathcal{L}^k is complete with respect to Lebesgue measure while \mathcal{F}^k is not known to be so. To see that \mathcal{F}^k is not complete it is sufficient to consider the product of a linear set which is not measurable in \mathbf{R} with $(k-1)$ single point sets. This set cannot be in the product σ -field by exercise 6.1 (7), but it is a subset of a line in \mathbf{R}^k and therefore it must be in \mathcal{L}^k . It follows that \mathcal{L}^k is a larger σ -field than \mathcal{F}^k . Since \mathcal{B}^k , the class of Borel sets in \mathbf{R}^k is the σ -field generated by \mathcal{P}^k , we also have $\mathcal{F}^k \supset \mathcal{B}^k$. If E is any set in \mathcal{L}^1 but not in \mathcal{B}^1 the Cartesian product of E with $(k-1)$ whole lines \mathbf{R} will be in \mathcal{F}^k but not in \mathcal{B}^k , so that \mathcal{F}^k is a larger σ -field than \mathcal{B}^k .

If we consider the case $k = 2$, a function $f(x, y)$ which is \mathcal{L}^2 -measurable need not be \mathcal{F}^2 -measurable. Thus we can only say that the function $f_x(y) = f(x, y)$ considered as a function of y for fixed x is \mathcal{L} -measurable for almost all x . Thus in Theorem 6.5 (ii), if $f(x, y)$ is Lebesgue integrable we can deduce that $\phi(x) = \int f(x, y) dy$ exists and is finite except for an exceptional set of x of zero measure. As $\phi(x)$ is thus defined a.e. it can be integrated and

$$\int \int f(x, y) dx dy = \int \phi(x) dx.$$

Exercises 6.3

1. Suppose Ω is any set of cardinal greater than \aleph_0 , and \mathcal{F} is the σ -field of sets in Ω which are either countable or have a countable complement. For $E \in \mathcal{F}$, put $\mu(E) = 0$ if E is countable, $\mu(E) = 1$ if $(\Omega - E)$ is countable. Consider the Cartesian product of two copies of Ω and let E be a set in $\Omega \times \Omega$ which has countable x -sections for every x and y -sections whose complement is countable for every y . If h is the indicator function of E , then

$$\int h_y(x) \mu(dx) = 1, \quad \int h_x(y) \mu(dy) = 0.$$

Why does this not contradict theorem 6.4?

2. Suppose (X, \mathcal{F}, μ) (Y, \mathcal{G}, ν) are σ -finite measure spaces and λ is the product measure on the product σ -field \mathcal{H} . Show that

(i) If $E, G \in \mathcal{H}$ are such that $\nu(E_x) = \nu(G_x)$ for almost all $x \in X$, then $\lambda(E) = \lambda(G)$.

(ii) If f, g are integrable functions on X, Y then $f(x)g(y)$ is integrable on $X \times Y$ and

$$\int f(x)g(y) d\lambda = \int f d\mu \int g d\nu.$$

3. $X = Y = [0, 1]$ and \mathcal{F}, \mathcal{G} are the Borel subsets. Let $\mu(E)$ be the Lebesgue measure of E , $\nu(E)$ the number of points in E . Form the product measure $\mu \times \nu$ on Borel subsets of the unit square. Then if D is the diagonal $\{(x, y); x = y\}$, D is measurable and

$$\int \nu(D_x) \mu(dx) = 1, \quad \int \mu(D^y) \nu(dy) = 0.$$

Why does this not contradict theorem 6.4?

4. If $f(x, y) = (x^2 - y^2)/(x^2 + y^2)^2$ show that

$$\int_0^1 \left\{ \int_0^1 f(x, y) dy \right\} dx = \frac{\pi}{4},$$

$$\int_0^1 \left\{ \int_0^1 f(x, y) dx \right\} dy = -\frac{\pi}{4},$$

where all the integrals are taken in the Lebesgue sense. Thus theorem 6.5 (iii) is not valid without the modulus sign. Similarly, show that

$$\int_0^1 \left\{ \int_1^\infty (e^{-xy} - 2e^{-2xy}) dx \right\} dy \neq \int_1^\infty \left\{ \int_0^1 (e^{-xy} - 2e^{-2xy}) dy \right\} dx.$$

5. If $f(x, y) = xy/(x^2 + y^2)^2$, then

$$\int_{-1}^{+1} \left\{ \int_{-1}^{+1} f(x, y) dx \right\} dy = 0 = \int_{-1}^{+1} \left\{ \int_{-1}^{+1} f(x, y) dy \right\} dx$$

but the integral over the unit square in \mathbf{R}^2 does not exist.

6. Given a countable collection of probability spaces $(X_i, \mathcal{F}_i, \mu_i)$ and the product measure μ on the product σ -field, we can form the finite product measures $\tau_n = \mu_1 \times \mu_2 \times \dots \times \mu_n$ and the product measure λ_n on the product space $\prod_{i=n+1}^\infty X_i$. Then, if $f(x_1, x_2, \dots)$ is any μ -integrable function on $\prod_{i=1}^\infty X_i$ we have

$$\int f d\mu = \int \left\{ \int f(x_1, x_2, \dots) d\lambda_n \right\} d\tau_n.$$

6.4 Radon-Nikodym theorem

We start with a definition.

Absolute continuity

Suppose \mathcal{F} is a σ -ring of subsets of Ω and μ is a measure on \mathcal{F} . Then the set function $\nu: \mathcal{F} \rightarrow \mathbf{R}^*$ is said to be absolutely continuous with respect to μ if $\nu(E) = 0$ for every E in \mathcal{F} with $\mu(E) = 0$. In this case we write $\nu \ll \mu$. If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $f: \Omega \rightarrow \mathbf{R}^*$ is μ -integrable, then it is clear that

$$\nu(E) = \int_E f d\mu \quad (E \in \mathcal{F}),$$

defines a finite valued absolutely continuous set function ν . In fact, in §5.4 we proved that ν was σ -additive, and that (corollary to theorem 5.6) given $\epsilon > 0$, there is a $\delta > 0$ such that for $E \in \mathcal{F}$,

$$\mu(E) < \delta \Rightarrow |\nu(E)| < \epsilon. \quad (6.4.1)$$

It is immediate that any set function ν which satisfies (6.4.1) is absolutely continuous with respect to μ . The conditions are equivalent for finite measures, but not in general (see exercise 6.4 (4)). There is a partial converse given by:

Lemma. *If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $\nu: \mathcal{F} \rightarrow \mathbf{R}$ is finite valued, σ -additive and absolutely continuous with respect to μ , then ν satisfies condition (6.4.1).*

Proof. By the decomposition of §3.2, any such ν is the difference of two finite measures, so it is sufficient to prove the result for a measure ν . Then if (6.4.1) is false, there is an $\epsilon > 0$ and a sequence $\{E_n\}$ of sets of \mathcal{F} such that $\nu(E_n) > \epsilon$ and $\mu(E_n) < 2^{-n}$. Put $E = \limsup E_n$. Then

$$\mu(E) \leq \mu\left(\bigcup_{r=n+1}^{\infty} E_r\right) \leq \sum_{n+1}^{\infty} \mu(E_r) < 2^{-n},$$

so that $\mu(E) = 0$ while

$$\nu(E) = \lim \nu\left(\bigcup_{r=n+1}^{\infty} E_r\right) \geq \limsup \nu(E_r)$$

so that $\nu(E) \geq \epsilon$. This contradicts $\nu \ll \mu$. **■**

Thus we see that the indefinite integral of an integrable function defines an absolutely continuous set function. Our object in the present section is to obtain the converse of this statement under suitable conditions. It is convenient at the same time to consider a more general σ -additive set function and to decompose it into a maximal absolutely continuous component and a remainder which has to be concentrated on a μ -null set. It is convenient to give a further definition.

Singular set function

Given a measure space $(\Omega, \mathcal{F}, \mu)$ a set function $\nu: \mathcal{F} \rightarrow \mathbf{R}^*$ is said to be singular with respect to μ if there is a set $E_0 \in \mathcal{F}$ for which $\mu(E_0) = 0$ and

$$\nu(E) = \nu(E \cap E_0), \quad \text{all } E \in \mathcal{F}. \quad (6.4.2)$$

This condition clearly means that the parts of Ω outside the null set E_0 make no contribution to ν . In fact if ν is also a measure we see that Ω can be dissected into two sets $E_0, E_1 \in \mathcal{F}$ such that

$$E_0 \cap E_1 = \emptyset, \quad E_0 \cup E_1 = \Omega, \quad \mu(E_0) = 0, \quad \nu(E_1) = 0.$$

The symmetry of the relationship in this case is sometimes stressed by saying that μ and ν are *mutually singular*.

Theorem 6.7. *Given a σ -finite measure space $(\Omega, \mathcal{F}, \mu)$ and a σ -additive, σ -finite set function ν , then there is a unique decomposition*

$$\nu = \nu_1 + \nu_2$$

into σ -additive set functions ν_i which are σ -finite and such that ν_1 is singular with respect to μ and $\nu_2 \ll \mu$. Further there is a finite valued measurable $f: \Omega \rightarrow \mathbf{R}$ such that

$$\nu_2(E) = \int_E f d\mu, \quad \text{all } E \in \mathcal{F}.$$

The function f is unique in the sense that if we also have

$$\nu_2(E) = \int_E g d\mu$$

for all E in \mathcal{F} , then $f(x) = g(x)$ except in a set of zero μ -measure.

Corollary. Under the conditions of the theorem if $\nu \ll \mu$ then there is a finite valued $f: \Omega \rightarrow \mathbf{R}$ such that

$$\nu(E) = \int_E f d\mu \quad \text{for } E \in \mathcal{F}.$$

Note. The decomposition of ν into absolutely continuous and singular components is often called the Lebesgue decomposition, while the integral representation is called the Radon–Nikodym theorem.

Proof. Since we can express Ω as a union of a countable set of disjoint sets on each of which both μ and ν are finite, there is no loss in generality in assuming that they are both finite on Ω . This applies to both the existence and uniqueness proofs. We first see that the decomposition is unique.

Let

$$\nu = \nu_1 + \nu_2 = \nu_3 + \nu_4,$$

where ν_1, ν_3 are singular and ν_2, ν_4 are absolutely continuous. Then

$$\nu_1 - \nu_3 = \nu_4 - \nu_2,$$

Taking the union of support sets of ν_1, ν_3 gives a set E_0 such that

$$(\nu_1 - \nu_3)(E) = (\nu_1 - \nu_3)(E \cap E_0), \quad \mu(E_0) = 0.$$

But $(\nu_4 - \nu_2)$ is absolutely continuous and therefore zero on any null set so that, for any $E \in \mathcal{F}$,

$$\begin{aligned} (\nu_4 - \nu_2)(E) &= (\nu_1 - \nu_3)(E) = (\nu_1 - \nu_3)(E \cap E_0) \\ &= (\nu_4 - \nu_2)(E \cap E_0) = 0. \end{aligned}$$

Thus $\nu_1(E) = \nu_3(E)$, $\nu_2(E) = \nu_4(E)$ for all E . The uniqueness of the integral representation of ν_2 was proved in §5.4. Thus it is sufficient to find any decomposition and integral representation.

By theorem 3.3 we can decompose ν into the difference of two measures. It is therefore sufficient to prove the theorem when ν is a measure. Now let \mathcal{H} be the class of non-negative measurable

$$f: \Omega \rightarrow \mathbf{R}^+$$

such that

$$\nu(E) \geq \int_E f d\mu \quad \text{for all } E \text{ in } \mathcal{F}$$

and put

$$\alpha = \sup \left\{ \int f d\mu : f \in \mathcal{H} \right\}.$$

Let $\{f_n\}$ be a sequence of functions in \mathcal{H} such that

$$\int f_n d\mu > \alpha - \frac{1}{n}.$$

Put $g_n(x) = \max\{f_1(x), f_2(x), \dots, f_n(x)\}$. Then if $E \in \mathcal{F}$, and n is fixed we can decompose E into a disjoint union $E_1 \cup E_2 \cup \dots \cup E_n$ of sets of \mathcal{F} such that $g_n = f_j$ on E_j . Hence

$$\int_E g_n d\mu = \sum_{j=1}^n \int_{E_j} g_n d\mu = \sum_{j=1}^n \int_{E_j} f_j d\mu \leq \sum_{j=1}^n \nu(E_j) = \nu(E),$$

so that $g_n \in \mathcal{H}$ for all n . But $\{g_n\}$ is monotone increasing, and by the monotone convergence theorem, $f_0(x) = \lim_{n \rightarrow \infty} g_n(x) \in \mathcal{H}$. Since

$$f_0(x) \geq f_n(x) \quad \text{for all } n,$$

we must have

$$\alpha = \int f_0(x) d\mu.$$

For each E in \mathcal{F} , put

$$\nu_2(E) = \int_E f_0 d\mu, \quad \nu_1(E) = \nu(E) - \nu_2(E).$$

Then ν_2 is absolutely continuous with respect to μ , so it only remains to show that ν_1 is singular.

Consider the σ -additive set function

$$\lambda_n = \nu_1 - (1/n)\mu$$

and decompose Ω , using theorem 3.3. into positive and negative sets P_n, N_n such that $P_n \cup N_n = \Omega$, $P_n \cap N_n = \emptyset$, $E \subset P_n \Rightarrow \lambda_n(E) \geq 0$, $E \subset N_n \Rightarrow \lambda_n(E) \leq 0$. Then, for $E \subset P_n$,

$$\nu(E) = \nu_1(E) + \nu_2(E) \geq \nu_2(E) + \frac{1}{n} \mu(E) = \int_E \left(f_0 + \frac{1}{n} \right) d\mu.$$

This shows that the function equal to f_0 on N_n and $[f_0 + (1/n)]$ on P_n is in \mathcal{H} . This will give a larger integral than α unless $\mu(P_n) = 0$. If $P = \bigcup_{n=1}^{\infty} P_n$, then $\mu(P) = 0$. Further $\Omega - P \subset N_n$ for all n so that $\nu_1(\Omega - P) = 0$ and

$$\nu_1(E) = \nu_1(E \cap P) \quad \text{for all } E \text{ in } \mathcal{F},$$

that is, ν_1 is μ -singular.

In the case where $\nu \ll \mu$, by the uniqueness of the decomposition we must have $\nu = \nu_2$, and the integral representation of ν now follows. **]**

Remark. In the statement of theorem 6.7 we do not assert that the function f is integrable. A necessary and sufficient condition that f be integrable is that ν be finite. However, the use of the symbol $\int_E f d\mu$ asserts that either f_+ or f_- has a finite integral. This corresponds to the result of theorem 3.2 that ν cannot take both the values $\pm \infty$.

Derivative of a set function

If $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space and

$$\nu(E) = \int_E f d\mu \quad \text{for } E \text{ in } \mathcal{F},$$

then we write $f = d\nu/d\mu$ and call f the Radon–Nikodym derivative of ν with respect to μ .

One should emphasise that the derivative $d\nu/d\mu$ is not defined uniquely at any given point, it has to be considered as a function and then it becomes uniquely defined in the sense that any two functions representing the same derivative can differ only on a μ -null set.

Exercises 6.4

1. Show that if μ, ν are any two measures on a σ -ring \mathcal{S} , then $\nu \ll \mu + \nu$.
2. Suppose $F(x)$ is the Cantor function defined in §2.7 and ν is the Lebesgue–Stieltjes measure with respect to F . Show that ν is singular with respect to Lebesgue measure.
3. Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space with $\mu(\Omega) < \infty$ and ν is a measure, $\nu \ll \mu$. Show there is a set E such that $(\Omega - E)$ has σ -finite ν measure and for every measurable $F \subset E$, $\nu(F)$ is either 0 or ∞ .
4. Let Ω be the set of positive integers,

$$\mu(E) = \sum_{n \in E} 2^{-n}, \quad \nu(E) = \sum_{n \in E} 2^n$$

then $\nu \ll \mu$, but (6.4.1) is not satisfied. This shows that (6.4.1) is a stronger condition than absolute continuity when ν is not finite.

5. Suppose Ω is an uncountable set, \mathcal{F} is the class of sets which are either countable or have countable complements. For $E \in \mathcal{F}$, put $\mu(E) =$ the number of points in E , $\nu(E) = 0$ or 1 according as E is countable or not. Then clearly $\nu \ll \mu$, but no integral representation is possible. This shows that in the Radon–Nikodym theorem we cannot do without the condition that μ be σ -finite.

6. If λ, μ, ν are σ -finite measures on \mathcal{F} and $\lambda \ll \mu, \mu \ll \nu$; show that $\lambda \ll \nu$ and

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

except on a set of zero λ -measure.

7. λ, μ are σ -finite measures on \mathcal{F} with $\mu \ll \lambda$. Then if f is μ -integrable

$$\int f d\mu = \int f \frac{d\mu}{d\lambda} d\lambda.$$

8. If λ, μ are σ -finite measures on \mathcal{F} such that $\mu \ll \lambda$ and $\lambda \ll \mu$ then

$$\frac{d\mu}{d\lambda} = \left(\frac{d\lambda}{d\mu} \right)^{-1},$$

except for a set of zero λ -measure.

9. If μ, ν are σ -finite measures on \mathcal{F} such that $\nu \ll \mu$, show that the set of points x at which $d\nu/d\mu$ is zero has zero ν -measure.

10. Suppose $\{\mu_i\}$ is a countable family of finite measures on a σ -field \mathcal{F} . Show that there exists a finite μ on \mathcal{F} such that each of the μ_i is absolutely continuous with respect to μ .

11. Suppose

$$\bar{\mu}_n = \sum_{k=1}^n \mu_k \rightarrow \bar{\mu}, \quad \bar{\nu}_n = \sum_{k=1}^n \nu_k \rightarrow \bar{\nu},$$

where all the μ, ν with suffices are finite measures on a σ -field \mathcal{F} and $\bar{\nu}_n$ is $\bar{\mu}_n$ -continuous for all n . Show that

- (i) $d\mu_1/d\bar{\mu}_n \rightarrow d\mu_1/d\bar{\mu}$ almost everywhere $(\bar{\mu})$.
- (ii) If each μ_n is ν -continuous then $d\bar{\mu}_n/d\nu \rightarrow d\bar{\mu}/d\nu$ a.e. (ν) .
- (iii) $\bar{\nu}$ is $\bar{\mu}$ -continuous and $d\bar{\nu}_n/d\bar{\mu}_n \rightarrow d\bar{\nu}/d\bar{\mu}$ a.e. $(\bar{\mu})$.

6.5 Mappings of measure spaces

In mathematical arguments one often needs to consider two spaces, X, Y with a mapping $f: X \rightarrow Y$. Such a mapping induces mappings on the classes of subsets of X and Y : if $E \subset X$, $f(E)$ denotes the set of y in Y with $y = f(x)$, $x \in E$, and if $F \subset Y$, $f^{-1}(F)$ denotes the set of x in X with $f(x) \in F$; further if \mathcal{C} is a class of subsets of X , $f(\mathcal{C})$ denotes the class of sets $f(E)$ with $E \in \mathcal{C}$, and similarly for $f^{-1}(\mathcal{C})$ where \mathcal{C} is a class of subsets of Y . We saw (§1.5) that f^{-1} preserves the structure of a class of subsets, so that if \mathcal{S} is a σ -field in Y , $f^{-1}(\mathcal{S})$ is a σ -field in X . Sometimes the two spaces X, Y already have classes of subsets defined, and one can then examine the relationship of the mapping f to these.

Measurable transformation

Suppose f is a mapping from X into Y , \mathcal{F} is a σ -field in X and \mathcal{G} is a σ -field in Y , then we say that f is a measurable transformation from (X, \mathcal{F}) into (Y, \mathcal{G}) if $f^{-1}(E) \in \mathcal{F}$ for every E in \mathcal{G} . This condition can also be written $f^{-1}(\mathcal{G}) \subset \mathcal{F}$.

In Chapter 5 we discussed 'measurable functions'. In our new terminology these are measurable transformations from (X, \mathcal{F}) to $(\mathbf{R}^*, \mathcal{G})$ in which \mathcal{G} is the σ -field of Borel sets in \mathbf{R}^* . Given mappings $f: X \rightarrow Y$, $g: Y \rightarrow Z$ we can consider the composition $g(f): X \rightarrow Z$ defined by $g(f)(x) = g(f(x))$. In particular if $g: Y \rightarrow \mathbf{R}^*$ is an extended real-valued function on Y , then $g(f)$ defines an extended real function on X .

Lemma. *If $f: X \rightarrow Y$ is a measurable transformation from (X, \mathcal{F}) into (Y, \mathcal{G}) and $g: Y \rightarrow \mathbf{R}^*$ is \mathcal{G} -measurable as a function with extended real values, then the composition $g(f)$ is \mathcal{F} -measurable.*

Proof. For any Borel set B in \mathbf{R}^* we have

$$\begin{aligned}\{x: g(f)(x) \in B\} &= f^{-1}\{y: g(y) \in B\} \\ &= f^{-1}(E) \quad \text{for some } E \in \mathcal{G},\end{aligned}$$

and is therefore in \mathcal{F} . **■**

Remark. We obtained a special case of this lemma when we proved that a Borel measurable function of a measurable function is measurable (see §5.2).

If we start with a measure space (X, \mathcal{F}, μ) and f is a measurable transformation from (X, \mathcal{F}) into (Y, \mathcal{G}) it is natural to use f to define a measure ν on \mathcal{G} by putting

$$\nu(E) = \mu(f^{-1}(E)) \quad \text{for } E \in \mathcal{G}. \quad (6.5.1)$$

With this definition of ν it is immediate that (Y, \mathcal{G}, ν) is a measure space. If (6.5.1) holds we will write $\nu = \mu f^{-1}$. This allows us to carry out a 'change of variable' in an integral.

Theorem 6.8. *Suppose f is a measurable transformation from a measure space (X, \mathcal{F}, μ) to (Y, \mathcal{G}) and $g: Y \rightarrow \mathbf{R}^*$ is \mathcal{G} -measurable: then*

$$\int_Y g d(\mu f^{-1}) = \int_X g(f) d\mu$$

in the sense that if either integral exists so does the other and the two are equal

Proof. It is clearly sufficient to consider non-negative functions

$g: Y \rightarrow \mathbf{R}^+$. Suppose first that $g = \chi_E$, the indicator function of a set E in \mathcal{G} . Then

$$\begin{aligned} g(f)(x) &= 1 & \text{if } x \in f^{-1}(E), \\ &= 0 & \text{if } x \notin f^{-1}(E); \end{aligned}$$

so that $g(f)$ is the indicator function of $f^{-1}(E)$, a set in \mathcal{F} . Thus, in this case, by (6.5.1)

$$\int g d(\mu f^{-1}) = \mu f^{-1}(E) = \mu(f^{-1}(E)) = \int g(f) d\mu.$$

By linearity, the result now follows for non-negative \mathcal{G} -simple functions g . If $\{g_n\}$ is an increasing sequence of non-negative simple functions converging to the measurable function g , then $g_n(f)$ will be an increasing sequence of simple functions converging to $g(f)$. The definition of the integral of a non-negative function now completes the proof. \blacksquare

Sometimes in integration, when the variable is changed, one wants to integrate with respect to a new measure $\nu \neq \mu f^{-1}$. We can do this easily when μf^{-1} is absolutely continuous with respect to ν .

Theorem 6.9. *Given σ -finite measure spaces (X, \mathcal{F}, μ) and (Y, \mathcal{G}, ν) and a measurable transformation f from (X, \mathcal{F}) into (Y, \mathcal{G}) such that μf^{-1} is absolutely continuous with respect to ν*

$$\int g(f) d\mu = \int g \cdot \phi d\nu,$$

where ϕ is the Radon-Nikodym derivative $d(\mu f^{-1})/d\nu$, for every measurable $g: Y \rightarrow \mathbf{R}^+$ in the sense that, if either integral exists, so does the other and the two are equal.

Corollary. *If $q: \mathbf{R} \rightarrow \mathbf{R}^+$ is Lebesgue integrable,*

$$F(x) = \int_{-\infty}^x q(t) dt,$$

and μ_F is the Lebesgue-Stieltjes measure generated by F , then

$$\int_A^B g(x) dx = \int_a^b g(F(t)) d\mu_F = \int_a^b g(F(t)) q(t) dt$$

where $A = F(a)$, $B = F(b)$.

Proof. By theorem 6.8 we have

$$\int g(f) d\mu = \int g d(\mu f^{-1}).$$

Since μf^{-1} is absolutely continuous with respect to ν , there exists a measurable ϕ such that, for every $E \in \mathcal{G}$

$$\int_E \phi d\nu = (\mu f^{-1})(E).$$

If g is the indicator function of a measurable set E it now follows that

$$\int g d(\mu f^{-1}) = (\mu f^{-1})(E) = \int g \cdot \phi d\nu$$

and the required result now follows by successive extension to functions g which are: (i) non-negative, simple, (ii) non-negative, measurable, (iii) measurable.

Under the conditions of the corollary we consider the mapping $F: \mathbf{R} \rightarrow \mathbf{R}$ given by the measure function F from the Lebesgue measure space $(\mathbf{R}, \mathcal{L}, \mu_F)$ to $(\mathbf{R}, \mathcal{B}, \mu)$. Theorem 6.8 then gives the first equality. If we define

$$\lambda(E) = \int_E q d\mu,$$

then $\lambda: \mathcal{L} \rightarrow \mathbf{R}$ is a measure which coincides with

$$\mu_F(E) = \int_E 1 d\mu_F$$

for intervals of \mathcal{P} and therefore for all sets $E \in \mathcal{L}$. Hence the measure μ_F is absolutely continuous with respect to Lebesgue measure μ and q is a possible definition of the Radon–Nikodym derivative $d\mu_F/d\mu$. The second equality now follows from theorem 6.9. \square

Remark. It is clear from the above that the function ϕ (or q in the corollary) plays the part of the Jacobian (or rather the absolute value of the Jacobian) in the theory of transformations of multiple integrals. In general it is not easy to obtain an explicit value for the Radon–Nikodym derivative $d(\mu f^{-1})/d\nu$, but in important special cases this can be done. In particular, if both spaces are $(\mathbf{R}^k, \mathcal{L}^k, \mu)$ with μ Lebesgue measure, and $f: \mathbf{R}^k \rightarrow \mathbf{R}^k$ is a linear transformation given by a non-singular matrix A so that $y = Ax$ one can prove that

$$|f(E)| = \|A\| \cdot |E|,$$

where $\|A\|$ denotes the absolute value of the determinant of A . (This can best be shown by expressing A as a product of elementary transformations, and proving the result for each elementary transformation.) This means that, in this case a possible Radon–Nikodym derivative is the constant function $\|A\|$.

Exercises 6.5

1. Show that the composition of two measurable transformations is measurable.

2. If f is a measurable transformation from (X, \mathcal{F}) into (Y, \mathcal{G}) and μ, ν are two measures on \mathcal{F} such that $\mu \ll \nu$, show that $\mu f^{-1} \ll \nu f^{-1}$.

3. (*Integration by parts.*) If $F(x), G(x)$ are non-negative continuous functions satisfying the conditions of §4.5 for a Stieltjes measure function and E is any Borel set, then

$$\int_E F(x) dG(x) + \int_E G(x) dF(x) = \mu_{FG}(E),$$

where μ_{FG} denotes the Lebesgue-Stieltjes measure generated by $F(x)G(x)$. In particular if

$$F(x) = \int_a^x f(t) dt, \quad G(x) = \int_a^x g(t) dt,$$

then
$$\int_a^b F(x) g(x) dx + \int_a^b f(x) G(x) dx = F(b)G(b) - F(a)G(a).$$

4. Suppose A is a non-singular $k \times k$ -matrix defining a mapping from \mathbf{R}^k to \mathbf{R}^k , then this is a measurable transformation from $(\mathbf{R}^k, \mathcal{L}^k)$ to itself. If μ_k denotes Lebesgue measure in \mathbf{R}^k show that

$$\int f d\mu_k = \|A\| \int f(A) d\mu_k,$$

for any Lebesgue measurable f , where $f(A)$ denotes the composite map $f(A)(x) = f(Ax)$.

6.6* Measure in function space

We saw that points in the product space $\prod_{i \in I} X_i$ can be thought of as functions $f: I \rightarrow \bigcup_{i \in I} X_i$ in which $f(i) \in X_i$. In the particular case where X_i is the same space X for all i , the space $\prod_{i \in I} X_i$ reduces to the set of functions: $I \rightarrow X$. For this reason such a product space is often denoted by X^I . Since theorem 6.3 clearly extends to arbitrary product spaces we can produce a product measure in X^I starting from any measure μ on X with $\mu(X) = 1$. However, for non-countable I , such product measures are rarely of interest. In applications, the space X^I usually describes a stochastic process (see Chapter 15), and the product measure in X^I would correspond to complete independence (see Chapter 11) between the values in each of the coordinate spaces. Usually one wants to be able to define and use measures in X^I which are not product measures.

In our account we restrict X to be the real line \mathbf{R} (it is easy to extend the theory to the case $X = \mathbf{C}$, but some restriction is needed for its validity), leaving the index set I completely arbitrary.

Borel sets in \mathbf{R}^I

If we assume the usual topology in \mathbf{R} , and denote the class of Borel sets in \mathbf{R} by \mathcal{B} , then the class \mathcal{C} of cylinder sets

$$\{f \in \mathbf{R}^I: f(i_k) \in B_k, \quad k = 1, 2, \dots, n\}, \quad B_k \in \mathcal{B}$$

is a semi-ring of subsets in \mathbf{R}^I . The σ -field generated by \mathcal{C} will be denoted by \mathcal{B}^I . If \mathcal{B}^n denotes the class of Borel sets in \mathbf{R}^n , it is immediate that \mathcal{B}^I can also be generated by the class of sets of the form

$$\{f \in \mathbf{R}^I: a_k < f(i_k) \leq b_k, \quad k = 1, 2, \dots, n\}, \quad (6.6.1)$$

or of the form

$$\{f \in \mathbf{R}^I: (f(i_1), f(i_2), \dots, f(i_n)) \in B^n\}, \quad B^n \in \mathcal{B}^n. \quad (6.6.2)$$

It is important to notice that no set in \mathcal{B}^I can have restrictions on an uncountable set of coordinates. For, if E is a countable subset of I and $F = I - E$, a set of the form

$$\{f \in \mathbf{R}^I: f_E \in \mathcal{B}^E\}, \quad (6.6.3)$$

where f_E denotes the restriction of f to E , contains functions f which are not restricted on F . The class of subsets of \mathbf{R}^I of the form (6.6.3) (for all possible countable sets $E \subset I$) is clearly a σ -field which contains the finite dimensional cylinder sets \mathcal{C} . Further, every set of the form (6.6.3) must be in \mathcal{B}^I , so that the Borel sets in \mathbf{R}^I are precisely the sets of the form (6.6.3).

Our object will be to extend a measure which is already defined on sets of the form (6.6.1) to the σ -field \mathcal{B}^I . For a fixed finite set $i_1, i_2, \dots, i_n \in I$, the sets of the form (6.6.1) clearly generate a σ -field containing those sets of \mathbf{R}^I obtained by taking a Borel set in $\mathbf{R}_{i_1} \times \mathbf{R}_{i_2} \times \dots \times \mathbf{R}_{i_n}$ and forming the cylinder with this set as base. If we are to have $\mu(\mathbf{R}^I) = 1$, then, for each fixed i_1, i_2, \dots, i_n , our set function on sets of the form (6.6.2) must define a measure on the Borel sets of the Euclidean n -space $\mathbf{R}_{i_1} \times \dots \times \mathbf{R}_{i_n}$ in which the whole space has measure 1.

It is clear that the measures given in the various Euclidean spaces of this type have to satisfy various *consistency relations*, if there is to be any hope of extending to a single measure on the whole of \mathcal{B}^I . For such a measure on \mathcal{B}^I must yield the original system on restriction to sets of the form (6.6.2). These consistency conditions can be stated

in terms of multidimensional distribution functions which generate the measures on sets (6.6.2), but we prefer to state them (equivalently) in terms of the measures.

We assume then that for each finite set of distinct indices i_1, i_2, \dots, i_n we have a measure $\mu_{i_1 i_2 \dots i_n}$ defined for the Borel sets in \mathbf{R}^n such that

(I) $\mu_{i_1 \dots i_n i_{n+1}}(A \times \mathbf{R}) = \mu_{i_1 \dots i_n}(A)$, $A \in \mathcal{B}^n$.

(II) If π is a permutation of $(1, 2, \dots, n)$ and $\phi: \mathbf{R}^n \rightarrow \mathbf{R}^n$ is the mapping

$$\phi(x_1, \dots, x_n) = (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_n}),$$

then

$$\mu_{i_{\pi_1} i_{\pi_2} \dots i_{\pi_n}} = \mu_{i_1 i_2 \dots i_n} \phi^{-1}.$$

The condition (I) says that putting on the additional condition $f(i_{n+1}) \in \mathbf{R}$ at a new index cannot effect the measure of the set since it imposes no restriction, and condition (II) makes precise the notion that the order in which the index set i_1, i_2, \dots, i_n is written should not have any effect on the measure of the (same) set. Both these consistency conditions are clearly necessary if there is to be any hope of extending the measures $\mu_{i_1 \dots i_n}$ to a single measure μ on \mathbf{R}^I . The fact that they are also sufficient was proved by Daniell in 1918 and rediscovered by Kolmogorov in 1933. We state it as

Theorem 6.10. *If I is any infinite index set, and for each finite set i_1, i_2, \dots, i_n of different indices in I there is a measure $\mu_{i_1 i_2 \dots i_n}$ defined on the Borel subsets of \mathbf{R}^n such that the family of all such measures satisfies the consistency conditions (I) and (II), there is a unique measure μ defined on \mathcal{B}^I in \mathbf{R}^I such that, for each $n \in \mathbf{Z}$, $B^n \in \mathcal{B}^n$,*

$$\mu\{f \in \mathbf{R}^I: (f(i_1), \dots, f(i_n)) \in B^n\} = \mu_{i_1 i_2 \dots i_n}(B^n).$$

Proof. Let \mathcal{S} denote the semi-ring of sets in \mathbf{R}^I of the form (6.6.1) for some finite value of n . Let \mathcal{R} denote the ring generated by \mathcal{S} consisting of finite unions of disjoint sets in \mathcal{S} . Now $\mu_{i_1 i_2 \dots i_n}$ defines the measure of the set

$$\{f \in \mathbf{R}^I: a_k < f(i_k) \leq b_k, \quad k = 1, \dots, n\} \quad (6.6.4)$$

and the consistency conditions (I) and (II) clearly ensure that the measure is uniquely defined and additive on \mathcal{S} (for the sets of any finite class of sets in \mathcal{S} can all be described by restrictions on the same finite set of coordinates, and therefore the measure can be given by a single measure of the family). It follows, by theorem 3.1, that there is a set function τ defined on the ring \mathcal{R} which is additive and coincides with the measure $\mu_{i_1 \dots i_n}$ on a set of the form (6.6.1).

Further \mathcal{B}^I is the σ -field generated by \mathbf{R} and we can obtain the required measure μ on \mathcal{B}^I by applying theorem 4.2 to the measure τ

—provided the conditions of that theorem are satisfied. It is immediate that \mathbf{R}^I is $\sigma\text{-}\mathcal{R}$, for $\mathbf{R}^I \in \mathcal{S}$, and $\tau(\mathbf{R}^I) = 1$; so that the only condition which requires proof is that τ is a measure on \mathcal{R} . The proof of this fact is an extension of the method used in §§3.4, 4.5.

If τ is not a measure on \mathcal{R} , we can find a decreasing sequence $\{E_n\}$ of sets in \mathcal{R} such that $\bigcap_{n=1}^{\infty} E_n = \emptyset$, but $\tau(E_n) \geq \delta > 0$ for all n . Now given any set C of the form (6.6.4), and $\epsilon > 0$, we can choose $\eta > 0$ such that

$$\tau(D) > \tau(C) - \epsilon$$

where

$$D = \{f \in \mathbf{R}^I: (f(i_1), f(i_2), \dots, f(i_n)) \in P_{\eta}\}$$

and

$$P_{\eta} = \{a_k + \eta < x_k \leq b_k, \quad k = 1, 2, \dots, n\},$$

since $\mu_{i_1 i_2 \dots i_n}$ is a measure. But now $\bar{P}_{\eta} \subset P_0$. This argument clearly extends to any non-empty subset in \mathcal{R} , and we can apply it by induction to the sequence $\{E_n\}$. Since in each of the sets E_n the value of f at only a finite set of indices is restricted, there is no loss of generality in assuming that in the sets E_1, E_2, \dots, E_n there is a restriction on f only at the first n of the indices in the sequence

$$i_1, i_2, \dots, i_n, \dots$$

(If this condition is not satisfied one need only add additional sets in \mathcal{R} to the sequence $\{E_n\}$ to obtain a new sequence of which the original is a subsequence.)

Thus we may assume that

$$E_n = \{f \in \mathbf{R}^I: (f(i_1), \dots, f(i_n)) \in Q_n\},$$

where $Q_n \in \mathcal{E}^n$ the class of elementary figures in \mathbf{R}^n . The condition that E_n be a decreasing sequence now means that $Q_{n+1} \subset Q_n \times \mathbf{R}$. We apply the above procedure to each of the sets E_n to give a sequence $\{D_n\}$ of sets

$$D_n = \{f \in \mathbf{R}^I: (f(i_1), \dots, f(i_n)) \in \bar{P}_n\}$$

such that $\bar{P}_n \subset Q_n$, $P_n \in \mathcal{E}^n$ and

$$\tau(D_n) > \tau(E_n) - \frac{\delta}{2^{n+1}}.$$

If we put

$$V_n = D_1 \cap D_2 \cap \dots \cap D_n$$

then $\tau(V_n) = \tau(E_n) - \tau(E_n - V_n) \geq \tau(E_n) - \sum_{i=1}^n \tau(E_i - D_i) > \frac{1}{2}\delta$

so that the sets $\{V_n\}$ form a monotone decreasing sequence of non-empty sets. In each V_n choose a point

$$f_n = \{f_n(i), \quad i \in I\}.$$

Now $(f_{n+p}(i_1), f_{n+p}(i_2), \dots, f_{n+p}(i_n)) \quad (p = 1, 2, \dots)$

defines a sequence of points in \mathbf{R}^n which is a subset of the bounded closed set

$$(\bar{P}_1 \times \mathbf{R}^{n-1}) \cap (\bar{P}_2 \times \mathbf{R}^{n-2}) \cap \dots \cap (\bar{P}_n) = F_n.$$

We can therefore find a subsequence of $\{f_{n+p}\}$ which, evaluated at the first n indices converges to a point of F_n . Since $\tau(V_n) > \frac{1}{2}\delta$, V_n is not empty and F_n is not empty since

$$V_n \subset \{f \in \mathbf{R}^I: (f(i_1), \dots, f(i_n)) \in F_n\}.$$

Further $F_n \times \mathbf{R} \subset F_{n+1}$ ($n = 1, 2, \dots$), and we can now employ a standard diagonalisation argument to obtain a point in $\bigcap_{n=1}^{\infty} E_n$.

Obtain successively, by induction, infinite increasing sequences of positive integers

$$\nu_1 \supset \nu_2 \supset \dots \supset \nu_k \supset \dots$$

such that $\{f_n\}$ restricted to the sequence ν_k gives a sequence whose values at i_1, i_2, \dots, i_k converge to a point in F_k . Form the sequence ν obtained by taking the k th integer in the sequence ν_k . Then, for each k , ν is a subsequence of ν_k except for a finite number of terms at the beginning so that $\{(f_n(i_1), f_n(i_2), \dots, f_n(i_k))\}, n \in \nu$ must converge to a point in $F_k \subset Q_k$. If we put $q_k = \lim_{n \in \nu} f_n(i_k)$ ($k = 1, 2, \dots$) the set

$$H = \{f \in \mathbf{R}^I: f(i_k) = q_k, k = 1, 2, \dots\}$$

is non-empty, and $H \subset V_n \subset E_n$ for all n . This contradicts $\bigcap_{n=1}^{\infty} E_n = \emptyset$.]

Remark. For a finite index set I , theorem 6.10 is still true, but lacks any content as the measure $\mu_{i_1} \dots i_n$ already is the required μ if

$$I = \{i_1, i_2, \dots, i_n\}.$$

Brownian motion

We can set up a mathematical model for Brownian motion by applying theorem 6.10 to a particular family of finite dimensional distributions. Use the index set $T = \{t \in \mathbf{R}, t \geq 0\}$ which can be thought of as time and, for

$$0 < t_1 < \dots < t_n,$$

define $\mu_{t_1} \dots t_n \{f \in \mathbf{R}^I: a_i < f(t_i) \leq b_i, i = 1, \dots, n\}$

$$\begin{aligned} &= \frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{a_n}^{b_n} \exp \left[-\frac{(\xi_n - \xi_{n-1})^2}{2(t_n - t_{n-1})} \right] d\xi_n \int_{a_{n-1}}^{b_{n-1}} \exp \left[-\frac{(\xi_{n-1} - \xi_{n-2})^2}{2(t_{n-1} - t_{n-2})} \right] d\xi_{n-1} \\ &\quad \dots \int_{a_2}^{b_2} \exp \left[-\frac{(\xi_2 - \xi_1)^2}{2(t_2 - t_1)} \right] d\xi_2 \int_{a_1}^{b_1} \exp \left(-\frac{\xi_1^2}{2t_1} \right) d\xi_1. \end{aligned}$$

The fact that this defines a consistent family of measures on \mathcal{S} which can be extended to all sets of the form (6.6.2) can be proved directly (it will follow from the discussion of the multinormal distribution in Chapter 14). Hence, we can apply theorem 6.10 to give a measure μ on \mathcal{B}^T the class of Borel sets in $\mathbf{R}^T = \Omega$. This is called *Wiener measure* in the space of functions $f: T \rightarrow \mathbf{R}$, and is an example of a stochastic process which will be discussed more fully in Chapter 15.

However, let us use the example of Wiener measure to illustrate the inadequacy of theorem 6.10. This follows from the fact that the σ -field \mathcal{B}^T is too small to contain interesting sets—for we have seen that it contains no set in which a non-countable set of time coordinates is restricted. Even if \mathcal{B}^T is completed with respect to μ to give a probability measure, the completed σ -field is still too small. For if $A \subset \mathbf{R}^T$ is a set in which f is restricted at a non-countable set, the only set of \mathcal{B}^T which is contained in A is the empty set. This means that A can only be measurable, if it has measure zero. But the same argument applies to the complement of A so that if both A and its complement involve restrictions on $f: T \rightarrow \mathbf{R}$ at a non-countable set of indices, then the outer measure of A must be 1, and the inner measure must be 0. In particular the set

$$\{f \in \mathbf{R}^T: a < f(t) \leq b \text{ for all } t \in [t_1, t_2]\} \quad (6.6.4)$$

is not measurable, and if C is the set of functions $f: T \rightarrow \mathbf{R}$ which are continuous for all $t \in T$, C also has outer measure 1 (and inner measure 0). Various methods can be used to extend μ from \mathcal{B}^T to a larger σ -field which includes C and (6.6.4) and other sets of interest. These have been studied in detail and the interested reader is advised to look in J. L. Doob, *Stochastic Processes* (Wiley, 1953).

6.7 Applications

In the second part of this book random variables will be defined as \mathcal{F} -measurable functions $f: \Omega \rightarrow \mathbf{R}^*$ where $(\Omega, \mathcal{F}, \mu)$ is a probability space. Although it is usual to work with general carrier spaces Ω , there is a sense in which the real line \mathbf{R} has a structure sufficiently complicated to reproduce all the probability properties of the function f . In fact, in many treatises on probability theory, the carrier space Ω is barely mentioned. This attitude is partially justified by the following considerations.

Suppose $(\Omega, \mathcal{F}, \mu)$ is any finite measure space and $f: \Omega \rightarrow \mathbf{R}$ is \mathcal{F} -measurable. For all real x , define

$$F(x) = \mu\{y: f(y) \leq x\}. \quad (6.7.1)$$

Then $F(x) \rightarrow 0$ as $x \rightarrow -\infty$, $F(x) \rightarrow \mu(\Omega)$ as $x \rightarrow +\infty$, and $F: \mathbf{R} \rightarrow \mathbf{R}$ is continuous on the right. Thus we can define a Stieltjes measure μ_F using this particular F .

Theorem 6.11. *Suppose $(\Omega, \mathcal{F}, \mu)$ is a finite measure space and $f: \Omega \rightarrow \mathbf{R}$ is \mathcal{F} -measurable, μ_F is the Lebesgue–Stieltjes measure in \mathbf{R} given by (6.7.1) and $g: \mathbf{R} \rightarrow \mathbf{R}$ is Borel measurable, then $g(f)$ is \mathcal{F} -measurable and $\mu\{x: g(f)(x) \in B\}$ is determined by μ_F for every Borel set B . Further*

$$\int g(f) d\mu = \int g(x) dF(x)$$

in the sense that, if either side exists so does the other, and the two are equal.

Proof. $\{x: g(f)(x) \in B\} = \{x: f(x) \in g^{-1}(B)\}$ and $g^{-1}(B) = C$ is a Borel set so that $\{x: f(x) \in C\}$ is in \mathcal{F} , and $\mu\{x: f(x) \in C\}$ is uniquely determined by $\mu\{x: a < f(x) \leq b\} = F(b) - F(a)$ for all real a, b since \mathcal{P} , the class of half-open intervals generates the σ -field \mathcal{B} of Borel sets in \mathbf{R} , and $F(b) - F(a) = \mu_F(a, b]$. Thus, for all B in \mathcal{B} ,

$$\mu\{x: g(f)(x) \in B\} = \mu_F(g^{-1}(B)).$$

Now suppose g is an indicator function of a Borel set B . Then

$$\begin{aligned} \int g(f) d\mu &= \mu\{x: f(x) \in B\} \\ &= \mu_F(B) = \int g d\mu_F. \end{aligned}$$

By linearity our result follows for non-negative simple functions and the monotone convergence theorem then gives it for non-negative Borel measurable g and then for all integrable g . \blacksquare

Corollary. *In the notation of the theorem*

$$\int f d\mu = \int x dF(x).$$

Remark. There is an n -dimensional form of theorem 6.11 and corollary which links the behaviour of n measurable functions with a Lebesgue–Stieltjes distribution in \mathbf{R}^n —see Chapter 14.

Marginal distributions

Not all measures in product spaces are product measures. Suppose X, Y are spaces, then the projection $X \times Y \rightarrow X$ given by $p(x, y) = x$ defines a mapping. This will be a measurable transformation on $(X \times Y, \mathcal{F})$ into (X, \mathcal{S}) provided $E \times Y \in \mathcal{F}$ for every $E \in \mathcal{S}$. In this case, if μ is a finite measure on \mathcal{F} , μp^{-1} defines a measure on \mathcal{S} . In

general it may not be a very interesting measure as there may be no sets of finite positive measure. However, if $(X \times Y, \mathcal{F}, \mu)$ is a finite measure space, then the measure μp^{-1} on \mathcal{S} is called the marginal measure on X . The marginal measure on Y is similarly defined using a projection on Y . If μ is a probability measure these marginal measures are called marginal (probability) distributions.

If $F(x, y)$ is a distribution function in \mathbf{R}^2 (see §4.5) then

$$\lim_{y \rightarrow +\infty} F(x, y) \quad \text{and} \quad \lim_{x \rightarrow +\infty} F(x, y)$$

will again define 1-dimensional distribution functions, and it is immediate from theorem 4.8 that the corresponding Lebesgue–Stieltjes measures will be the marginal distributions of μ_F . If

$$F(x, y) = F_1(x) \cdot F_2(y)$$

is the product of two 1-dimensional distributions, then μ_F will be the completion of the product measure $\mu_{F_1} \times \mu_{F_2}$ and F_1, F_2 will be the marginal distributions for F . Conversely, if μ_F is a product measure, then it must be the product of its marginal distributions so that $F(x, y) = F_1(x) F_2(y)$ is a necessary as well as a sufficient condition for μ_F to be a product of two probability measures.

Thick subsets

For any finite measure space $(\Omega, \mathcal{F}, \mu)$ we can generate the outer measure

$$\mu^*(E) = \inf \sum_{i=1}^{\infty} \mu(E_i) \quad (6.7.2)$$

the infimum being taken over all sequences of sets $\{E_i\}$ in \mathcal{F} with $E \subset \bigcup_{i=1}^{\infty} E_i$. (Since μ is a measure on the σ -field \mathcal{F} , (6.7.2) is the same as $\mu^*(E) = \inf \mu(F)$ for $F \supset E, F \in \mathcal{F}$.) A subset E_0 of Ω is said to be *thick* in Ω if $\mu^*(E_0) = \mu(\Omega)$. Thus a subset E_0 is thick if and only if $(\Omega - E_0)$ contains no set in \mathcal{F} of positive μ -measure. There is a sense in which the measure space can be projected onto any thick subset.

Theorem 6.12. *If E_0 is a thick subset of the finite measure space $(\Omega, \mathcal{F}, \mu)$, $\mathcal{F}_0 = \mathcal{F} \cap E_0$, and $\mu_0(E \cap E_0) = \mu(E)$ for any $E \in \mathcal{F}$, then $(E_0, \mathcal{F}_0, \mu_0)$ is a measure space.*

Proof. We first see that μ_0 is defined uniquely on \mathcal{F}_0 . If $A_1, A_2 \in \mathcal{F}$ are such that $A_1 \cap E_0 = A_2 \cap E_0$, then we must have

$$(A_1 \triangle A_2) \cap E_0 = \varnothing,$$

so that $\mu(A_1 \triangle A_2) = 0$ and $\mu(A_1) = \mu(A_2)$.

Now suppose $\{B_n\}$ is a disjoint sequence of sets in \mathcal{F}_0 so that there is a sequence of sets $\{C_n\}$ in \mathcal{F} with

$$B_n = C_n \cap E_0 \quad (n = 1, 2, \dots).$$

Put
$$D_n = C_n - \bigcup_{i=1}^{n-1} C_i \quad (n = 1, 2, \dots).$$

Then
$$D_n \cap E_0 = C_n \cap E_0,$$

so that $\mu(D_n \triangle C_n) = 0$. It follows that

$$\sum_{n=1}^{\infty} \mu_0(B_n) = \sum_{n=1}^{\infty} \mu(C_n) = \sum_{n=1}^{\infty} \mu(D_n) = \mu\left(\bigcup_{n=1}^{\infty} D_n\right) = \mu\left(\bigcup_{n=1}^{\infty} C_n\right) = \mu_0\left(\bigcup_{n=1}^{\infty} B_n\right)$$

so that μ_0 is a measure. **]**

Remark. This theorem shows that in a probability space (Ω, \mathcal{F}, P) , the σ -field \mathcal{F} can be extended to include any set E_0 not in it whose outer measure is 1. The effect of this extension is to discard all the points of Ω which are not in E_0 . The device turns out to be useful in the theory of stochastic processes where, by a careful choice of E_0 , one can obtain a probability on a useful class of subsets. In particular, for Wiener measure in \mathbf{R}^T described in §6.6, it can be shown that the set C of continuous functions is thick and that the extension given by putting $E_0 = C$ is a useful one—see Chapter 15.

Exercises 6.7

1. Formulate and prove a theorem of the form of theorem 6.11 for n \mathcal{F} -measurable functions $f_i: \Omega \rightarrow \mathbf{R}$ ($i = 1, 2, \dots, n$).

2. Find the 2-dimensional distribution function $F(x, y)$ which generates the measure μ_F such that $\mu_F(R)$ is $1/\sqrt{2}$ (length of diagonal D in R) for any rectangle R , where D is the segment joining $(0, 0)$ to $(1, 1)$. Calculate the marginal distributions of μ_F , and show that μ_F is not a product measure.

3. If $(\Omega, \mathcal{F}, \mu)$ is a complete σ -finite measure space, and the outer measure μ^* is defined by (6.7.2) show that a set E is μ^* -measurable if and only if it is in \mathcal{F} .

4. Suppose $(\Omega, \mathcal{F}, \mu)$ is a finite measure space and E_0 is a subset of Ω such that, for $A_1, A_2 \in \mathcal{F}$,

$$A_1 \cap E_0 = A_2 \cap E_0 \Rightarrow \mu(A_1) = \mu(A_2).$$

Prove that E_0 is thick in Ω .

THE SPACE OF MEASURABLE FUNCTIONS

Throughout this chapter we will assume (unless stated otherwise) that $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space, and that the σ -field \mathcal{F} is complete with respect to μ . This implies that if $f: \Omega \rightarrow \mathbf{R}^*$, $g: \Omega \rightarrow \mathbf{R}^*$ are functions such that f is \mathcal{F} -measurable and $f = g$ a.e., then g is also \mathcal{F} -measurable. Thus, if M is the class of functions $f: \Omega \rightarrow \mathbf{R}^*$ which are \mathcal{F} -measurable, we say that f_1, f_2 in M are equivalent if $f_1 = f_2$ a.e. This clearly defines an equivalence relation in M and we can form the space \mathcal{M} of equivalence classes with respect to this relation. When we think of a function f of M as an element of \mathcal{M} we are really thinking of f as a representative of the class of \mathcal{F} -measurable functions which are equal to f a.e. As is usual we will use the same notation f for an element of M and \mathcal{M} . We can think of M or \mathcal{M} as an abstract space, and the definition of convergence if given in terms of a metric will then impose a topological structure on the space. We will consider several such notions of convergence of which some, but not all, can be expressed in terms of a metric in \mathcal{M} . We will obtain the relationships between different notions of convergence, and in each case prove that the space is complete in the sense that for any Cauchy sequence there is a limit function to which the sequence converges. The main strategy used to prove completeness will be to find a suitable subsequence of the given sequence which clearly converges to a limit f and then show that f is a limit of the whole sequence. This extends the method used in § 2.2 to show that \mathbf{R} is complete.

7.1 Point-wise convergence

Given a sequence $\{f_n\}$ of functions where $f_n: E \rightarrow \mathbf{R}^*$ and a function $f: E \rightarrow \mathbf{R}^*$ ($E \subset \Omega$), we say that f_n converges to f point-wise on E if, for each x in E , $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$. This notion has a meaning if we restrict consideration to \mathcal{M} . If E is such that $\mu(\Omega - E) = 0$, and $f_n \rightarrow f$ point-wise on E , then we say that $f_n \rightarrow f$ a.e. For if $f_n \rightarrow f$ a.e., $f_n = g_n$ a.e. for each n , and $g_n \rightarrow g$ a.e., then

$$\begin{aligned} \{x: f(x) \neq g(x)\} &\subset \{x: f_n(x) \not\rightarrow f(x)\} \\ &\cup \{x: g_n(x) \not\rightarrow g(x)\} \cup \bigcup_{n=1}^{\infty} \{x: f_n(x) \neq g_n(x)\}, \end{aligned}$$

and each of these sets has zero measure so $f(x) = g(x)$ a.e. which means that $f = g$ in \mathcal{M} . $\{f_n\}$ is a Cauchy sequence (point-wise) on E if, given $x \in E$, $\epsilon > 0$, there is an integer N such that

$$|f_n(x) - f_m(x)| < \epsilon \quad \text{for } n, m > N. \quad (7.1.1)$$

(This has meaning only if $f_n: E \rightarrow \mathbf{R}$ is finite valued.) Because \mathbf{R} is complete it is clear that if $\{f_n\}$ is a Cauchy sequence on E , there must be an $f: E \rightarrow \mathbf{R}$ such that $f_n \rightarrow f$ point-wise on E .

Uniform convergence

If the sequence $\{f_n\}$ and the function f are finite valued functions on E to \mathbf{R} , we say that f_n converges uniformly to f on E if for each $\epsilon > 0$, there is an integer N such that

$$x \in E, \quad n \geq N \Rightarrow |f_n(x) - f(x)| < \epsilon.$$

Similarly, we say that the sequence is a Cauchy sequence uniformly on E if given $\epsilon > 0$, a single integer N can be chosen so that (7.1.1) is satisfied for all $x \in E$. Since a Cauchy sequence uniformly on E is certainly a Cauchy sequence on E and the existence of $\lim_{n \rightarrow \infty} f_m(x) = f(x)$

follows for each x , we can let $m \rightarrow \infty$ in (7.1.1) to deduce that a Cauchy sequence uniformly on E must have a limit function $f: E \rightarrow \mathbf{R}$ such that $f_n \rightarrow f$ uniformly on E .

If $\mu(\Omega - E) = 0$ and $f_n \rightarrow f$ uniformly on E , then we say that $f_n \rightarrow f$ uniformly a.e. All these notions have a meaning for functions which need not be measurable. However, the notion of convergence uniformly a.e. can be expressed in terms of a metric on the restricted class of measurable functions.

Essentially bounded functions

An \mathcal{F} -measurable function $f: \Omega \rightarrow \mathbf{R}^*$ is said to be essentially bounded if $\mu\{x: |f(x)| > a\} = 0$ for some real number a . In this case we define the essential supremum of f by

$$\text{ess sup } |f| = \inf \{a: \mu\{x: |f(x)| > a\} = 0\}.$$

Notice that, if $\text{ess sup } |f| = C$, then

$$E = \{x: |f(x)| > C\} = \bigcup_{k=1}^{\infty} \left\{x: |f(x)| > C + \frac{1}{k}\right\}$$

so that $\mu(E) = 0$ and $|f(x)| \leq C$ outside E . Thus $|f(x)| \leq C$ a.e., and if we define

$$f^*(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq C, \\ 0 & \text{if } |f(x)| > C, \end{cases}$$

then $|f^*(x)| \leq C$ for all x and $f^* = f$ a.e. Further $\{x: |f^*(x)| > C - \epsilon\}$ has positive measure for all $\epsilon > 0$, so that it is non-empty and we must have $\sup |f^*| = C$. It is clear that, if $f = g$ a.e., then $\text{ess sup } f = \text{ess sup } g$, so that we can think of ess sup as a functional on the subset \mathcal{L}_∞ of the essentially bounded functions of \mathcal{M} . If we define $(\alpha f + \beta g)$ by

$$\begin{aligned} (\alpha f + \beta g)(x) &= \alpha f(x) + \beta g(x) & \text{when } f(x), g(x) \in \mathbf{R}, \\ &= 0 & \text{otherwise;} \end{aligned}$$

it is clear that $(\alpha f + \beta g) \in \mathcal{L}_\infty$ if $f, g \in \mathcal{L}_\infty$ for any $\alpha, \beta \in \mathbf{R}$ so that \mathcal{L}_∞ is a linear subspace of \mathcal{M} (over the reals) Further

$$\rho_\infty(f, g) = \text{ess sup } |f - g|$$

defines a metric in \mathcal{L}_∞ for

- (i) $\rho_\infty(f, g) = \rho_\infty(g, f)$;
- (ii) $\rho_\infty(f, g) = 0$ if only if $f = g$ a.e.;
- (iii) $\text{ess sup } |f + g| \leq \text{ess sup } |f| + \text{ess sup } |g|$ so that

$$\rho_\infty(f, g) \leq \rho_\infty(f, h) + \rho_\infty(h, g).$$

Now it is clear that, if $\{f_n\}$ and f are functions in \mathcal{L}_∞ such that $f_n \rightarrow f$ uniformly a.e., then $\rho_\infty(f_n, f) \rightarrow 0$ as $n \rightarrow \infty$. Conversely suppose $\rho_\infty(f_n, f) \rightarrow 0$, and let E_n be a set of \mathcal{F} with $\mu(E_n) = 0$ and

$$\text{ess sup } |f_n - f| = \sup_{x \in \Omega - E_n} |f_n(x) - f(x)|.$$

Put $E = \bigcup E_n$, then for $x \in \Omega - E$

$$|f_n(x) - f(x)| \leq \sup_{x \in \Omega - E_n} |f_n(x) - f(x)| = \text{ess sup } |f_n - f|$$

so that $f_n \rightarrow f$ uniformly on $\Omega - E$ and $\mu(E) = 0$. A similar, but slightly more complicated argument shows that, in \mathcal{L}_∞ , a Cauchy sequence uniformly a.e. is the same as a Cauchy sequence in ρ_∞ norm.

Almost uniform convergence

Given functions $f_n: E \rightarrow \mathbf{R}^*$ ($n = 1, 2, \dots$) and $f: E \rightarrow \mathbf{R}^*$ each of which is finite a.e. on E we say that f_n converges almost uniformly to f on E if, for each $\epsilon > 0$, there is a set $F_\epsilon \subset E$, $F_\epsilon \in \mathcal{F}$, $\mu(F_\epsilon) < \epsilon$ such that $f_n \rightarrow f$ uniformly on $(E - F_\epsilon)$. The example $E = [0, 1] \subset \mathbf{R}$, $f_n(x) = x^n$ μ Lebesgue measure shows that it is possible for a sequence to converge almost uniformly on E while it does not converge uniformly a.e. on E . However, it is immediate from the definitions that convergence uniformly a.e. implies almost uniform convergence. What is more surprising is that, under suitable conditions, convergence a.e. implies almost uniform convergence.

Theorem 7.1. (Egoroff). Suppose $E \in \mathcal{F}$, $\mu(E) < \infty$, and $\{f_n\}$ is a sequence of measurable functions on $E \rightarrow \mathbf{R}^*$ which are finite a.e. and converge a.e. to a function $f: E \rightarrow \mathbf{R}^*$ which is also finite a.e. Then $f_n \rightarrow f$ almost uniformly in E .

Proof. By omitting a subset of E of zero measure, we may assume that all the functions f_n and f are finite and that

$$f_n(x) \rightarrow f(x) \quad \text{for all } x \in E.$$

For positive integers, m, n put

$$A_n^m = \bigcap_{i=n}^{\infty} \left\{ x: |f_i(x) - f(x)| < \frac{1}{m} \right\}.$$

Then, for fixed m , $A_1^m, A_2^m, \dots, A_n^m, \dots$ is an increasing sequence of measurable sets converging to E . Since $\mu(E)$ is finite, by theorem 3.2 there is a positive integer $N_m = N_m(m)$ such that

$$\mu(E - A_i^m) < \epsilon/2^m \quad \text{for } i \geq N_m.$$

If we put
$$F_\epsilon = \bigcup_{m=1}^{\infty} (E - A_{N_m}^m),$$

then $\mu(F_\epsilon) < \epsilon$. Further given $\delta > 0$ we can choose m so that $1/m < \delta$ and then

$$|f_i(x) - f(x)| < \delta \quad \text{for all } i \geq N_m, \quad x \in (E - F_\epsilon),$$

so that $f_n \rightarrow f$ uniformly on $(E - F_\epsilon)$. \blacksquare

Remark. The converse to theorem 7.1 is true and almost trivial. For if $\{f_n\}, f$ are finite a.e. on E , measurable, and $f_n \rightarrow f$ almost uniformly, this means we can find sets F_n with $\mu(F_n) < 1/n$ such that $f_n \rightarrow f$ uniformly on $(E - F_n)$ and so $f_n \rightarrow f$ point-wise on $(E - F_n)$. Put

$$F = \bigcap_{n=1}^{\infty} F_n, \quad \text{then } \mu(F) = 0$$

and $f_n \rightarrow f$ point-wise on $(E - F)$ so that $f_n \rightarrow f$ a.e. on E .

Exercises 7.1

1. Let X be the space of positive integers, \mathcal{F} the class of all subsets of X , and $\mu(E)$ the number of integers in $E \subset X$. If $f_n(x)$ is the indicator function of $\{1, 2, \dots, n\}$, then $f_n(x) \rightarrow 1$ for all x . However, f_n does not converge almost uniformly to 1, showing that theorem 7.1 is false without $\mu(E) < \infty$.

2. Suppose the conditions of theorem 7.1 are satisfied except that $\mu(E) = \infty$, show that given $P > 0$, there is a subset $F_P \subset E$ with $\mu(F_P) > P$ such that $f_n \rightarrow f$ uniformly on F_P but that there need not be a subset F with $\mu(F) = +\infty$ with $f_n \rightarrow f$ uniformly on F .

3. Suppose $E \in \mathcal{F}$, $\mu(E) < \infty$, $f_n: E \rightarrow \mathbf{R}^*$ ($n = 1, 2, \dots$) is a Cauchy sequence a.e. of measurable functions each finite a.e. Prove there is a finite c and a measurable $F \subset E$ with $\mu(F) > 0$ such that, for every integer n , all $x \in F$, $|f_n(x)| \leq c$.

4. Suppose $E \in \mathcal{F}$, E has σ -finite measure, f_n ($n = 1, 2, \dots$) and f are finite a.e. on E and $f_n \rightarrow f$ a.e. on E . Show that there exists a sequence $\{E_i\}$ of sets in \mathcal{F} such that

$$\mu\left(E - \bigcup_{i=1}^{\infty} E_i\right) = 0 \quad \text{and} \quad f_n \rightarrow f$$

uniformly on each E_i . By considering the measure of example 2, §3.1, and a suitable sequence of functions show that the condition that E has σ -finite measure is essential.

5. In §4.4 we produced a sequence of sets $\{Q_n\}$ each of which was not Lebesgue measurable. If we put $f_n(x) =$ indicator function of

$$[0, 1) - \bigcup_{i=1}^n Q_i, \quad \text{then} \quad f_n(x) \rightarrow 0 \quad \text{for all } x \text{ in } [0, 1].$$

Show that f_n does not converge almost uniformly so that theorem 7.1 fails if the functions are not measurable.

6. Suppose $f_h: E \rightarrow \mathbf{R}$, $h > 0$ is a continuous family of measurable functions, each finite valued, $\mu(E) < \infty$ and for each $x \in E$, $f_h(x) \rightarrow f(x)$ as $h \rightarrow 0$ where f is finite valued. Then if a continuous parameter version of Egoroff's theorem were valid we would have given $\epsilon > 0$, there exists $F_\epsilon \subset \mathcal{F}$, $F_\epsilon \subset E$, $\mu(F_\epsilon) < \epsilon$ such that $f_h(x) \rightarrow f(x)$ as $h \rightarrow 0$ uniformly on $(E - F_\epsilon)$. The following example shows that this extension is false. In Chapter 4, we saw that there is a non-measurable set $E \subset [0, 1)$ such that every point $x \in [0, 1)$ has a unique representation $x = y + q \pmod{1}$, $y \in E$, q rational.

Prove that, if M is a measurable subset of $[0, 1]$ such that $M \cap E(r)$ is non-void for finitely many rationals r , then $|M| = 0$.

Arrange the rationals Q as a sequence $\{r_n\}$. For $x \in [0, 1)$ let $n(x)$ be the integer such that $x = y + r_{n(x)}$, $y \in E$. If $x/n(x) = \cdot\alpha_1\alpha_2\dots$ (decimal representation not ending in 9 recurring), put $\phi(x) = \cdot\beta_1\beta_2\dots$ where $\beta_{2k} = \alpha_k$ ($k = 1, 2, \dots$); and $\beta_{2k-1} = 1$ for $k = n(x)$, 0 otherwise. Put $f_h(x) = 1$, for $x = \phi(h)$, $f_h(x) = 0$ otherwise. Prove $f_h(x) \rightarrow 0$ as $h \rightarrow 0$ for each x . Show that if M any measurable set, $|M| > 0$, then $f_h(x) \not\rightarrow 0$ uniformly on M .

7. Suppose $\{f_n\}$ is a sequence of functions in \mathcal{M} , $f_n \rightarrow f$ a.e. and there is an integrable function g such that $|f_n| \leq g$ a.e. for all n . Show that $f_n \rightarrow f$ almost uniformly.

8. Define what is meant by saying that a sequence $\{f_n\}$ of a.e. finite valued functions is a Cauchy sequence almost uniformly, and show that this implies the existence of a limit function f such that $f_n \rightarrow f$ almost uniformly.

7.2 Convergence in measure

We now consider a different kind of 'nearness' in \mathcal{M} in which the measure of the set where two functions differ by more than a fixed positive number is relevant. This time we make the definitions relative to the whole space Ω . Obvious changes give the corresponding concepts relative to a set E in \mathcal{F} . Given \mathcal{F} -measurable functions $f: \Omega \rightarrow \mathbf{R}^*$, $f_n: \Omega \rightarrow \mathbf{R}^*$ ($n = 1, 2, \dots$) we say that f_n converges in measure (μ) to f if, for each $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mu\{x: |f_n(x) - f(x)| \geq \epsilon\} = 0.$$

Note that the definition only makes sense for functions in \mathcal{M} which are finite a.e. We first see that the limit in measure is unique in \mathcal{M} . For suppose $f_n \rightarrow f$ in measure, $f_n \rightarrow g$ in measure; then if $\delta > 0$,

$$\{x: |f(x) - g(x)| > \delta\} \subset \{x: |f_n(x) - f(x)| > \tfrac{1}{2}\delta\} \cup \{x: |f_n(x) - g(x)| > \tfrac{1}{2}\delta\}$$

and both sets on the right can be made of arbitrarily small measure by choosing n large. This means that

$$\mu\{x: |f(x) - g(x)| > \delta\} = 0 \quad \text{for each } \delta > 0,$$

and it follows that $f = g$ a.e. (by taking a sequence δ_n decreasing to zero).

We say that the sequence $\{f_n\}$ of functions in \mathcal{M} is a Cauchy sequence in measure if, given $\epsilon > 0$, $\delta > 0$ there is an integer N such that

$$n \geq N, \quad m \geq N \Rightarrow \mu\{x: |f_n(x) - f_m(x)| > \epsilon\} < \delta.$$

The argument used to prove uniqueness of the limit also shows that

$$f_n \rightarrow f \text{ in measure} \Rightarrow \{f_n\} \text{ is a Cauchy sequence in measure.}$$

The converse is included in the following theorem.

Theorem 7.2. *Suppose f and f_n ($n = 1, 2, \dots$) are functions in \mathcal{M} which are finite a.e. Then*

- (i) $f_n \rightarrow f$ almost uniformly $\Rightarrow f_n \rightarrow f$ in measure;
- (ii) $\{f_n\}$ is a Cauchy sequence almost uniformly $\Rightarrow \{f_n\}$ is a Cauchy sequence in measure;
- (iii) $\{f_n\}$ is a Cauchy sequence in measure \Rightarrow there is a subsequence $\{n_k\}$ such that $\{f_{n_k}\}$ is a Cauchy sequence almost uniformly;
- (iv) $\{f_n\}$ is a Cauchy sequence in measure \Rightarrow there is a function $g \in \mathcal{M}$ such that $f_n \rightarrow g$ in measure.

Proof. (i) If $f_n \rightarrow f$ almost uniformly, for each $\epsilon > 0$, $\delta > 0$, we can find a set $E_\delta \in \mathcal{F}$ such that $\mu(E_\delta) < \delta$ and $f_n \rightarrow f$ uniformly on $(\Omega - E_\delta)$. Hence there is an N such that

$$|f_n(x) - f(x)| < \epsilon \quad \text{for } n \geq N, \quad x \in (\Omega - E_\delta)$$

and then

$$\mu\{x: |f_n(x) - f(x)| \geq \epsilon\} \leq \mu(E_\delta) < \delta \quad \text{for } n \geq N.$$

(ii) An argument similar to that in (i) will work.

(iii) Now suppose that f_n is a Cauchy sequence in measure. For each positive integer k , choose an integer m_k such that $m_k > m_{k-1}$ and

$$n \geq m_k, \quad m \geq m_k \Rightarrow \mu\{x: |f_n(x) - f_m(x)| \geq 2^{-k}\} \leq 2^{-k}.$$

Put

$$E_k = \{x: |f_{m_k}(x) - f_{m_{k+1}}(x)| \geq 2^{-k},$$

$$F_k = \bigcup_{i=k}^{\infty} E_i.$$

Then

$$\mu(F_k) \leq \sum_{i=k}^{\infty} \mu(E_i) \leq 2^{1-k}.$$

Given $\epsilon > 0$ we can choose k so that $\epsilon > 2^{1-k}$; then $\mu(F_k) < \epsilon$ and for all $x \in (\Omega - F_k)$ we have

$$|f_{m_i}(x) - f_{m_{i+1}}(x)| < 2^{-i} \quad \text{for all } i \geq k.$$

Thus

$$j \geq i \geq k \Rightarrow |f_{m_i}(x) - f_{m_j}(x)| \leq \sum_{s=i}^{j-1} |f_{m_s}(x) - f_{m_{s+1}}(x)| < 2^{1-i}$$

so that the sequence f_{m_i} converges uniformly on $(\Omega - F_k)$; that is, since $\mu(F_k) \rightarrow 0$ as $k \rightarrow \infty$, it is a Cauchy sequence almost uniformly.

(iv) By (iii) we can obtain a subsequence f_{m_k} of the given sequence which is a Cauchy sequence almost uniformly. This means we can find a function $g \in \mathcal{M}$ such that $f_{m_k} \rightarrow g$ almost uniformly as $k \rightarrow \infty$. Now, for $\epsilon > 0$,

$$\begin{aligned} \{x: |f_n(x) - g(x)| \geq \epsilon\} &\subset \{x: |f_n(x) - f_{m_k}(x)| \geq \tfrac{1}{2}\epsilon\} \\ &\cup \{x: |f_{m_k}(x) - g(x)| \geq \tfrac{1}{2}\epsilon\}. \end{aligned}$$

Given $\delta > 0$ we can find a set $E_\delta \in \mathcal{F}$ and integers k_0, N such that $\mu(E_\delta) < \tfrac{1}{2}\delta$,

$$|f_{m_k}(x) - g(x)| < \tfrac{1}{2}\epsilon \quad \text{for } k \geq k_0, \quad x \in \Omega - E_\delta,$$

and $\mu\{x: |f_n(x) - f_{m_k}(x)| \geq \tfrac{1}{2}\epsilon\} < \tfrac{1}{2}\delta \quad \text{for } n \geq N, \quad m_k \geq N.$

It follows that

$$n \geq \max\{N, m_{k_0}\} \Rightarrow \mu\{x: |f_n(x) - g(x)| \geq \epsilon\} < \delta. \blacksquare$$

It is not difficult to see that convergence in measure does not necessarily imply convergence point-wise at any point, and so it certainly cannot imply almost uniform convergence of the whole sequence. For let

$$E_{r,k} = \left[\frac{r-1}{2^k}, \frac{r}{2^k} \right] \quad (r = 1, 2, \dots, 2^k; \quad k = 1, 2, \dots),$$

and arrange these intervals as a single sequence of sets $\{F_n\}$ by taking first those for which $k = 1$, then those with $k = 2$, etc. If μ denotes Lebesgue measure on $[0, 1]$, and $f_n(x)$ is the indicator function of F_n , then, for $0 < \epsilon < 1$,

$$\{x: |f_n(x)| \geq \epsilon\} = F_n$$

so that, for any $\epsilon > 0$, $\mu\{x: |f_n(x)| \geq \epsilon\} \leq \mu(F_n) \rightarrow 0$. This means that $f_n \rightarrow 0$ in measure in $[0, 1]$. However, at no point x in $[0, 1]$ does $f_n(x) \rightarrow 0$; in fact, since every x is in infinitely many of the sets F_n and infinitely many of the sets $(\Omega - F_n)$ we have

$$\liminf f_n(x) = 0, \quad \limsup f_n(x) = 1 \quad \text{for all } x \in [0, 1].$$

Exercises 7.2

1. Suppose $\{f_n\}$ is a Cauchy sequence in measure, and f_n, f_m are two subsequences which converge to f, g , respectively. Prove that $f = g$ a.e.

2. Show that if $\{f_n\}$ is a Cauchy sequence in measure then every subsequence of $\{f_n\}$ is also a Cauchy sequence in measure.

3. If Ω is the set of positive integers and μ is the counting measure on the class \mathcal{F} of all subsets, show that convergence in measure is equivalent to uniform convergence.

4. If $\mu(\Omega) = \infty$ can we say that convergence a.e. implies convergence in measure?

5. Suppose $\{A_n\}$ is a sequence of sets in \mathcal{F} , χ_n is the indicator function of A_n , and $d(A, B) = \mu(A \Delta B)$ for $A, B \in \mathcal{F}$. Show that $\{\chi_n\}$ is a Cauchy sequence in measure if and only if $d(A_n, A_m) \rightarrow 0$ as $n, m \rightarrow \infty$.

6. Suppose $\{f_n\}$ is a sequence of functions of M which are finite a.e. and $f_n \rightarrow f$ a.e. with f finite a.e. Show that, if (i) $\mu(\Omega) < \infty$, or (ii) $|f_n| \leq g_0$ for all n where g_0 is integrable; then $f_n \rightarrow f$ in measure.

7. Suppose $(\Omega, \mathcal{F}, \mu)$ is a finite measure space and $\{f_n\}, \{g_n\}$ are finite valued \mathcal{F} -measurable functions which converge in measure to f, g respectively. Show

(i) $|f_n|$ converges in measure to $|f|$;

(ii) for all real α, β the sequence $\{\alpha f_n + \beta g_n\}$ converges in measure to $(\alpha f + \beta g)$;

- (iii) if $f = 0$ a.e., then f_n^2 converges in measure to f^2 ;
- (iv) the sequence $\{f_n g\}$ converges in measure to fg ;
- (v) the sequence $\{f_n^2\}$ converges in measure to f^2 ;
- (vi) the sequence $\{f_n g_n\}$ converges in measure to fg ;
- (vii) if $f_n \neq 0$ a.e. all n , $f \neq 0$ a.e., the sequence $\{1/f_n\}$ converges in measure to $1/f$.

Is the condition $\mu(\Omega) < \infty$ essential for all these results?

7.3 Convergence in p th mean

All the definitions of the present section can be made relative to an arbitrary E in \mathcal{F} . Since we could restrict μ to the σ -field $\mathcal{F} \cap E$ of subsets of E , there is no loss in generality in making our definitions in terms of Ω , the whole space. In Chapter 5 we saw that $f \in \mathcal{M}$ is μ -integrable (over Ω) if and only if $|f|$ is μ -integrable. Further we saw that the subset of L_1 of M consisting of μ -integrable functions is a linear space (here we define $(\alpha f + \beta g)(x)$ arbitrarily on the set of zero measure where it is not defined because it involves $+\infty + (-\infty)$). Further for $f, g \in L_1$,

$$\rho_1(f, g) = \int |f - g| d\mu$$

is finite. By theorem 5.6, $\rho(f, g) = 0$ if and only if $f = g$ a.e., so that if we take equivalence classes of functions equal a.e. to form the linear space $\mathcal{L}_1 \subset \mathcal{M}$ we see that

$$\rho_1(f, g) = \rho_1(g, f) \quad \text{for all } f, g \in \mathcal{L}_1,$$

$$\rho_1(f, g) = 0 \quad \text{if and only if } f = g \text{ in } \mathcal{L}_1.$$

The triangle inequality

$$\rho_1(f, h) \leq \rho_1(f, g) + \rho_1(g, h) \quad \text{for } f, g, h \in \mathcal{L}_1$$

also follows by integrating

$$|f(x) - h(x)| \leq |f(x) - g(x)| + |g(x) - h(x)|,$$

so that ρ_1 defines a metric in the space \mathcal{L}_1 .

Convergence in mean

A sequence $\{f_n\}$ of functions in L_1 (or in \mathcal{L}_1) is said to converge in mean to a function f in L_1 if $\rho_1(f_n, f) \rightarrow 0$ as $n \rightarrow \infty$. A sequence $\{f_n\}$ of functions in L_1 is a Cauchy sequence in mean if $\rho_1(f_n, f_m) \rightarrow 0$ as $n, m \rightarrow \infty$.

Convergence in mean is the special case $p = 1$ of convergence in p th mean. Since most of the proofs are the same for $p = 1$ and $p \geq 1$, it is convenient to consider this at the same time.

The class \mathcal{L}_p

For $p \geq 1$, a function f in M is said to be of class L_p if $|f|^p$ is μ -integrable. Since

$$|f(x) + g(x)| \leq \begin{cases} 2|f(x)|, & \text{if } |f(x)| \geq |g(x)|, \\ 2|g(x)|, & \text{if } |g(x)| \geq |f(x)|; \end{cases}$$

we have, for all x ,

$$|f(x) + g(x)|^p \leq 2^p\{|f(x)|^p + |g(x)|^p\}. \quad (7.3.1)$$

Thus, if $f, g \in L_p$ we must have $(f \pm g) \in L_p$. With the usual convention about the set of zero measure where $(\alpha f + \beta g)$ may not be defined, it follows that L_p is a linear space. For $f, g \in L_p$ we define

$$\rho_p(f, g) = \left[\int |f - g|^p d\mu \right]^{1/p}$$

and notice again that $\rho_p(f, g) = 0$ if and only if $f = g$ a.e. so that in the space $\mathcal{L}_p \subset \mathcal{M}$ of equivalence classes we have

$$\rho_p(f, g) = \rho_p(g, f),$$

$$\rho_p(f, g) = 0 \quad \text{if and only if} \quad f = g \text{ in } \mathcal{L}_p.$$

We will prove in the next section that ρ_p satisfies the triangle inequality, which shows that it is a metric in \mathcal{L}_p . However, we can now define:

Convergence in p th mean

A sequence $\{f_n\}$ of functions in L_p (or in \mathcal{L}_p) is said to converge in p th mean to a function f in L_p if $\rho_p(f_n, f) \rightarrow 0$ as $n \rightarrow \infty$. A sequence $\{f_n\}$ of functions in L_p is a Cauchy sequence in p th mean if $\rho_p(f_n, f_m) \rightarrow 0$ as $n, m \rightarrow \infty$.

It is immediate, by (7.3.1) that convergence in p th mean to a function implies that we have a Cauchy sequence in p th mean. Completeness for this type of convergence can now be proved.

Theorem 7.3. *For $p \geq 1$, if $\{f_n\}$ is a sequence of functions in L_p which is a Cauchy sequence in p th mean, then there is an f in L_p such that $f_n \rightarrow f$ in p th mean.*

Proof. We again use the device of obtaining a subsequence which will converge a.e. to f . For any $\epsilon > 0$, let $N(\epsilon)$ denote an integer such that

$$\int |f_r - f_s|^p d\mu < \epsilon^{p+1} \quad \text{for } r, s \geq N(\epsilon).$$

Put $N_k = N(\epsilon 2^{-k})$, and assume that $N_{k+1} > N_k$ for each integer k . Then

$$\mu(E(\epsilon, r, s)) < \epsilon \quad \text{for } r, s \geq N(\epsilon),$$

where

$$E(\epsilon, r, s) = \{x: |f_r - f_s| \geq \epsilon\}.$$

If we put

$$E_k = E(\epsilon 2^{-k}, N_{k+1}, N_k),$$

$$F_k = \bigcup_{i=k}^{\infty} E_i,$$

we have $\mu(E_k) < 2^{-k}\epsilon$, $\mu(F_k) < 2^{1-k}\epsilon$, and if x is not in F_k ,

$$|f_{N_{i+1}}(x) - f_{N_i}(x)| < \epsilon 2^{-i} \quad \text{for all } i \geq k.$$

Hence the series $\sum_{i=1}^{\infty} (f_{N_{i+1}} - f_{N_i})$ converges outside $F = \bigcap_{k=1}^{\infty} F_k$ and $\mu(F) = 0$. Suppose then that $f_{N_i} \rightarrow f$ a.e. For a fixed integer r , if we put $g_i = |f_{N_i} - f_r|^p$, $g = |f - f_r|^p$ we obtain a sequence g_i of non-negative measurable functions with $\liminf g_i = \lim g_i = g$ a.e. By theorem 5.7 (Fatou) we have

$$\int g d\mu \leq \liminf_{i \rightarrow \infty} \int |f_{N_i} - f_r|^p d\mu < \epsilon \quad \text{if } r > N(\epsilon).$$

Hence, g is integrable, so that $(f - f_r) \in L_p$ which implies that $f \in L_p$. We have also proved that

$$\int |f - f_r|^p d\mu < \epsilon \quad \text{if } r > N(\epsilon)$$

so that $f_r \rightarrow f$ in p th mean. \blacksquare

It is worth remarking at this stage that the theorem corresponding to theorem 7.3 for Riemann integrals over a finite interval is false. It is not difficult to construct an example of a sequence of functions whose p th powers are Riemann integrable and which Cauchy converges in p th mean, but for which the limit is necessarily discontinuous on a set of positive measure and so cannot be Riemann integrable by theorem 5.9 (see exercise 7.3(10)). Thus theorem 7.3 exhibits another way in which the Lebesgue integral is a big improvement on the Riemann integral.

We now relate convergence in p th mean to convergence in measure.

Theorem 7.4. *If $\{f_n\}$ is a sequence of functions of L_p ($p \geq 1$) which is a Cauchy sequence in p th mean then $\{f_n\}$ is a Cauchy sequence in measure. If $f_n \rightarrow f$ in p th mean, then $f_n \rightarrow f$ in measure.*

Proof. For any h in L_p , $\eta > 0$

$$\mu\{x: |h(x)| \geq \eta^{1/2p}\} > \eta^{\frac{1}{2}} \Rightarrow \int |h|^p d\mu \geq \eta.$$

If $\{f_n\}$ is not a Cauchy sequence in measure, then there is an $\epsilon > 0$, $\delta > 0$ for which

$$\mu\{x: |f_n(x) - f_m(x)| \geq \epsilon\} > \delta$$

for infinitely many n, m . If now $\eta > 0$ is small enough to ensure that $\epsilon \geq \eta^{1/2p}$, $\delta > \eta^{\frac{1}{2}}$ we have

$$\int |f_n(x) - f_m(x)|^p d\mu \geq \eta > 0$$

for infinitely many n, m so that $\{f_n\}$ is not a Cauchy sequence in p th mean. This proves the first statement: the second part of the theorem is proved similarly. \blacksquare

Remark. The example after theorem 7.2 shows that $\{f_n\}$ may converge in p th mean but not converge a.e., though theorems 7.2, 7.3 together show that there must be a subsequence $\{f_{n_i}\}$ which converges a.e. If we consider Lebesgue measure in \mathbf{R} and put

$$f_n(x) = \begin{cases} n^{-1/p} & \text{for } x \text{ in } [0, n], \\ 0 & \text{otherwise,} \end{cases}$$

$$g_n(x) = \begin{cases} n^{1/p} & \text{for } x \text{ in } [0, 1/n], \\ 0 & \text{otherwise.} \end{cases}$$

we see that $f_n \rightarrow 0$ uniformly (and therefore almost uniformly, a.e., and in measure) but not in p th mean. If $\Omega = [0, 1]$, then $g_n \rightarrow 0$ almost uniformly, a.e. and in measure, but not in p th mean so that even in a finite measure space we cannot deduce convergence in mean from other types of convergence without some additional condition, even if the functions concerned are all in \mathcal{L}_p . The next definition turns out to be appropriate:

Set functions equicontinuous at \emptyset

Suppose ν_i ($i \in I$) is a family of set functions defined on a σ -field \mathcal{F} . The family is said to be equicontinuous at \emptyset if, given $\epsilon > 0$ and any sequence $\{B_n\}$ of sets of \mathcal{F} which decreases to \emptyset , there is an integer N such that

$$|\nu_i(B_n)| < \epsilon \quad \text{for all } i \in I, \quad n \geq N.$$

In §6.4 we saw that a set function ν was absolutely continuous with respect to μ if, given $\epsilon > 0$ there is a $\delta > 0$ such that, for

$$E \in \mathcal{F}, \quad \mu(E) < \delta \Rightarrow |\nu(E)| < \epsilon;$$

and that this condition was also necessary if ν was a finite valued measure. This makes the following definition reasonable:

Uniform absolute continuity

Any family $\nu_i (i \in I)$ of set functions defined on \mathcal{F} is said to be uniformly absolutely continuous with respect to μ if, given $\epsilon > 0$ there is a $\delta > 0$ such that, for $E \in \mathcal{F}$, $\mu(E) < \delta \Rightarrow |\nu_i(E)| < \epsilon$ for all i . To see what this condition means, suppose $\nu_i (i \in I)$ is a family of measures each of which is absolutely continuous with respect to μ , but such that the family is *not* uniformly absolutely continuous. Then there is an $\epsilon > 0$, and a sequence $\{B_n\}$ of sets of \mathcal{F} with indices $\{i_n\}$ such that

$$\mu(B_n) < 2^{-n}, \quad \nu_{i_n}(B_n) \geq \epsilon.$$

Put
$$A_k = \bigcup_{n=k}^{\infty} B_n, \quad C = \lim_{k \rightarrow \infty} A_k.$$

Then $\mu(C) = 0$ and $\lim (A_k - C) = \emptyset$. It follows that

$$\nu_{i_k}(A_k - C) = \nu_{i_k}(A_k) \geq \nu_{i_k}(B_k) \geq \epsilon > 0$$

so that, by considering the sequence $\{A_k - C\}$ which decreases to \emptyset , we see that the family $\nu_i (i \in I)$ is not equicontinuous at \emptyset . Thus we have proved

Lemma. Suppose $\nu_i (i \in I)$ is a family of measures on \mathcal{F} each of which is absolutely continuous w.r.t. μ . Then if the family is equicontinuous at \emptyset , it is uniformly absolutely continuous w.r.t. μ .

Theorem 7.5. Suppose $\{f_n\}$ is a sequence of functions of L_p and

$$\nu_n(E) = \int_E |f_n|^p d\mu, \quad (E \in \mathcal{F}, n = 1, 2, \dots).$$

(i) $\{f_n\}$ is a Cauchy sequence in p th mean if and only if $\{f_n\}$ is a Cauchy sequence in measure and the family $\{\nu_n\}$ of measures is equicontinuous at \emptyset .

(ii) The sequence $\{f_n\}$ converges to f in p th mean if f_n converges to f in measure and $\{\nu_n\}$ is equicontinuous at \emptyset .

Proof. (i) Suppose first that $\{f_n\}$ is a Cauchy sequence in p th mean. Then by theorem 7.4 $\{f_n\}$ is a Cauchy sequence in measure. For each $\epsilon > 0$, there is an N such that

$$\int |f_n - f_N|^p d\mu < \frac{\epsilon}{2^{p+1}} \quad \text{for } n \geq N.$$

Now suppose $\{B_k\}$ is a sequence of sets of \mathcal{F} decreasing to \emptyset . Since ν_n is absolutely continuous for $n = 1, 2, \dots, N$ we can find, by theorem 5.6 an integer k_0 such that

$$\int_{B_k} |f_n|^p < \frac{\epsilon}{2^{p+1}} \quad \text{for } k \geq k_0 \quad (n = 1, 2, \dots, N).$$

By (7.3.1) we obtain, for $n \geq N$, $k \geq k_0$,

$$\begin{aligned} \int_{B_k} |f_n|^p d\mu &\leq 2^p \int_{B_k} |f_N|^p d\mu + 2^p \int_{B_k} |f_n - f_N|^p d\mu \\ &< \frac{\epsilon}{2} + 2^p \int_{\Omega} |f_n - f_N|^p d\mu < \epsilon, \end{aligned}$$

so that the sequence $\{\nu_n\}$ is equicontinuous at \emptyset .

In the other direction, since we assume that μ is σ -finite on Ω , there must be a sequence $\{E_n\}$ in \mathcal{F} which decreases to \emptyset and is such that $\mu(\Omega - E_n)$ is finite for all n . Given $\epsilon > 0$, the equicontinuity of ν_n now ensures that there is a set $E = E_N$ with $\mu(\Omega - E) < \infty$ and

$$\int_E |f_n|^p d\mu < \frac{\epsilon}{2^{p+2}} \quad \text{for all } n.$$

Thus, for all m, n , by (7.3.1)

$$\int_E |f_n - f_m|^p d\mu < \frac{1}{2}\epsilon. \quad (7.3.2)$$

Now put $\Omega - E = F$, $\mu(F) = \lambda$. By the lemma, the sequence $\{\nu_n\}$ of measures must be uniformly absolutely continuous. We can therefore find an $\eta > 0$ such that, for $B \in \mathcal{F}$, $\mu(B) < \eta$,

$$\nu_n(B) = \int_B |f_n|^p d\mu < \frac{\epsilon}{2^{p+3}}. \quad (7.3.3)$$

For each m, n put

$$C_{m,n} = \left\{ x: |f_m(x) - f_n(x)| \geq \left(\frac{\epsilon}{4\lambda} \right)^{1/p} \right\}.$$

Then
$$\int_{F - C_{m,n}} |f_m - f_n|^p d\mu \leq \frac{\epsilon}{4\lambda} \mu(F - C_{m,n}) \leq \frac{\epsilon}{4\lambda} \mu(F) = \frac{1}{4}\epsilon.$$

Since $\{f_n\}$ is a Cauchy sequence in p th mean we can find an n_0 such that $\mu(C_{m,n}) < \eta$ for $m \geq n_0$, $n \geq n_0$. This gives, by (7.3.3),

$$\int_{C_{m,n}} |f_m - f_n|^p d\mu \leq 2^p \int_{C_{m,n}} |f_m|^p d\mu + 2^p \int_{C_{m,n}} |f_n|^p d\mu < \frac{1}{4}\epsilon,$$

so that
$$\int_F |f_m - f_n|^p d\mu < \frac{1}{2}\epsilon \quad \text{for } m, n \geq n_0.$$

This, together with (7.3.2) gives

$$\int_{\Omega} |f_m - f_n|^p d\mu < \epsilon \quad \text{for } m, n \geq n_0.$$

(ii) If $f_n \rightarrow f$ in measure, then $\{f_n\}$ is a Cauchy sequence in measure so that by (i) the condition that $\{\nu_n\}$ is equicontinuous implies that $\{f_n\}$ is a Cauchy sequence in p th mean. By theorem 7.3, there exists a $g \in L_p$ such that $f_n \rightarrow g$ in p th mean. By theorem 7.4 (i), $f_n \rightarrow g$ in measure so that we have $f = g$ a.e. and it follows that $f_n \rightarrow f$ in p th mean. **■**

We can now slightly strengthen the dominated convergence theorem (5.8).

Theorem 7.6. Suppose $p \geq 1$, and $\{f_n\}$ is a sequence of measurable functions with $|f_n|^p \leq h \in L_1$ for each n . If either $f_n \rightarrow f_0$ in measure or $f_n \rightarrow f_0$ a.e., then $f_n \rightarrow f_0$ in p th mean.

Proof. We must have $\nu_n(E) = \int_E |f_n|^p d\mu \leq \int_E h d\mu$, so that the family $\{\nu_n\}$ is equicontinuous at \emptyset by theorem 5.6. If $f_n \rightarrow f_0$ in measure we can apply theorem 7.5 (ii) to obtain the result. On the other hand, in the proof of theorem 7.5 (i) we only use convergence in measure on the subset F of Ω with $\mu(F)$ finite. On F , $f_n \rightarrow f_0$ a.e. implies $f_n \rightarrow f_0$ in measure by theorems 7.1, 7.2, so that the condition $f_n \rightarrow f_0$ a.e., together with equicontinuity at \emptyset of $\{\nu_n\}$, implies convergence in p th mean of $\{f_n\}$. **■**

We have now defined convergence to a limit for sequences of functions in several different ways, and have proved completeness in each case. It may help to summarise the relationships by a number of diagrams (Figures 2 to 4). In each of these an arrow from A to B

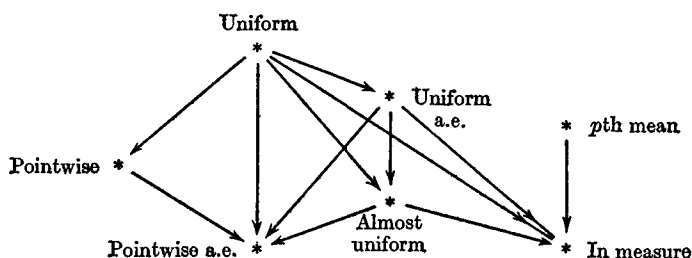
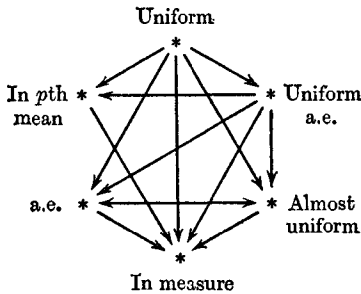
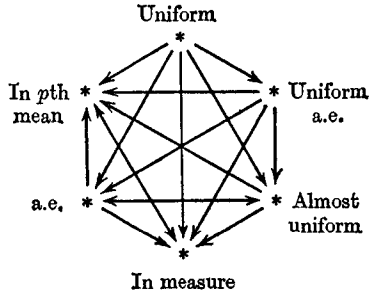


Fig. 2. No additional conditions.

indicates that convergence in sense A implies convergence in sense B . The lack of an arrow from A to B indicates that there is an example of a sequence satisfying the stated conditions which converges in sense A , but not in sense B . We assume throughout that we are considering functions in \mathcal{M} which are a.e. finite.

Fig. 3. $\mu(\Omega) < \infty$.Fig. 4. $|f_n|^p < h$ a.e., $h \in L_1$.

Exercises 7.3

1. Check Figures 2, 3, 4, stating in each case the theorem or theorems which justifies $A \rightarrow B$, or the example which justifies the exclusion of $A \rightarrow B$.

2. Show that, if $\mu(\Omega) < \infty$, then the condition in theorem 7.5 that $\{\nu_n\}$ be equicontinuous at \emptyset can be replaced by a condition that $\{\nu_n\}$ is uniformly absolutely continuous.

3. Show that if $\{f_n\}$ Cauchy converges uniformly a.e. and each f_n is integrable over E with $\mu(E)$ finite, then $f(x) = \lim f_n(x)$ a.e. is integrable over E and

$$\int_E |f_n - f| d\mu \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

4. Suppose Ω is set of positive integers and μ is the counting measure. Then

(i) If
$$f_n(k) = \begin{cases} 1/n & \text{for } 1 \leq k \leq n, \\ 0 & \text{for } k > n, \end{cases}$$

show that $f_n(x) \rightarrow 0$ uniformly on Ω , each f_n is integrable but

$$\int f_n d\mu \nrightarrow \int \lim f_n d\mu.$$

This shows $\mu(E) < \infty$ is essential in question (3).

(ii) For the same sequence $\{f_n\}$ show that

$$\nu_n(E) = \int_E f_n d\mu$$

is uniformly absolutely continuous, but not equicontinuous at \emptyset . This shows that the condition $\mu(\Omega) < \infty$ is essential in question (2).

(iii) If

$$f_n(k) = \begin{cases} 1/k & \text{for } 1 \leq k \leq n, \\ 0 & \text{for } k > n, \end{cases}$$

show that $\{f_n\}$ is uniformly convergent on Ω , each f_n is integrable, but the limit is not.

5. Show that, if $\{f_n\}$ converges in p th mean to f , and g is essentially bounded, then $\{f_n g\}$ converges in p th mean to fg .

6. Show that if
$$\nu_n(E) = \int_E f_n d\mu \quad (n = 1, 2, \dots),$$

defines a sequence of set functions which is uniformly absolutely continuous then so does

$$\lambda_n(E) = \int_E |f_n| d\mu.$$

7. Suppose $\{f_n\}$ is a sequence of functions in L_1 . Show that $\{f_n\}$ is a Cauchy sequence in mean if and only if

$$\int_E f_n d\mu = x_n$$

is a Cauchy sequence of real numbers for every $E \in \mathcal{F}$, and $\{f_n\}$ is a Cauchy sequence in measure. Give an example of a sequence which does not converge in measure, for which

$$\lim_{n \rightarrow \infty} \int_E f_n d\mu = 0$$

for all E .

8. Suppose $\mu(\Omega) < \infty$, and for $f, g \in \mathcal{M}$, and a.e. finite;

$$\rho(f, g) = \int \frac{|f - g|}{1 + |f - g|} d\mu.$$

Show that ρ defines a metric in the space of a.e. finite functions of \mathcal{M} , and that convergence in this metric is equivalent to convergence in measure.

9. Suppose $\mu(\Omega) < \infty$ ($1 \leq q < p$). Show that $\mathcal{L}_1 \supset \mathcal{L}_q \supset \mathcal{L}_p \supset \mathcal{L}_\infty$, and that $\rho_\infty(f, 0) = \lim_{p \rightarrow \infty} \rho_p(f, 0)$ for $f \in \mathcal{L}_\infty$. Show that the finite measure condition is essential.

By considering a suitable function on $[0, 1]$ show that $\mathcal{L}_\infty \neq \bigcap_{p>1} \mathcal{L}_p$, but that if $f \in \bigcap_{p>1} \mathcal{L}_p - \mathcal{L}_\infty$, then $\rho_p(f, 0) \rightarrow \infty$ as $p \rightarrow \infty$.

10. Suppose $\Omega = [0, 1]$, μ is Lebesgue measure. Let K be a nowhere dense perfect set with positive measure and let $\{E_k\}$ be the set of disjoint open intervals such that $(0, 1) - K = \bigcup_{i=1}^{\infty} E_i$. Let f_n be the indicator function

of $F_n = \bigcup_{i=1}^n E_i$. Prove that f_n ($n = 1, 2, \dots$) is Riemann integrable and con-

verges in mean to the indicator function f of $(0, 1) - K$. By considering the construction of K , show that f is discontinuous a.e. on the set K of positive measure, and so cannot be Riemann integrable. This shows that the class of Riemann integrable functions is not complete with respect to convergence in mean.

7.4 Inequalities

We now obtain some inequalities which turn out to be important in several branches of analysis. We need to use the algebraic inequality

$$x^\alpha y^\beta \leq \alpha x + \beta y \quad (7.4.1)$$

for $x > 0, y > 0, \alpha > 0, \beta > 0, \alpha + \beta = 1$, which is strict unless $x = y$. This is most easily proved by taking logarithms to give

$$\alpha \log x + \beta \log y \leq \log(\alpha x + \beta y)$$

and using the fact that $\log: \mathbf{R}^+ \rightarrow \mathbf{R}$ is strictly concave since it has a negative second derivative.

Conjugate indices

If $p > 1, q > 1$ and $1/p + 1/q = 1$, we say that p and q are conjugate indices.

Theorem 7.7 (Hölder). *If p, q are conjugate indices, and $f \in L_p, g \in L_q$ then fg is integrable and, for each E in \mathcal{F} ,*

$$\int_E |fg| d\mu \leq \left(\int_E |f|^p d\mu \right)^{1/p} \left(\int_E |g|^q d\mu \right)^{1/q}.$$

The inequality is strict unless there exist real numbers a, b such that $a|f|^p = b|g|^q$ a.e. on E .

Proof. If

$$\int_E |fg| d\mu = 0,$$

then the loose inequality is certainly satisfied and if the right-hand side is also zero then either $f = 0$ a.e. on E or $g = 0$ a.e. on E . In either case the condition $a|f|^p = b|g|^q$ a.e. is satisfied with $b = 0$ or $a = 0$, respectively. Hence we may assume that

$$\int_E |fg| d\mu > 0.$$

Put

$$E_1 = \{x: |g(x)| \leq |f(x)|^{p-1}\},$$

$$E_2 = \Omega - E_1.$$

Then

$$\text{for } x \in E_1, \quad |f(x)g(x)| \leq |f(x)|^p,$$

$$\text{for } x \in E_2, \quad |f(x)g(x)| \leq |g(x)|^q,$$

so that

$$|f(x)g(x)| \leq |f(x)|^p + |g(x)|^q, \quad \text{for all } x,$$

this implies that fg is integrable.

Given $E \in \mathcal{F}$, put $E_0 = E \cap \{x: f(x)g(x) = 0\}$. Then by our assumption $\mu(E - E_0) > 0$. For $x \in E - E_0$, we can apply (7.4.1) replacing

$$\alpha \text{ by } \frac{1}{p}, \quad \beta \text{ by } \frac{1}{q}, \quad x \text{ by } \left[\frac{|f(x)|^p}{\int_{E-E_0} |f|^p d\mu} \right] \quad \text{and} \quad y \text{ by } \left[\frac{|g(x)|^q}{\int_{E-E_0} |g|^q d\mu} \right].$$

This gives

$$\frac{|f(x)g(x)|}{\left(\int_{E-E_0} |f|^p d\mu\right)^{1/p} \left(\int_{E-E_0} |g|^q d\mu\right)^{1/q}} \leq \frac{|f(x)|^p}{p \int_{E-E_0} |f|^p d\mu} + \frac{|g(x)|^q}{q \int_{E-E_0} |g|^q d\mu}. \quad (7.4.2)$$

If we now integrate over $(E - E_0)$ and note that the right-hand side gives $1/p + 1/q = 1$, we obtain the desired inequality for the integral over $(E - E_0)$. We can only obtain equality for the integrals if we have equality in (7.4.2) for almost all $x \in (E - E_0)$. The condition for equality in (7.4.1) now shows that we must have $a|f|^p = b|g|^q$ a.e. on $(E - E_0)$ where

$$\frac{1}{a} = \int_{E-E_0} |f|^p d\mu \quad \text{and} \quad \frac{1}{b} = \int_{E-E_0} |g|^q d\mu.$$

The inequality for the integrals over E now follows, and we can again only have equality if $f = g = 0$ a.e. on E_0 , since otherwise the right-hand side is increased while the left-hand side remains the same on replacing $(E - E_0)$ by E . Thus equality can only occur if $a|f|^p = b|g|^q$ a.e. on E . **]**

Remark 1. The special case $p = q = 2$ of theorem 7.7 is called Schwartz's inequality. A simpler proof of this case is possible. See exercise 5.4 (13).

Remark 2. In the sense of theorem 7.7 the index conjugate to $p = 1$ is $q = \infty$. It is easy to prove directly that, if $f \in L_1$, $g \in L_\infty$ then

$$\int_E |fg| d\mu \leq \left(\int_E |f| d\mu \right) (\text{ess sup } |g\chi_E|).$$

Theorem 7.8 (Minkowski). For $p \geq 1$, if $f, g \in L_p$ then $(f+g) \in L_p$ and for any $E \in \mathcal{F}$,

$$\left\{ \int_E |f+g|^p d\mu \right\}^{1/p} \leq \left(\int_E |f|^p d\mu \right)^{1/p} + \left(\int_E |g|^p d\mu \right)^{1/p}.$$

For $p > 1$, equality is strict unless there are real numbers $a > 0$, $b > 0$ such that $af = bg$ a.e. on E .

Proof. We already proved in §7.3 that L_p is a linear space. For $p = 1$, the result is immediate. For $p > 1$,

$$\begin{aligned}\int_E |f+g|^p d\mu &= \int_E |f+g| |f+g|^{p-1} d\mu \\ &\leq \int_E |f| |f+g|^{p-1} d\mu + \int_E |g| |f+g|^{p-1} d\mu,\end{aligned}$$

with equality if and only if f and g have the same sign a.e. in E . If we now apply theorem 7.7 to each of these integrals we obtain

$$\begin{aligned}\int_E |f+g|^p d\mu &\leq \left(\int_E |f|^p d\mu \right)^{1/p} \left(\int_E |f+g|^p d\mu \right)^{1/q} \\ &\quad + \left(\int_E |g|^p d\mu \right)^{1/p} \left(\int_E |f+g|^p d\mu \right)^{1/q}\end{aligned}$$

with strict inequality unless there are numbers α, β, γ such that $\alpha|f|^p = \beta|f+g|^p = \gamma|g|^p$ a.e. We can only have equality in both inequalities if there is $a \geq 0, b \geq 0$ with $af = bg$ a.e. Provided it is not zero we now divide both sides by

$$\left(\int_E |f+g|^p d\mu \right)^{1/q}$$

to obtain the desired result. If

$$\int_E |f+g|^p d\mu = 0,$$

then the inequality is trivially satisfied, and equality is only possible if $f = g = 0$ a.e. **■**

The above theorem shows that

$$\rho_p(f, g) = \left(\int_E |f-g|^p d\mu \right)^{1/p}$$

defines a metric in \mathcal{L}_p .

We have proved the Hölder and Minkowski inequalities for general measure spaces. They are of course valid for Lebesgue measure in \mathbf{R} either over a finite interval or over the whole line.

However, we can also apply these general theorems to the case where Ω is the set of positive integers \mathbf{Z} , and μ is the counting measure

$$\mu(E) = \text{number of integers in } E,$$

which makes all subsets $E \subset \mathbf{Z}$ measurable. Then functions $f: \mathbf{Z} \rightarrow \mathbf{R}$ and $g: \mathbf{Z} \rightarrow \mathbf{R}$ reduce to

$$f(i) = a_i, \quad g(i) = b_i \quad (i = 1, 2, \dots),$$

where $\{a_i\}$ and $\{b_i\}$ are sequences of real numbers, and we can apply theorems 7.7, 7.8 to give:

Hölder.
$$\sum_{i=1}^{\infty} |a_i b_i| \leq \left(\sum_{i=1}^{\infty} |a_i|^p \right)^{1/p} \left(\sum_{i=1}^{\infty} |b_i|^q \right)^{1/q}$$

in the sense that the convergence of both series on the right implies the convergence of the series on the left and the inequality. Further, equality is only possible if there is a constant k such that $|a_i|^p = k|b_i|^q$ for all i .

Minkowski.

$$\left(\sum_{i=1}^{\infty} |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{\infty} |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^{\infty} |b_i|^p \right)^{1/p}$$

with equality if and only if there is a $k > 0$ such that $a_i = kb_i$ for all i .

It is interesting to see how these elementary inequalities (which can of course be proved directly) fall out of the general theorems by using a simple special measure.

Exercises 7.4

1. Suppose $\alpha > 0, \beta > 0, \gamma > 0, \alpha + \beta + \gamma = 1$ and $f \in L_\alpha, g \in L_\beta, h \in L_\gamma$. Show that

$$\int_{\mathbb{E}} |fgh| d\mu \leq \left(\int_{\mathbb{E}} |f|^{1/\alpha} d\mu \right)^\alpha \left(\int_{\mathbb{E}} |g|^{1/\beta} d\mu \right)^\beta \left(\int_{\mathbb{E}} |h|^{1/\gamma} d\mu \right)^\gamma.$$

Generalise to $n \geq 2$ functions.

2. If μ is Lebesgue measure on $I = (a, b)$ and $f \in L_p(I, \mu)$ show that there is a continuous g such that

$$\int_a^b |f(x) - g(x)|^p dx < \epsilon.$$

Deduce that $\int_a^b |f(x+h) - f(x)|^p dx \rightarrow 0$ as $h \rightarrow 0$

(here f is given the value zero outside I). *Hint.* See theorem 5.10.

3. If $\mu(\Omega) < \infty, 1 \leq p < q$ and $f \in L_q$ show that

$$[\rho_p(f, 0)] \leq k\rho_q(f, 0)$$

for a suitable constant k . Show that $\mu(\Omega) < \infty$ is essential.

4. Show by an example that theorem 7.8 is false for $p < 1$.

5. If p, q are conjugate indices, $f_n \rightarrow f_0$ in p th mean, $g_n \rightarrow g_0$ in q th mean, show that $f_n g_n \rightarrow f_0 g_0$ in mean.

6. By considering intervals of \mathcal{P}^k whose coordinates are rational, and linear combinations of indicator functions of such intervals obtain a countable dense set in \mathcal{L}_p for $\Omega = \mathbf{R}^k$, μ any Lebesgue–Stieltjes measure. Such a space \mathcal{L}_p is therefore separable.

7.5* Measure preserving transformations from a space to itself

In §6.5 we discussed measurable transformations T from (X, \mathcal{F}, μ) to (Y, \mathcal{S}) and defined the measure μT^{-1} on \mathcal{S} in terms of the transformation. A special case of this is obtained when $T: X \rightarrow X$ maps X into itself. We then say that T is *measure preserving* if, for every $E \in \mathcal{F}$, $\mu(T^{-1}(E)) = \mu(E)$. Given a mapping $T: X \rightarrow X$, we can define the iterates T^n obtained by composing T with itself n times. For convenience T^0 will denote the identity mapping, and T^{-n} will be defined as a set mapping

$$T^{-n}(E) = \{x: T^n(x) \in E\}$$

even if T^{-n} is not a point function.

If \mathcal{F} is the σ -ring generated by a semi-ring \mathcal{P} , then it is immediate on applying the extension theorems of Chapters 3 and 4 that T is measure preserving if, and only if,

$$\mu(T^{-1}(E)) = \mu(E) \quad \text{for every } E \text{ in } \mathcal{P}.$$

If T is a $(1, 1)$ transformation from X to itself, then it is said to be *invertible* and the condition for T to be measure preserving in this case can be written as $\mu(T(E)) = \mu(E)$ for all E in \mathcal{F} .

In §4.5 we considered the geometrical properties of Lebesgue measure and showed that the transformations of Euclidean space defined in terms of translations, rotations or reflexions are measure preserving. One can also prove that a matrix transformation of determinant 1 defines a measure preserving transformation in Euclidean space. All these are easily seen to be invertible.

If $\Omega = [0, 1)$ and $Tx = 2x$ reduced mod 1 then, for Lebesgue measure, T is seen to be measure preserving by considering the effect of T^{-1} on the dyadic intervals $[p/2^q, (p+1)/2^q)$ which form a semi-ring generating \mathcal{B} . If $x = .\alpha_1\alpha_2\dots$ is the expansion of x as a binary ‘decimal’, then $Tx = .\alpha_2\alpha_3\dots$. This T is not invertible.

It is worth remarking that the study of measure preserving transformations started with certain considerations in statistical mechanics. Suppose we have a system of k particles whose present state is described by a point in ‘phase space’ \mathbf{R}^{6k} in which each particle determines

3 coordinates for position and 3 coordinates for momentum. Then the entire history of the system can be represented by a trajectory in phase space which is completely determined (assuming the laws of classical mechanics) by a single point on it. Thus for any (time) t we can define an invertible transformation T_t by saying that, for x in phase space, $T_t x$ denotes the state of a system which starts at x after a time t . One of the basic results in statistical mechanics (due to Liouville) states that, if the coordinates in phase space are correctly chosen, then the 'flow' in phase space leaves all volumes (i.e. Lebesgue measure in \mathbf{R}^{6k}) unchanged. This means that T_t becomes a measure preserving transformation in $(\mathbf{R}^{6k}, \mathcal{L}^{6k}, \mu)$. In practice k is enormous, and it is not possible to observe at any one moment all the particles of the system. Instead one asks questions like 'what is the probability that at time t the state of the system belongs to a given subset of phase space?' One then imposes conditions which ensure that this can be calculated by considering the 'average' behaviour of $T_t x$ as $t \rightarrow \infty$. To be more precise $T_{s+t} = T_s T_t$ so that $T_{nt} = (T_t)^n$ and one can consider a discrete model, count the proportion of times up to n that $T^i x \in E$ where $T = T_{t_0}$ and then let $n \rightarrow \infty$. In practice a set E in phase space is replaced by a function $f(x)$ (representing some physical measurement) and one considers the average behaviour in terms of the sequence

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \quad (n = 1, 2, \dots).$$

This discussion of phase space provides a physical interpretation for the mathematical results which we now formulate precisely.

For the remainder of this section, T will denote a measure preserving transformation of Ω to itself, and $f: \Omega \rightarrow \mathbf{R}^*$ will denote an integrable function. We define

$$f_k(x) = f(T^k x) \quad (k = 0, 1, 2, \dots).$$

Then f_k will be integrable and theorem 6.8 shows immediately that

$$\int f_k d\mu = \int f d\mu.$$

Before giving the proof (due to F. Riesz) of the point-wise ergodic theorem, we obtain a lemma which is an important step towards it.

Lemma. (sometimes called the maximal ergodic theorem). Suppose E is the set of points $x \in \Omega$ such that

$$\sum_{i=0}^n f_i(x) \geq 0$$

for at least one n : then

$$\int_E f d\mu \geq 0.$$

Proof. We first need a result about finite sequences of real numbers. Suppose $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbf{R}$ and $m \leq n$. A term α_i of this sequence is called an m -leader if there is an integer p , $1 \leq p \leq m$, such that

$$\alpha_i + \alpha_{i+1} + \dots + \alpha_{i+p-1} \geq 0.$$

For a fixed m , let α_k be the first m -leader, and let $(\alpha_k + \dots + \alpha_{k+p-1})$ be the shortest non-negative sum that it leads. Then for every integer h with $k \leq h \leq k+p-1$, we must have $\alpha_h + \alpha_{h+1} + \dots + \alpha_{k+p-1} \geq 0$, so that α_h is an m -leader. Now continue with the first m -leader in $\alpha_{k+p}, \dots, \alpha_n$ and repeat the argument until all the m -leaders have been found. It follows that the sum of *all* the m -leaders of the original sequence must be non-negative, as it is the same as the sum of the non-negative shortest sums obtained by the above procedure.

We can now turn to the proof of the lemma and notice that, since f is integrable, we may assume that it is everywhere finite valued. If E_m denotes the set of x such that

$$\sum_{i=0}^n f_i(x) \geq 0$$

for at least one $n \leq m$, then E_m increases to E , so it is sufficient to prove

$$\int_{E_m} f d\mu \geq 0 \quad \text{for all } m.$$

For a positive integer n , let $s(x)$ be the sum of the m -leaders of the finite sequence $f_0(x), f_1(x), \dots, f_{n+m-1}(x)$. Let A_k be the set of x for which $f_k(x)$ is an m -leader and let χ_k be the indicator function of A_k . Since A_k is measurable, and $s(x) = \sum_{k=0}^{n+m-1} \chi_k(x) f_k(x)$, s is measurable and integrable and $s(x) \geq 0$ so that

$$\sum_{k=0}^{n+m-1} \int_{A_k} f_k d\mu \geq 0. \quad (7.5.1)$$

Now notice that, for $k = 1, 2, \dots, n-1$, $T(x) \in A_{k-1}$ if and only if $f_{k-1}(Tx) + \dots + f_{k-1+p-1}(Tx) \geq 0$ for some $p \leq m$, which is equivalent to $f_k(x) + \dots + f_{k+p-1}(x) \geq 0$ for some $p \leq m$ which in turn is the condition for $x \in A_k$. Thus $A_k = T^{-1}A_{k-1} = T^{-k}A_0$ for $k = 1, \dots, n-1$. Hence by theorem 6.8,

$$\int_{A_k} f_k(x) d\mu = \int_{T^{-k}A_0} f(T^k x) d\mu = \int_{A_0} f(x) d\mu$$

so that the first n terms of (7.5.1) are all equal. Now $A_0 = E_m$, so that (7.5.1) implies

$$n \int_{E_m} f d\mu + m \int |f| d\mu \geq 0.$$

Divide by n , keep m fixed and let $n \rightarrow \infty$ to give

$$\int_{E_m} f d\mu \geq 0. \blacksquare$$

Theorem 7.9 (Birkhoff). Suppose T is a measure preserving transformation on a σ -finite measure space $(\Omega, \mathcal{F}, \mu)$ to itself and $f: \Omega \rightarrow \mathbf{R}^*$ is integrable. Then

- (i) $\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)$ converges point-wise a.e.;
- (ii) the limit function $f^*(x)$ is integrable and invariant under T (i.e. $f^*(Tx) = f^*(x)$ a.e.);
- (iii) if $\mu(\Omega) < \infty$, then $\int f^* d\mu = \int f d\mu$.

Proof (i). Suppose r, s are rational numbers $r < s$ and $B = B(r, s)$ is the set of points x for which

$$\liminf \left[\frac{1}{n} \sum_{i=0}^{n-1} f_i(x) \right] < r < s < \limsup \left[\frac{1}{n} \sum_{i=1}^{n-1} f_i(x) \right].$$

It is immediate that B is measurable and invariant under T . Our result will now follow if we can prove that $\mu(B(r, s)) = 0$ for all rational r, s . The first step in this direction is to show that $\mu(B) < \infty$.

We may assume without loss of generality that $s > 0$, for otherwise the argument can be carried out with f replaced by $-f$. Suppose $C \in \mathcal{F}$, $\mu(C) < \infty$, $C \subset B$, and χ is the indicator function of C . Apply the lemma to the function $(f - s\chi)$ to give

$$\int_E (f - s\chi) d\mu \geq 0,$$

where E is defined in the lemma. If $x \in B$, then at least one of the averages

$$\frac{1}{n} \sum_{i=0}^{n-1} f_i(x) > s > 0$$

so that at least one of the sums

$$\sum_{i=0}^{n-1} \{f(T^i x) - s\chi(T^i x)\} \geq 0,$$

and it follows that $x \in E$. Thus

$$\int_E f d\mu \geq \int_E s\chi d\mu \quad \text{so that} \quad \mu(C) \leq \frac{1}{s} \int_E |f| d\mu.$$

Since B has σ -finite measure and its subsets of finite measure have bounded measure, it follows that $\mu(B)$ is finite. Since B is invariant under T we can restrict our σ -field and measure to B and think of T as a measure preserving transformation on B . Apply the lemma again to the integrable function $(f-s)$, and note that in this case the set E of the lemma is the whole space B . This gives

$$\int_B (f-s) d\mu \geq 0.$$

Similarly, we can obtain $\int_B (r-f) d\mu \geq 0$.

Together these give $\int_B (r-s) d\mu \geq 0$.

Since $r < s$, we must have $\mu(B) = 0$.

(ii) Put

$$f^*(x) = \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} f_i(x) \right\}$$

when the sequence converges. Then it is immediate that f^* is measurable and invariant. Further

$$\int \left| \frac{1}{n} \sum_{i=0}^{n-1} f_i(x) \right| d\mu \leq \frac{1}{n} \sum_{i=0}^{n-1} \int |f_i(x)| d\mu = \int |f(x)| d\mu$$

so that, by theorem 5.7 (Fatou) f^* is integrable, and

$$\int |f^*| d\mu \leq \int |f| d\mu.$$

(iii) For fixed n , put

$$D(k, n) = \left\{ x: \frac{k}{2^n} \leq f^*(x) < \frac{k+1}{2^n} \right\}$$

and apply the lemma to the transformation T on the set $D(k, n)$ which can be assumed to be invariant. Then $f^*(x) \geq k/2^n$ in $D(k, n)$, so that at least one of the sums

$$\sum_{i=0}^{n-1} \left(f_i(x) - \frac{k}{2^n} + \epsilon \right) > 0$$

for each $\epsilon > 0$. Hence

$$\int_{D(k, n)} f d\mu \geq \left(\frac{k}{2^n} - \epsilon \right) \mu(D(k, n)),$$

and so we must have

$$\int_{D(k, n)} f d\mu \geq \frac{k}{2^n} \mu(D(k, n)).$$

Similarly

$$\int_{D(k,n)} f d\mu \leq \frac{k+1}{2^n} \mu(D(k,n))$$

and
$$\frac{k}{2^n} \mu(D(k,n)) \leq \int_{D(k,n)} f d\mu \leq \frac{k+1}{2^n} \mu(D(k,n)).$$

For each integer k , it follows that

$$\left| \int_{D(k,n)} f^* d\mu - \int_{D(k,n)} f d\mu \right| \leq \frac{1}{2^n} \mu(D(k,n));$$

and if we sum over k

$$\left| \int_{\Omega} f^* d\mu - \int_{\Omega} f d\mu \right| \leq \frac{1}{2^n} \mu(\Omega).$$

Since n is arbitrary we must have $\int f^* d\mu = \int f d\mu$. \blacksquare

For applications to statistical mechanics one would expect the equilibrium value $f^*(x)$ to be independent of the point x , so that the limit function f^* of theorem 7.9 is a constant. Unfortunately this is not true without imposing an additional condition.

Ergodic transformation

A measure preserving transformation T is said to be ergodic (or metrically transitive or metrically invariant) if for all invariant sets E (sets for which $T^{-1}(E) = (E)$), $\mu(E) = 0$ or $\mu(\Omega - E) = 0$.

Lemma. *T is ergodic if and only if every measurable invariant function is constant a.e.*

Proof. Suppose g is measurable and invariant. Then $\{x: g(x) \geq \alpha\}$ is invariant for all real α , and must either have zero measure or be the complement of a set of zero measure. Hence $g = \text{constant a.e.}$, if T is ergodic. Conversely, if every measurable invariant function is constant a.e., since the indicator function of an invariant set is an invariant function, there cannot be any invariant sets other than null sets and complements of null sets. \blacksquare

We can now apply this lemma to theorem 7.9 when T is ergodic. There are two cases:

(i) $\mu(\Omega) = +\infty$. Since the only constant which is integrable over a space of infinite measure is zero we deduce that

$$\frac{1}{n} \sum_{i=0}^{n-1} f_i(x) \rightarrow 0 \quad \text{a.e.}$$

(ii) $\mu(\Omega) < \infty$. We can integrate $f^* = c$ a.e. by (iii) to obtain

$$\frac{1}{n} \sum_{i=0}^{n-1} f_i(x) \rightarrow \frac{1}{\mu(\Omega)} \int f d\mu \quad \text{a.e.}$$

This last result ties up with our remarks about statistical mechanics, since it shows that the average value of f on the discrete trajectory approaches the average value of f in phase space for all starting points x except for a possible null set.

The reader who wishes to learn more about ergodic theory is advised to read P. R. Halmos, *Lectures on Ergodic Theory* (Chelsea, 1956).

Exercises 7.5

1. Suppose Ω is the real line, T is the translation $Tx = x + 1$, f is the indicator function of $[0, 1]$. What is

$$f^*(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_i(x)$$

in this case? Show that theorem 7.9 (iii) is not satisfied without the condition $\mu(\Omega) < \infty$.

2. Suppose T is measure preserving and ergodic on $(\Omega, \mathcal{F}, \mu)$ and

$$\mu(\Omega) < \infty.$$

If f is non-negative measurable and

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \rightarrow c \in \mathbb{R} \text{ a.e.,}$$

deduce that f is integrable.

3. Suppose Ω is five point space $\{a, b, c, d, e\}$, \mathcal{F} is the set of all subsets, $\mu\{a\} = \mu\{b\} = \mu\{c\} = 1$ and $\mu\{d\} = \mu\{e\} = 2$, T is the permutation $(a, b, c)(d, e)$. Show that T is measure preserving but not ergodic. Find the f^* of theorem 7.9 if f is the indicator function of $\{a, b, e\}$.

4. Suppose (Ω, \mathcal{F}, P) is a probability measure. Form the doubly infinite Cartesian product of copies of (Ω, \mathcal{F}, P) labelled $(\dots, -2, -1, 0, 1, \dots, n, \dots)$ and the product measure by the process of theorem 6.3. If a point of this product space is $w = (\dots, w_{-1}, w_0, w_1, \dots)$ and T is the shift $(Tw)_n = w_{n+1}$; show that T is measure preserving and ergodic.

5. If $\Omega = [0, 1)$, $Tx = 2x \pmod{1}$, and μ is Lebesgue measure, show that T is ergodic. By applying the ergodic theorem to the indicator function of $[0, \frac{1}{2})$ deduce the Borel normal number theorem which states that in the binary expansion $0 \cdot \alpha_1 \alpha_2 \dots \alpha_n \dots$ of real numbers in $[0, 1)$, the density

$$\frac{1}{n} \sum_{i=1}^n \alpha_i \rightarrow \frac{1}{2} \quad \text{for almost all } x.$$

6. Suppose T is ergodic and measure preserving on a finite measure space (X, \mathcal{F}, μ) and f, g are the indicator functions of measurable sets F, G . Show that

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=0}^{n-1} \mu((T^{-i}F) \cap G) \right] = \frac{\mu(F)\mu(G)}{\mu(X)}.$$

LINEAR FUNCTIONALS

In this chapter all measure spaces $(\Omega, \mathcal{F}, \mu)$ will be σ -finite, and \mathcal{F} will be complete with respect to μ , unless stated otherwise. In Chapter 7 we saw that \mathcal{L}_p ($1 \leq p < \infty$) with the metric

$$\rho_p(f, g) = \left\{ \int |f - g|^p d\mu \right\}^{1/p},$$

and \mathcal{L}_∞ with the metric

$$\rho_\infty(f, g) = \text{ess sup } |f - g|,$$

were complete metric spaces. We also proved they were linear spaces (over the reals); and it is immediate that the metric defines a norm

$$\|f\|_p = \rho_p(f, 0) \quad (1 \leq p \leq \infty),$$

for which the spaces are normed linear spaces. Thus

$$\|f\|_p > 0 \quad \text{if } f \neq 0, \quad \|0\| = 0,$$

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p,$$

$$\|\alpha f\|_p = |\alpha| \|f\|_p \quad \text{for } \alpha \in \mathbf{R}.$$

We will omit the suffix p in $\|\cdot\|_p$ when it is clear which \mathcal{L}_p space is being considered.

It turns out that the space \mathcal{L}_2 has some special properties not shared by other \mathcal{L}_p spaces. These can be postulated in terms of the difference between Hilbert space and Banach space, but we prefer to examine, in the first three sections of this chapter, the structure of \mathcal{L}_2 and then discuss later the analogous properties of more general normed linear spaces.

8.1 Dependence of \mathcal{L}_2 on the underlying $(\Omega, \mathcal{F}, \mu)$

In general, the structure of the space \mathcal{L}_2 depends on the underlying space $(\Omega, \mathcal{F}, \mu)$: when we want to emphasise this we use the notation $\mathcal{L}_2(\Omega, \mu)$. We first examine conditions on $(\Omega, \mathcal{F}, \mu)$ which will ensure that $\mathcal{L}_2(\Omega, \mu)$ is separable (in the topology of the norm). We later define (real) *Hilbert space* in terms of its abstract properties, and show that $\mathcal{L}_2(\Omega, \mu)$ is always a realisation of Hilbert space.

Countable basis for measure

In the measure space $(\Omega, \mathcal{F}, \mu)$ we can use the equivalence relation $A \sim B \Leftrightarrow \mu(A \Delta B) = 0$ to identify the subsets in \mathcal{F} which differ only by a set of measure zero. If we denote the resulting quotient space by \mathcal{F}_μ , it is clear that, when $\mu(\Omega) < \infty$, \mathcal{F}_μ is a metric space with the metric $\rho(A, B) = \mu(A \Delta B)$, and one can further show that \mathcal{F}_μ is complete. In this case we can define a dense subset by means of the topology of this metric. However, the notion of a dense subset in \mathcal{F}_μ can be extended to include the case $\mu(\Omega) = \infty$ by a device which makes sense provided μ is σ -finite on Ω . Thus we say that μ has a *countable basis* if there is a sequence $\{E_n\}$ of sets in \mathcal{F} such that, given $\epsilon > 0$ and any $A \in \mathcal{F}$ with $\mu(A) < \infty$, there is a set E_k of the sequence for which

$$\mu(A \Delta E_k) < \epsilon.$$

In Chapters 3 and 4 we saw how measures could be obtained by extending a measure already defined on a semi-ring. If μ can be defined by extending a finite measure on a semi-ring \mathcal{P} which contains only a countable number of sets, then μ has a countable basis. For the ring \mathcal{R} generated by \mathcal{P} is countable, and forms a basis for \mathcal{F} by theorem 4.4. In particular, in the definition of Lebesgue measure, we could have used the countable semi-ring of $\frac{1}{2}$ -open intervals, whose bounding coordinates are rationals, to generate the σ -field \mathcal{B}^k of Borel sets in \mathbf{R}^k ; so it follows that Lebesgue measure in \mathbf{R}^k has a countable basis.

We first obtain a condition equivalent to the existence of a countable basis for μ .

Lemma. *A measure μ has a countable basis if and only if, for each $\epsilon > 0$, any collection $\mathcal{C} \subset \mathcal{F}$ of subsets of finite measure such that*

$$A, B \in \mathcal{C}, \quad A \neq B \Rightarrow \mu(A \Delta B) \geq \epsilon \quad (8.1.1)$$

is countable.

Proof. Suppose first that $\epsilon > 0$ is such that there is a non-countable \mathcal{C} satisfying (8.1.1); and suppose if possible that μ has a countable basis \mathcal{D} . Then, for each $A \in \mathcal{C}$ we can find a set $E_A \in \mathcal{D}$ with

$$\mu(A \Delta E_A) < \frac{1}{3}\epsilon.$$

Then, if $A \neq B$,

$$\mu(E_A \Delta E_B) \geq \mu(A \Delta B) - \mu(A \Delta E_A) - \mu(B \Delta E_B) > \frac{1}{3}\epsilon > 0,$$

so that $E_A \neq E_B$. Thus if \mathcal{D} is dense, it contains a non-countable subclass, which is a contradiction.

Conversely, suppose the condition is satisfied. Then for each positive integer n , the set Γ_n of those classes \mathcal{C}_n which satisfy (8.1.1) with $\epsilon = 1/n$ can be partially ordered by inclusion. Clearly if $\Delta_n \subset \Gamma_n$ is a totally ordered set of classes, the union of all the classes in Δ_n is a class \mathcal{C}_n which is a maximal element of Δ_n . By Zorn's lemma (see §1.6) it follows that we can obtain a maximal element in Γ_n with respect to this ordering. Thus we can find a class $\mathcal{C}_n^0 \subset \mathcal{F}$ satisfying (8.1.1), with $\epsilon = 1/n$, and such that, given $E \in \mathcal{F}$, there is at least one $A \in \mathcal{C}_n^0$ with $\mu(A \triangle E) < 1/n$, as otherwise E could be added to \mathcal{C}_n^0 to form a larger collection. But \mathcal{C}_n^0 is countable so $\mathcal{C} = \bigcup_{n=1}^{\infty} \mathcal{C}_n^0$ is countable and forms a basis for μ . \blacksquare

Theorem 8.1. *The space $\mathcal{L}_2(\Omega, \mu)$ of functions $f: \Omega \rightarrow \mathbf{R}^*$ which are square integrable is separable (in the norm topology) if and only if the measure μ has a countable basis.*

Proof. Suppose first that \mathcal{L}_2 is separable, so that there is a sequence $\{f_n\}$ in \mathcal{L}_2 such that for any $\epsilon > 0$, $f \in \mathcal{L}_2$ we can find an integer k with $\|f - f_k\| < \epsilon$. Let \mathcal{C} be any collection of measurable sets of finite measure. Then for each $A \in \mathcal{C}$, the indicator function $\chi_A \in \mathcal{L}_2$, so there is an integer k_A such that

$$\|f_{k_A} - \chi_A\| < \frac{1}{3}\epsilon.$$

Then, if \mathcal{C} satisfies (8.1.1), we must have

$$\|f_{k_A} - f_{k_B}\| > \frac{1}{3}\epsilon \quad \text{for } A \neq B$$

so that $k_A \neq k_B$, and \mathcal{C} must be countable. By the lemma this implies that μ has a countable basis.

Conversely suppose that μ has a countable basis \mathcal{C} . The set \mathcal{S} of all simple functions

$$h = \sum_{i=1}^n r_i \chi_i$$

which are finite sums of rational multiples of indicator functions of sets of \mathcal{C} is then countable. In order to prove \mathcal{L}_2 separable, it is sufficient to show that this set \mathcal{S} is dense in \mathcal{L}_2 .

From the definition of the integral, for any $f \in \mathcal{L}_2$, $\epsilon > 0$ we can find a set $E \in \mathcal{F}$ with $\mu(E) < \infty$ such that f is bounded on E and

$$\int_{\Omega-E} |f|^2 d\mu < \frac{1}{3}\epsilon^2.$$

On the set E , we can use the process of theorem 5.2 to approximate f uniformly by a simple function g taking only rational values

$$g = \sum_{k=1}^n r_k \chi_{E_k} \begin{cases} E_i \cap E_j = \emptyset & (i \neq j), \\ E = \bigcup_{i=1}^n E_i. \end{cases}$$

Using $\mu(E) < \infty$, this means that such a function g can be found with

$$\int_E |f-g|^2 d\mu < \frac{1}{3}\epsilon^2.$$

Then

$$\begin{aligned} \|f-g\|^2 &= \int_{\Omega-E} |f-g|^2 d\mu + \int_E |f-g|^2 d\mu \\ &= \int_{\Omega-E} |f|^2 d\mu + \int_E |f-g|^2 d\mu \\ &< \frac{1}{4}\epsilon^2; \end{aligned}$$

so that $\|f-g\| < \frac{1}{2}\epsilon$. If all the r_k in the representation of g are zero we are finished, so there is no loss in generality in assuming they are all non-zero. Since \mathcal{C} is a basis for μ and $\mu(E_k) < \infty$, we can find sets $C_k \in \mathcal{C}$ such that

$$\mu(E_k \triangle C_k) < \left(\frac{\epsilon}{2r_k n}\right)^2 \quad (k = 1, 2, \dots, n).$$

Then

$$\|r_k \chi_{E_k} - r_k \chi_{C_k}\|^2 = r_k^2 \mu(E_k \triangle C_k) < \left(\frac{\epsilon}{2n}\right)^2,$$

so that, if

$$h = \sum_{k=1}^n r_k \chi_{C_k}, \quad \text{we have} \quad \|g-h\| \leq \sum_{k=1}^n \|r_k \chi_{E_k} - r_k \chi_{C_k}\| < \frac{1}{2}\epsilon$$

and

$$\|f-h\| \leq \|f-g\| + \|g-h\| < \epsilon. \blacksquare$$

Corollary. If μ denotes Lebesgue measure in \mathbf{R}^k , then $\mathcal{L}_2(\mathbf{R}^k, \mu)$ is separable.

To prove this we use the observation that Lebesgue measure in \mathbf{R}^k has a countable basis. It is worth remarking that the classical method of proving this corollary is to approximate $f \in \mathcal{L}_2$ by a continuous function vanishing outside a finite interval, and then approximate this continuous function by a rational polynomial.

We end this section by making an important definition which is essential to a geometric understanding of linear spaces. We will see later than it is *not* possible to define an inner product in \mathcal{L}_p for $p \neq 2$.

Inner product

For any normed linear space K over the reals a function (f, g) on $K \times K \rightarrow \mathbf{R}$ is called an inner product (or scalar product) if

- (i) $(f, g) = (g, f)$;
- (ii) $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$;
- (iii) $(\lambda f, g) = \lambda(f, g)$, for $\lambda \in \mathbf{R}$;
- (iv) $(f, f) = \|f\|^2$.

For the normed linear space \mathcal{L}_2 we can define

$$(f, g) = \int fg d\mu, \quad f, g \in \mathcal{L}_2,$$

since, by theorem 7.7, the product fg is integrable. It is a simple matter to check that, with this definition, (f, g) satisfies all the conditions (i)–(iv) for an inner product.

Exercises 8.1

1. For any normed linear space with an inner product, prove that

$$|(f, g)| \leq \|f\| \|g\|.$$

Hint. Consider $(f + \theta g, f + \theta g) \geq 0$ for all real θ .

2. If $(f, g) = 0$ in a normed linear space, show that

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2.$$

3. Suppose $(\Omega, \mathcal{F}, \mu)$ is a discrete measure space, i.e. there is a sequence $\{p_i\}$ of reals with $\sum |p_i| < \infty$ and a sequence $\{x_i\}$ in Ω such that $\mu(E) = \sum_{x_i \in E} p_i$.

Prove that $\mathcal{L}_2(\Omega, \mu)$ is separable.

4. If $(X, \mathcal{F}, \mu), (Y, \mathcal{G}, \nu)$ are σ -finite measure spaces each with countable bases, show that the product measure $\lambda = \mu \times \nu$ on $X \times Y$ also has a countable basis.

Hint. Consider finite unions of rectangles which are products of basic sets.

5. Generalise the above to countable products of spaces $(X_i, \mathcal{F}_i, \mu_i)$ with $\mu_i(X_i) = 1$. The example (8) below shows that it does not extend to arbitrary products.

6. Let Ω be any set and define the counting measure $\mu(E) =$ number of points in E when E is finite; $\mu(E) = +\infty$ otherwise. Show (i) if Ω is countable, the finite subsets of Ω form a countable basis; (ii) if Ω is uncountable, there is no countable basis for μ .

7. Show that any Lebesgue–Stieltjes measure $(\mathbf{R}^k, \mathcal{L}_F^k, \mu_F)$ has a countable basis.

8. Suppose I is a non-countable index set and for $\alpha \in I$, X_α denotes a 2-point space $\{0, 1\}$ with $\mu_\alpha\{0\} = \mu_\alpha\{1\} = \frac{1}{2}$. Form the product measure μ on the Cartesian product $\prod_{\alpha \in I} \{0, 1\} = \Omega$. Show that there is no countable basis for μ .

9. Show that, if μ has a countable basis, then $\mathcal{L}_p(\Omega, \mu)$, $1 \leq p < \infty$ is a separable space.

8.2 Orthogonal systems of functions

We now examine the part of the structure of $\mathcal{L}_2(\Omega, \mu)$ which is more intimately related to the inner product.

Linear dependence

In a linear space K , the finite set $\phi_1, \phi_2, \dots, \phi_n$ is said to be linearly dependent if there are real numbers c_i , not all zero, such that

$$\sum_{i=1}^n c_i \phi_i = 0. \quad (8.2.1)$$

On the other hand, if (8.2.1) implies that all the c_i are zero, then we say that ϕ_1, \dots, ϕ_n are *linearly independent*. A set $E \subset K$ is said to be linearly independent if each of its finite subsets is linearly independent. Note that, when $K = \mathcal{L}_2$ the relation (8.2.1) becomes

$$\sum_{i=1}^n c_i \phi_i(x) = 0 \text{ a.e.}$$

Closed linear span

In a normed linear space K , given a family $\{\phi_\alpha\}$, $\alpha \in I$ of points of K the set $M\{\phi_\alpha\}$ of all finite linear combinations

$$\sum_{k=1}^n c_{\alpha_k} \phi_{\alpha_k}, \quad c_{\alpha_k} \in \mathbf{R} \quad (8.2.2)$$

is called the span of $\{\phi_\alpha\}$, and its closure (in the norm topology) is called the *closed linear span* of $\{\phi_\alpha\}$ and denoted by $\bar{M}\{\phi_\alpha\}$. Thus this set \bar{M} consists of precisely those elements of K which can be approximated in norm by elements of the form (8.2.2).

Complete set

A family $\{\phi_\alpha\}$ ($\alpha \in I$) in a normed linear space K is said to form a *complete set* if its closed linear span is the whole space.

Suppose now that a normed linear space K is separable, so that there is a sequence $x_1, x_2, \dots, x_n, \dots$ of points dense in K . By omitting, in turn, any point in the sequence which can be expressed as a linear

combination of the previous ones we obtain a sequence g_1, g_2, \dots which is linearly independent, and has the same linear span as $\{x_n\}$. Since $\{x_n\}$ is dense, the closed linear span of $\{g_n\}$ must therefore be the whole space. Thus in any separable normed linear space there is a complete set of independent points which is either finite or enumerable. If there is a finite complete set (g_1, g_2, \dots, g_k) of independent points and K has an inner product, then we will see that K is isomorphic to Euclidean k -space. For $K = \mathcal{L}_2(\Omega, \mu)$, it is easy to see that K is finite-dimensional if μ is a discrete measure concentrated on a finite set of points, for then the indicator functions of these individual points will form a finite complete set. However, the interesting \mathcal{L}_2 -spaces are infinite-dimensional. Any $(\Omega, \mathcal{F}, \mu)$ for which \mathcal{F} contains infinitely many disjoint sets, each of finite positive measure, clearly generates an infinite-dimensional $\mathcal{L}_2(\Omega, \mu)$ since the indicator functions of this sequence form an independent set.

Orthogonal system

Two points x, y in a normed linear space K with an inner product are said to be orthogonal if $(x, y) = 0$. Any class $\{\phi_i\} (i \in I)$ of points of K which are different from zero and pairwise orthogonal is called an *orthogonal system*. A non-zero element $x \in K$ is said to be normalized if $\|x\| = 1$, i.e. $(x, x) = 1$. An orthogonal system of normalized points is said to be an *orthonormal system* in K . Thus $\{\phi_i\} (i \in I)$ is an orthonormal system if

$$(\phi_i, \phi_j) = \begin{cases} 1 & \text{for } i = j \in I, \\ 0 & \text{for } i \neq j. \end{cases}$$

Now any orthogonal system of points is certainly linearly independent for, if we take the inner product of (8.2.1) with ϕ_j we obtain $c_j(\phi_j, \phi_j) = 0$, so that $c_j = 0$. Further, if K is separable, any orthogonal system in K is countable. For any such system can be normalised to give an orthonormal system $\{\phi_i\} (i \in I)$, and then

$$\|\phi_i - \phi_j\| = \sqrt{2} \quad \text{for } i \neq j;$$

and, if $\{x_n\}$ is a countable dense set, we can find for each $i \in I$ an integer n_i such that $\|x_{n_i} - \phi_i\| < \frac{1}{2}$ and this gives

$$\|x_{n_i} - x_{n_j}\| > \sqrt{2} - 1 > 0 \quad \text{for } i \neq j, \quad \text{so that } n_i \neq n_j.$$

In the study of finite-dimensional normed linear spaces it is helpful to use a (finite) orthogonal normalized basis. In the general case, at least for K separable, it is also possible to find a complete orthonormal

sequence for K . This can be done by first obtaining a complete independent sequence and then orthogonalising it by the process of the next theorem.

Theorem 8.2 (*Gram-Schmidt orthogonalisation process*). *If K is a normed linear space with an inner product and $x_1, x_2, \dots, x_n, \dots$ is a linearly independent sequence in K , then there is an orthonormal sequence $y_1, y_2, \dots, y_n, \dots$ such that*

$$(i) \quad y_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n, \quad a_{nn} \neq 0;$$

$$(ii) \quad x_n = b_{n1}y_1 + b_{n2}y_2 + \dots + b_{nn}y_n, \quad b_{nn} \neq 0;$$

where a_{ij}, b_{ij} are real numbers. Further each y_i is uniquely determined (up to the sign) by these conditions.

Proof. If $y_1 = \alpha x_1$, then

$$(y_1, y_1) = \alpha^2(x_1, x_1) = 1$$

if α is suitably chosen. The conditions are therefore satisfied with $n = 1$ if $b_{11} = 1/a_{11} = \sqrt{(x_1, x_1)}$. (Note that the linear independence condition ensures that $\|x_1\| \neq 0$.) For $n > 1$, suppose y_1, y_2, \dots, y_{n-1} have been found to satisfy all the conditions. Then

$$x_n = b_{n1}y_1 + \dots + b_{n,n-1}y_{n-1} + z_n,$$

where $b_{ni} = (x_n, y_i)$ ($i = 1, 2, \dots, n-1$) so that $(z_n, y_i) = 0$ for $i < n$. We must have $(z_n, z_n) > 0$, since otherwise $z_n = 0$ and x_1, x_2, \dots, x_n would be linearly dependent. If we put

$$y_n = \frac{z_n}{\sqrt{(z_n, z_n)}}, \quad b_{nn} = \sqrt{(z_n, z_n)},$$

then $(y_n, y_n) = 1$, $(y_n, y_i) = 0$ for $i < n$ and (ii) is satisfied. We can then deduce (i) since $b_{nn} \neq 0$. By induction the method of orthogonalisation is established.

The uniqueness of the process (apart from sign) follows since the values of the constants are all determined except for the \pm sign in the square root which occurs at each stage. \blacksquare

Corollary. *If $\mathcal{L}_2(\Omega, \mu)$ is separable, then there is a complete orthonormal sequence in \mathcal{L}_2 .*

Proof. Start with a sequence $\{f_n\}$ which is dense in \mathcal{L}_2 , and replace it first by a sequence $\{g_n\}$ of linearly independent functions with the same closed linear span. If this is an infinite sequence, use the process of theorem 8.2 to obtain the orthonormal sequence $\{h_n\}$. It is clear that this sequence has the same closed linear span as $\{g_n\}$ so that it is a complete set. On the other hand, if \mathcal{L}_2 is finite dimensional, we will obtain

a finite set $\{g_1, g_2, \dots, g_n\}$ whose linear span is \mathcal{L}_2 . This finite set can be replaced by a finite orthonormal set using the process of theorem 8.2. **■**

In practice it is not always easy to prove that a given orthonormal sequence $\{\phi_1, \phi_2, \dots\}$ is complete. Various methods for proving completeness will be given in the next section.

Exercises 8.2

1. Suppose $\Omega = [0, 1)$, μ is Lebesgue measure, $f_0(x) \equiv 1$,

$$f_n(x) = \begin{cases} +1 & \text{if } 2^{n-1}x \equiv y \in [0, \frac{1}{2}) \pmod{1}, \\ -1 & \text{if } 2^{n-1}x \equiv y \in [\frac{1}{2}, 1) \pmod{1}. \end{cases}$$

The functions $f_n: [0, 1) \rightarrow \mathbf{R}$ are called the Rademacher functions. Show that they form an orthonormal sequence in $\mathcal{L}_2(\Omega, \mu)$.

2. For $\Omega = [-\pi, \pi]$, μ Lebesgue measure, show that the trigonometric functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos x, \quad \frac{1}{\sqrt{\pi}} \sin x, \dots, \quad \frac{1}{\sqrt{\pi}} \cos nx, \frac{1}{\sqrt{\pi}} \sin nx, \dots$$

form an orthonormal sequence in $\mathcal{L}_2(\Omega, \mu)$.

3. For $\Omega = [-1, 1]$, μ Lebesgue measure, show that the Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n \{(x^2 - 1)^n\}}{dx^n} \quad (n = 0, 1, 2, \dots),$$

are orthogonal in $\mathcal{L}_2(\Omega, \mu)$ and that the sequence $\sqrt{\frac{1}{2}(2n+1)} P_n(x)$ is orthonormal.

8.3 Riesz–Fischer theorem

This theorem is formulated and proved in Hilbert space. Since \mathcal{L}_2 -spaces will be shown to be realisations of Hilbert space, we will deduce the classical theorem about the Fourier expansion as a trigonometric series of a function in \mathcal{L}_2 as a special case.

Hilbert space

Suppose H is a normed linear space with an inner product which is complete in the topology of the norm; then H is said to be a Hilbert space. (Note that some older text-books require separability in addition.) If H is finite-dimensional, then theorem 8.2 allows us to choose a finite orthogonal basis e_1, e_2, \dots, e_n for H . It is then clear that

$$x = \sum_{k=1}^n c_k e_k, \quad c_k = (x, e_k) \quad (8.3.1)$$

is the unique expansion of $x \in H$ in terms of this basis. For separable infinite-dimensional H , we have seen that there is always an orthonormal sequence $\{e_i\}$ which forms a complete set. The main objective of the present section is to obtain the extension of (8.3.1) to this infinite-dimensional case. However, in formulating the results we will not assume that H is separable. It will turn out that an expansion in the form (8.3.1) is still possible, and that at most countably many terms in any such expansion will be non-zero. Before leaving the finite-dimensional H , we should observe that any Hilbert space of dimension n is isomorphic to Euclidean n -space \mathbf{R}^n with the usual scalar product. For (8.3.1) determines the point $(c_1, c_2, \dots, c_n) \in \mathbf{R}^n$ and defines a $(1, 1)$ mapping which then preserves the inner product.

Fourier coefficients

Given an orthonormal family (e_j) , $j \in J$ in a Hilbert space H , and any point $x \in H$, the real numbers

$$c_j = (x, e_j) \quad (j \in J),$$

are called the *Fourier coefficients* of x on the orthonormal family, and the series

$$\sum_{j \in J} c_j e_j$$

is called the *Fourier series* of x . (Note we have not yet said in what sense (if any) this series converges.)

The choice of the Fourier coefficients c_j can be justified as follows. If I is a finite subset of the index set J , re-label the indices $1, 2, \dots, n$ and consider the partial sum

$$s_n = \sum_{i=1}^n \alpha_i e_i \quad (n = 1, 2, \dots).$$

$$\begin{aligned} \text{Then} \quad \|s_n - x\|^2 &= \left\| x - \sum_{i=1}^n \alpha_i e_i, \quad x - \sum_{i=1}^n \alpha_i e_i \right\|^2, \\ &= \|x\|^2 - 2 \sum_{i=1}^n (x, \alpha_i e_i) + \sum_{i=1}^n \sum_{j=1}^n (\alpha_i e_i, \alpha_j e_j), \\ &= \|x\|^2 - 2 \sum_{i=1}^n \alpha_i c_i + \sum_{i=1}^n \alpha_i^2, \end{aligned}$$

$$\text{so that} \quad \|s_n - x\|^2 = \|x\|^2 - \sum_{i=1}^n c_i^2 + \sum_{i=1}^n (\alpha_i - c_i)^2. \quad (8.3.2)$$

Thus $\|s_n - x\|$ will be a minimum when all the terms of the last series in (8.3.2) are zero, and $\sum_{i=1}^n \alpha_i e_i$ is the best approximation (in norm)

to x when all the α_i are Fourier coefficients. This generalises the well-known geometrical theorem (for \mathbf{R}^n) which states that the length of the perpendicular from a point to a plane is smaller than the length of any other line joining the point to the plane: for $\left(x - \sum_{i=1}^k c_i e_i\right)$ is orthogonal to all linear combinations of the form $\sum_{i=1}^k \beta_i e_i$.

Bessel's inequality

We can make another deduction from (8.3.2) by noting that

$$\|s_n - x\|^2 \geq 0.$$

If we put $\alpha_i = c_i$ we obtain

$$\sum_{k=1}^n c_k^2 \leq \|x\|^2.$$

If we now define $\sum_{j \in J} c_j^2$ to be the supremum of $\sum_{j \in I} c_j^2$ for all finite subsets $I \subset J$ we find that

$$\sum_{j \in J} c_j^2 \leq \|x\|^2, \quad (8.3.3)$$

and this is known as *Bessel's inequality*. It follows as an immediate corollary that at most countably many Fourier coefficients of a point in H can be non-zero.

Theorem 8.3. *If $\{e_j\}$ ($j \in J$) is an orthonormal family in a Hilbert space, each of the following conditions is equivalent to $\{e_j\}$ being a complete family*

- (i) $\sum_{j \in J} c_j^2 = \|x\|^2$ for every $x \in H$ (Parseval),

where $\{c_j\}$ are the Fourier coefficients of x ;

- (ii) The finite partial sums $s_I = \sum_{k \in I} c_k e_k$ of the Fourier series of x converge to x in norm for all $x \in H$.

Note. For any arbitrary index set J we say that $\sum_{j \in J} x_j$ converges in norm to x if, given $\epsilon > 0$, there is a finite set K such that if I is finite and $K \subset I \subset J$ then

$$\left\| \sum_{j \in I} x_j - x \right\| < \epsilon.$$

It is easy to check that, when J is countable and the x_j are real so that we have a real series this notion reduces to the usual definition of absolute convergence.

Proof. The conditions (i) and (ii) are clearly equivalent by (8.3.2). Now suppose that (ii) is satisfied. Then any x can be approximated in norm by a finite sum s_n which is a linear combination of e_1, e_2, \dots, e_n .

Hence, each x is in the closed linear span of $\{e_j\}$ and the sequence must form a complete system.

Conversely suppose $\{e_j\}$ is a complete system. Then given $\epsilon > 0$, $x \in H$, there is a finite sum $y = \sum_{i=1}^N \alpha_i e_i$ for which $\|x - y\| < \epsilon$. But, if S_N is the corresponding partial Fourier sum, we know

$$\|x - y\|^2 \geq \|x - s_N\|^2,$$

so that, by (8.3.2)

$$\sum_{i=1}^N c_i^2 \geq \|x\|^2 - \epsilon.$$

Since ϵ is arbitrary, we can combine this with (8.3.3) to give

$$\sum_{j \in J} c_j^2 = \|x\|^2. \blacksquare$$

From (8.3.3) we know that a given set $\{\beta_j\} (j \in J)$ of real numbers can only be the Fourier coefficients of a point in H if $\sum_{j \in J} \beta_j^2$ converges.

It turns out that this condition is sufficient as well as necessary.

Theorem 8.4 (Riesz-Fischer). *Let $\{e_j\} (j \in J)$ be any orthonormal system (not necessarily complete) in a Hilbert space H , and let $\{\beta_j\} j \in J$ be any set of real numbers such that $\sum_{j \in J} \beta_j^2$ converges. Then there is a point $x \in H$ with Fourier coefficients $\beta_j = (x, e_j)$ such that the finite partial sums $s_I = \sum_{i \in I} \beta_i e_i$ converge to x in norm.*

Proof. Since $\sum_{j \in J} \beta_j^2$ converges the set of j for which $\beta_j \neq 0$ is countable and we may suppose then that these indices are renamed

$$\beta_1, \beta_2, \dots, \beta_n, \dots$$

(it may be only a finite set). Put

$$s_n = \sum_{i=1}^n \beta_i e_i.$$

Then

$$\|s_{n+p} - s_n\|^2 = \sum_{i=n+1}^{n+p} \beta_i^2.$$

Since $\sum \beta_i^2$ converges, it follows that $\{s_n\}$ is a Cauchy sequence in norm. Since H is complete, there must be an $x \in H$ such that $s_n \rightarrow x$ in norm. Further

$$\begin{aligned} (x, e_i) &= (s_n, e_i) + (x - s_n, e_i) \\ &= \beta_i + (x - s_n, e_i) \quad \text{for } n \geq i \end{aligned}$$

and, by exercise 8.1 (1)

$$|(x - s_n, e_i)| \leq \|e_i\| \cdot \|x - s_n\| = \|x - s_n\| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since (x, e_i) is independent of n , we have $\beta_i = (x, e_i)$ for all i . \blacksquare

Corollary. An orthonormal system $\{e_j\}$ ($j \in J$) in a Hilbert space H forms a complete system if and only if the only point $x \in H$ which is orthogonal to all the $\{e_j\}$ is the point $x = 0$.

Proof. Suppose $\{e_j\}$ is a complete orthonormal set and $(x, e_j) = 0$ for all j . Then all the Fourier coefficients of x are zero and so

$$\|x\|^2 = \sum_{j \in J} c_j^2 = 0.$$

Conversely suppose $\{e_j\}$ is not a complete system. Then by theorem 8.3 (i), there is a point $x \in H$ with

$$\|x\|^2 > \sum_{j \in J} c_j^2, \quad \text{where } c_j = (x, e_j).$$

By theorem 8.4, there is a $y \in H$ such that

$$c_j = (y, e_j), \quad \|y\|^2 = \sum_{j \in J} c_j^2.$$

Then the point $(x - y) \in H$ is orthogonal to all the e_j . But $\|x - y\| \neq 0$, since $\|x\| > \|y\|$. **]**

Remark. If the Hilbert space is separable we have already observed that any orthonormal system is countable. For a separable Hilbert space, therefore, it is natural to state and prove theorem 8.4 and corollary in terms of an arbitrary orthonormal sequence. No essential modifications to the proof are needed.

The space l_2

The class of all infinite sequences $c_1, c_2, \dots, c_n, \dots$ of real numbers for which $\sum_{k=1}^{\infty} c_k^2$ converges is called the space l_2 . By using the discrete version of theorem 7.7 one can check that if $\{c_i\}, \{d_i\} \in l_2$,

$$(c, d) = \sum_{i=1}^{\infty} c_i d_i \tag{8.3.4}$$

converges and defines an inner product. Alternatively, if Ω is the set of positive integers, μ is the counting measure, $c_i = f(i)$, then $f \in \mathcal{L}_2(\Omega, \mu)$ if and only if $\sum_{k=1}^{\infty} c_k^2 < \infty$, and (8.3.4) is the usual inner product $(f, g) = \int fg d\mu$ in \mathcal{L}_2 . Completeness and separability for l_2 can be proved directly, or they can be deduced from the corresponding properties of $\mathcal{L}_2(\Omega, \mu)$. Thus l_2 is a separable Hilbert space according to our definition—and historically l_2 was the space first considered in detail.

The justification for our abstract definition of Hilbert space is contained in the following theorem.

Theorem 8.5. *An n -dimensional Hilbert space is isomorphic to \mathbf{R}^n . Any separable infinite-dimensional Hilbert space is isomorphic to l_2 .*

Proof. The finite-dimensional case was considered earlier. If H is separable and infinite-dimensional we can select a complete orthonormal sequence $\{e_n\}$ and obtain the Fourier coefficients $\{c_n\}$ of a point $x \in H$. Then since $x \in H \Rightarrow \sum c_n^2 < \infty$, this defines a mapping from H to l_2 . Conversely, every sequence in l_2 determines a point in H with these Fourier coefficients, by theorem 8.4. There is only one such function by the corollary to theorem 8.4. Thus to prove that we have an isomorphism it is sufficient to prove that the linear structure and inner product are preserved by the mapping. Suppose

$$x^{(1)}, x^{(2)} \in H$$

correspond to $\{c_i^{(1)}\}, \{c_i^{(2)}\} \in l_2$. Then it is immediate that

$$\alpha x^{(1)} \leftrightarrow \{\alpha c_i^{(1)}\} \quad (\alpha \in \mathbf{R});$$

$$x^{(1)} + x^{(2)} \leftrightarrow \{(c_i^{(1)} + c_i^{(2)})\};$$

$$\|x^{(1)}\|^2 = \sum (c_i^{(1)})^2, \quad \|x^{(2)}\|^2 = \sum (c_i^{(2)})^2;$$

$$\begin{aligned} \|x^{(1)} + x^{(2)}\|^2 &= \|x^{(1)}\|^2 + 2(x^{(1)}, x^{(2)}) + \|x^{(2)}\|^2 \\ &= \sum (c_i^{(1)})^2 + 2\sum c_i^{(1)} c_i^{(2)} + \sum (c_i^{(2)})^2, \end{aligned}$$

so that $(x^{(1)}, x^{(2)}) = \sum c_i^{(1)} c_i^{(2)}$. \blacksquare

Corollary. *If $(\Omega, \mathcal{F}, \mu)$ is any measure space with a countable basis, then $\mathcal{L}_2(\Omega, \mu)$ is either finite-dimensional, when it is isomorphic to a Euclidean space \mathbf{R}^n , or it is infinite-dimensional when it is isomorphic to l_2 . If $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ are such that $\mathcal{L}_2(\Omega_1, \mu_1)$ and $\mathcal{L}_2(\Omega_2, \mu_2)$ are both infinite-dimensional and separable, then $\mathcal{L}_2(\Omega_1, \mu_1)$ and $\mathcal{L}_2(\Omega_2, \mu_2)$ are isomorphic.*

The theorems of this section were first obtained for trigonometric series of functions f in $\mathcal{L}_2([-\pi, \pi], \mu)$ where μ is Lebesgue measure. In order to obtain these special theorems one only has to prove that the functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos x, \quad \frac{1}{\sqrt{\pi}} \sin x, \quad \dots, \quad \frac{1}{\sqrt{\pi}} \cos nx, \quad \frac{1}{\sqrt{\pi}} \sin nx, \dots$$

form a complete orthonormal sequence in this \mathcal{L}_2 -space. The steps necessary for this proof are contained in exercise 8.3 (2).

Exercises 8.3

1. Prove that a series $\sum_{n=1}^{\infty} a_n$ of real terms converges absolutely to s if and only if, for each $\epsilon > 0$ there is a finite set $I \subset \mathbb{Z}$ such that for every finite K with $I \subset K \subset \mathbb{Z}$ we have

$$|s - \sum_{n \in K} a_n| < \epsilon.$$

2. If $\Omega = [-\pi, \pi]$, μ is Lebesgue measure, $f \in \mathcal{L}_2(\Omega, \mu)$

$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos mx \, dx \quad (m = 0, 1, 2, \dots);$$

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin mx \, dx \quad (m = 1, 2, \dots),$$

then the a_m, b_m are the classical Fourier coefficients of f . Prove:

(i) $\frac{1}{2}a_0^2 + \sum_{m=1}^{\infty} (a_m^2 + b_m^2) \leq \frac{1}{\pi} \int_{-\pi}^{\pi} \{f(x)\}^2 \, dx;$

(ii) if $\{a_n\}, \{b_n\}$ are such that

$$\frac{1}{2}a_0^2 + \sum_1^{\infty} (a_m^2 + b_m^2) < \infty,$$

then there is a function $f \in \mathcal{L}_2$ for which these are the Fourier coefficients and such that

$$s_n(x) = \frac{1}{2}a_0 + \sum_{m=1}^n (a_m \cos mx + b_m \sin mx) \rightarrow f$$

in second mean;

(iii) if $\{a_n\}, \{b_n\}$ are the Fourier coefficients of f in the above sense

$$s_n(x) = \frac{1}{2}a_0 + \sum_{m=1}^n (a_m \cos mx + b_m \sin mx),$$

$$\sigma_n(x) = \frac{1}{n+1} [s_0(x) + s_1(x) + \dots + s_n(x)],$$

then $\sigma_n(x) \rightarrow f(x)$ in \mathcal{L}_2 norm (and in fact $\sigma_n \rightarrow f$ a.e.);

(iv) $\frac{1}{\pi} \int_{-\pi}^{\pi} \sigma_n^2(x) \, d\mu = \frac{1}{2}a_0^2 + \sum_1^n (a_k^2 + b_k^2) \left(1 - \frac{k}{n+1}\right)^2$

$$\leq \frac{1}{2}a_0^2 + \sum_1^n (a_k^2 + b_k^2) \leq \frac{1}{2}a_0^2 + \sum_1^{\infty} (a_k^2 + b_k^2);$$

(v) since $\int_{-\pi}^{\pi} \sigma_n^2(x) \, d\mu \rightarrow \int_{-\pi}^{\pi} f^2(x) \, d\mu$ we have for $f \in \mathcal{L}_2$

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \{f(x)\}^2 \, d\mu = \frac{1}{2}a_0^2 + \sum_1^{\infty} (a_k^2 + b_k^2);$$

(vi) the trigonometric system of exercise 8.2 (2) is complete.

3. If $\{\phi_n\}$ is any orthonormal sequence in a Hilbert space H , then for $x \in H$, $(x, \phi_n) \rightarrow 0$ as $n \rightarrow \infty$.

4. If $f \in \mathcal{L}_2([-\pi, \pi], \mu)$ then, as $n \rightarrow \infty$,

$$\int_{-\pi}^{\pi} f(x) \sin nx \, dx \rightarrow 0, \quad \int_{-\pi}^{\pi} f(x) \cos nx \, dx \rightarrow 0.$$

5. If \mathbf{l}_p is the space of real sequences $\{x_i\}$ for which

$$\sum_{i=1}^{\infty} |x_i|^p < \infty \quad \text{for } 1 \leq p < \infty,$$

show that \mathbf{l}_p is separable. \mathbf{l}_{∞} is the space of bounded real sequences with $\|x\| = \sup |x_i|$. By considering sequences of 0's and 1's show that \mathbf{l}_{∞} is not separable. Deduce that \mathcal{L}_{∞} is not separable if Ω has infinite but σ -finite measure.

8.4* Space of linear functionals

We start by defining a more general type of normed linear space.

Banach space

Any normed linear space over the reals which is complete in the topology determined by the norm is called a (real) Banach space.

We saw already that the \mathcal{L}_p ($1 \leq p \leq +\infty$) spaces are normed linear spaces and that each of them is complete in the norm topology so that each \mathcal{L}_p is a Banach space. Euclidean n -space \mathbf{R}^n with the usual metric provides a simpler example of a Banach space. $C[a, b]$, the space of continuous functions on a finite closed interval, with $\|f\| = \sup |f(x)|$, can also be easily seen to be a Banach space.

From our definition of Hilbert space, it follows that any Banach space will be a Hilbert space provided there is an inner product defined satisfying $\|f\|^2 = (f, f)$. The question immediately arises as to whether or not all Banach spaces are Hilbert spaces; or is it always possible to define an inner product in a Banach space? We can settle this as follows. If there is to be an inner product, then

$$\|f+g\|^2 = (f+g, f+g) = (f, f) + 2(f, g) + (g, g),$$

$$\|f-g\|^2 = (f-g, f-g) = (f, f) - 2(f, g) + (g, g),$$

so that on adding

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2. \quad (8.4.1)$$

Thus, a relation (8.4.1) for all f, g in the space is a necessary condition for the Banach space to have an inner product. One can also easily check that, if (8.4.1) is always satisfied, then

$$(f, g) = \frac{1}{2}(\|f+g\|^2 - \|f\|^2 - \|g\|^2) \quad (8.4.2)$$

satisfies all the conditions for an inner product, so that any Banach space which satisfies (8.4.1) is a Hilbert space if we define the inner product by (8.4.2). We can think of the condition (8.4.1) as a generalisation of the Euclidean theorem that in any parallelogram the sum of the squares on the diagonals is twice the sum of the squares on two adjacent sides. If this theorem is not valid in the Banach space K , then it is not possible to define an inner product on K . This allows us to show that \mathcal{L}_p is not a Hilbert space for $p \neq 2$ —see exercise 8.4 (2).

Linear functional

Given a linear space K over the reals a function $T: K \rightarrow \mathbf{R}$ is called a linear functional if, for all $x_1, x_2 \in K$, $\alpha, \beta \in \mathbf{R}$,

$$T(\alpha x_1 + \beta x_2) = \alpha T(x_1) + \beta T(x_2).$$

If K is a normed linear space, then T is continuous at $x_0 \in K$ if, given $\epsilon > 0$, there is a $\delta > 0$ with

$$\|x - x_0\| < \delta \Rightarrow |T(x) - T(x_0)| < \epsilon.$$

For a normed K the functional is said to be bounded if there is a real constant C such that

$$|T(x)| \leq C\|x\| \quad \text{for all } x \in K.$$

It is immediate that a linear functional on a normed linear space is continuous everywhere if it is continuous at any one point. The connexion between continuity and boundedness is not quite so obvious.

Lemma. *A linear functional T on a normed linear space K is continuous if and only if it is bounded.*

Proof. Suppose first that T is bounded, then given $x_1 \in K$, $\epsilon > 0$, put $\delta = \epsilon \cdot C^{-1}$ and we have

$$|T(x) - T(x_1)| = |T(x - x_1)| \leq C\|x - x_1\| < \epsilon$$

if $\|x - x_1\| < \delta$. Conversely, if T is continuous at $0 \in K$ we can choose $B > 0$ such that

$$|T(x)| \leq 1 \quad \text{for } \|x\| \leq B.$$

Then for $x \in K$, $\left\| \frac{B}{\|x\|} x \right\| = B$, so that

$$|T(x)| = \frac{\|x\|}{B} T\left(\frac{Bx}{\|x\|}\right) \leq \frac{1}{B} \|x\|,$$

and T is bounded. \blacksquare

Norm of a bounded functional

If T is a bounded linear functional on a normed linear space K , the smallest number C satisfying $|T(x)| \leq C\|x\|$ for all $x \in K$ is called the norm of T and we denote it by $\|T\|$. Because of linearity,

$$\|T\| = \sup_{x \neq 0} \frac{|T(x)|}{\|x\|} = \sup_{\|x\|=1} |T(x)|.$$

If T_1, T_2 are two linear functionals on a linear space K , $\alpha, \beta \in \mathbf{R}$ then

$$(\alpha T_1 + \beta T_2)(x) = \alpha T_1(x) + \beta T_2(x)$$

is also a linear functional on K ; and the set of all linear functionals on K is a linear space. We can say more if K is normed.

Lemma. *If K^* denotes the set of bounded linear functionals on a normed linear space K , then K^* is a Banach space.*

Proof. $T \in K^*$, $\|T\| = 0$ implies that $T(x) = 0$ for all $x \in K$ which means that T is the null transformation,

$$\begin{aligned} \|T_1 + T_2\| &= \sup_{\|x\|=1} |T_1(x) + T_2(x)| \leq \sup_{\|x\|=1} |T_1(x)| + \sup_{\|x\|=1} |T_2(x)| \\ &= \|T_1\| + \|T_2\|, \end{aligned}$$

$$\text{and} \quad \|\alpha T\| = \sup_{\|x\|=1} |\alpha T(x)| = |\alpha| \sup_{\|x\|=1} |T(x)| = |\alpha| \cdot \|T\|.$$

This shows that K^* is a normed linear space with the norm

$$\|T\| = \sup_{\|x\|=1} |T(x)|.$$

It remains to show that K^* is complete. Suppose $\{T_n\}$ is a sequence in K^* such that

$$\|T_m - T_n\| \rightarrow 0 \quad \text{as } m, n \rightarrow \infty.$$

Then, for each $x \in K$, $|T_m(x) - T_n(x)| \rightarrow 0$ as $m, n \rightarrow \infty$. The completeness of \mathbf{R} now implies that there is a real number

$$y = T(x) = \lim_{n \rightarrow \infty} T_n(x).$$

Now T is clearly a linear transformation and since

$$|\|T_m\| - \|T_n\|| \leq \|T_m - T_n\|,$$

the real sequence $\{\|T_n\|\}$ must be bounded, say by C . Then

$$|T_n(x)| \leq C\|x\| \quad \text{for all } x \in K$$

and all integers n , so that $|T(x)| \leq C\|x\|$ and T is a bounded linear functional, that is $T \in K^*$. Given $\epsilon > 0$, there is an integer $N = N(\epsilon)$ such that

$$|T_m(x) - T_n(x)| < \epsilon \quad \text{for } \|x\| = 1, \quad x \in K, \quad m, n \geq N.$$

If we let $m \rightarrow \infty$, then

$$|T(x) - T_n(x)| \leq \epsilon \quad \text{for } \|x\| = 1 \quad (n \geq N),$$

so that $\|T - T_n\| \rightarrow 0$ as $n \rightarrow \infty$, and K^* is complete. \blacksquare

Conjugate space

For any normed linear space K (in particular if K is a Banach space), the Banach space K^* of bounded linear functionals on K is called the conjugate space (or dual space) of K .

Linear subspace

A set H contained in a linear space K such that H is itself a linear space is called a linear subspace of K .

If K contains a point $x \neq 0$, it is clear that the set of all points αx , $\alpha \in \mathbf{R}$ is a linear subspace H of K . Then, if we put,

$$T(\alpha x) = \alpha \quad \text{for all } \alpha x \in H$$

it is immediate that $T \neq 0$ is a linear functional on H . However, it is not immediately obvious that the set of bounded linear functionals defined on the whole of K contains any $T \neq 0$. The existence of such a non-trivial T will follow if we can prove that linear functionals defined on a subspace can always be extended to the whole of K .

Theorem 8.6 (Hahn-Banach extension). *Suppose K is a linear subspace of a linear space H . Then any bounded linear functional on K can be extended to a bounded linear functional on H with the same norm.*

Proof. Suppose $f: K \rightarrow \mathbf{R}$ is the given functional and

$$a = \sup_{x \in K} |f(x)| / \|x\|.$$

Then $|f(x)| \leq a\|x\|$ for all x in K . Consider the class \mathcal{C} of all linear functionals T defined on spaces J such that (i) $K \subset J \subset H$; (ii) $T(x) = f(x)$ for $x \in K$; (iii) $T(x) \leq a\|x\|$ for $x \in J$. We can partially order \mathcal{C} by putting $g_1 < g_2$ if g_1 is defined on J_1 , g_2 on J_2 , $K \subset J_1 \subset J_2 \subset H$ and

$$g_1(x) = g_2(x) = f(x) \quad \text{for } x \in K, \quad g_1(x) = g_2(x) \quad \text{for } x \in J_1.$$

By Zorn's lemma (§1.6) we can find a maximal element in this partial ordering. This must be an extension T of f defined on a subspace $J \subset H$ such that no further extension to a larger subspace is possible. It is clearly sufficient to show that, for this maximal $T \in \mathcal{C}$, we must have $J = H$.

Suppose not, then there is a point $z \in H - J$. We will obtain a contradiction by showing that T can be extended to the linear space J_z consisting of all points of the form $j + \alpha z$, $j \in J$, $\alpha \in \mathbf{R}$. Note first that,

since $z \notin J$, the representation $(j + \alpha z)$ is unique. The extension to J_z will therefore be determined by its value at z . Now if $x, y \in J$,

$$T(x) - T(y) = T(x - y) \leq a\|x - y\| \leq a\|x + z\| + a\|-y - z\|$$

so that $-a\|-y - z\| - T(y) \leq a\|x + z\| - T(x)$.

$$\text{Hence } \sup_{y \in J} [-a\|-y - z\| - T(y)] \leq \inf_{x \in J} [a\|x + z\| - T(x)].$$

Let t be any real number satisfying

$$\sup_{y \in J} [-a\|-y - z\| - T(y)] \leq t \leq \inf_{x \in J} [a\|x + z\| - T(x)], \quad (8.4.3)$$

and put $\bar{T}(z) = t$: this implies

$$\bar{T}(k + \alpha z) = T(k) + \alpha t \quad \text{for } k \in J.$$

Now put $y = x/\alpha$ in (8.4.3) and let $w = x + \alpha z$:

$$-a\left\|-\frac{w}{\alpha}\right\| - T\left(\frac{x}{\alpha}\right) \leq t \leq a\left\|\frac{w}{\alpha}\right\| - T\left(\frac{x}{\alpha}\right).$$

If $\alpha > 0$ multiply the right-hand inequality by α , while if $\alpha < 0$ multiply the left-hand inequality by α . Both cases give

$$a\|w\| - T(x) \geq \alpha t$$

so that

$$\bar{T}(w) \leq a\|w\|$$

and $\bar{T} \in \mathcal{C}$. Since J is a proper subset of J_z this establishes the existence of the extension. To see that the extension F has the same norm as f we need only note that $|F(x)| \leq a\|x\|$ for all x in H so that

$$\|F\| \leq a = \|f\|;$$

and

$$\|F\| = \sup_{\substack{x \in H \\ \|x\|=1}} |F(x)| \geq \sup_{\substack{x \in K \\ \|x\|=1}} |f(x)| = \|f\|. \quad \blacksquare$$

Remark. In the above theorem, the only property of the norm which we used was that

$$\|x + y\| \leq \|x\| + \|y\| \quad \text{for all } x, y \in H.$$

It is possible to state and prove the extension theorem in terms of any subadditive bounding functional. This gives

Theorem 8.6A (Hahn-Banach extension). Suppose K is a linear subspace of a linear space H , p is a subadditive functional on H such that $p(ax) = ap(x)$ for $a \geq 0$, $x \in H$; and f is a linear functional on K such that $f(x) \leq p(x)$ for all $x \in K$. Then there is a linear functional $\bar{f}: H \rightarrow \mathbf{R}$ such that

$$\bar{f}(x) = f(x) \quad \text{for } x \in K, \quad \bar{f}(x) \leq p(x) \quad \text{for } x \in H.$$

Exercises 8.4

1. If $(\Omega, \mathcal{F}, \mu)$ is such that there are two sets $E_1, E_2 \in \mathcal{F}$ with $\mu(E_1), \mu(E_2)$ positive and finite, show by considering the indicator functions of E_1, E_2 that $\mathcal{L}_p(\Omega, \mu)$ does not satisfy (8.4.1) if $p \neq 2$; and therefore is not a Hilbert space.

2. Prove that m , the set of bounded real sequences $\{x_i\}$, is a Banach space with $\|x\| = \sup_i |x_i|$.

3. Suppose K is a Banach space, K^* is its dual, and K^{**} is the dual of K^* . Prove:

(i) if x is a fixed element in K , $X(f) = f(x)$ for $f \in K^*$ defines a linear functional on K^* ;

(ii) for the above function $\|X\| = \|x\|$, so that $T(x) = X$ is a norm preserving map from K to K^{**} ;

(iii) this map T preserves the linear structure;

(iv) The set of elements X of K^{**} such that $X(f) = f(x)$ for some $x \in K$ forms a closed linear subspace of K^{**} .

4. If K is a linear subspace of a Banach space H show that a point y of H is in the closure of K if and only if $f(y) = 0$ for every linear functional $f \in H^*$ which vanishes on K .

5. In §4.4 we showed that it is not possible to define a *measure* on $[0, 1)$ which is defined for all subsets and invariant for translations (mod 1). The following steps will show that we can define such a *finitely additive* set function on all subsets of $[0, 1)$ with $\nu[0, 1) = 1$ and $\nu(E) = |E|$ when E is Lebesgue measurable.

(i) Let H be the set of all bounded functions $f: [0, 1) \rightarrow \mathbf{R}$ which are extended to be periodic in the whole of \mathbf{R} by $f(x+1) = f(x)$. Prove H is a linear space.

(ii) Put
$$M(f; \alpha_1, \alpha_2, \dots, \alpha_n) = \sup_{x \in \mathbf{R}} \frac{1}{n} \sum_{i=1}^n f(x + \alpha_i),$$

$p(f) = \inf M(f; \alpha_1, \dots, \alpha_n)$ for all such finite sequences of real α_i . Prove that p is subadditive and $p(af) = ap(f)$ for $a \geq 0$.

(iii) If $f: [0, 1) \rightarrow \mathbf{R}$ is bounded and Lebesgue measurable, show that the Lebesgue integral $\mathcal{J}(f) \leq M(f; \alpha_1, \dots, \alpha_n)$.

(iv) Show that the set of bounded measurable $f: [0, 1) \rightarrow \mathbf{R}$ is a linear subspace of H , and $\mathcal{J}(f)$ is a bounded linear functional on this subspace.

(v) Use theorem 8.6A to extend \mathcal{J} to a linear functional F defined on all of H .

(vi) Show that $F\{f(x + x_0)\} = F\{f(x)\}$ for all $x_0 \in \mathbf{R}$.

(vii) By considering indicator functions of subsets of $[0, 1)$, put $\nu(E) = F(\chi_E)$ for $E \subset [0, 1)$.

Prove $\nu(E_1 \cup E_2) = \nu(E_1) + \nu(E_2)$ if E_1, E_2 are disjoint,

$$\nu(E) = |E| \quad \text{if } E \text{ is Lebesgue measurable,}$$

$$\nu(E) \quad \text{is invariant under translations.}$$

6. Similar arguments to those used in (5) above can be applied to the linear space V of bounded real functions $f: [0, +\infty) \rightarrow \mathbf{R}$. Show that there exists a bounded linear functional $\text{Lim } f(x)$ on V such that, for $a, b \in \mathbf{R}$.

$$(i) \quad \text{Lim } \{af(x) + bg(x)\} = a \text{Lim } f(x) + b \text{Lim } g(x);$$

$$(ii) \quad f(x) \geq 0 \text{ in } [0, \infty) \Rightarrow \text{Lim } f(x) \geq 0;$$

$$(iii) \quad \text{Lim } f(x + x_0) = \text{Lim } f(x) \text{ for any } x_0 \geq 0;$$

$$(iv) \quad \text{Lim } f(x) = \lim_{x \rightarrow +\infty} f(x) \text{ if this exists.}$$

Deduce a corresponding result for the space m of bounded real sequences.

8.5* The space conjugate to \mathcal{L}_p

We have seen that \mathcal{L}_p ($1 \leq p \leq +\infty$) is a Banach space and, *a fortiori*, a normed linear space. It follows from the lemma on p. 211 that the space of bounded linear functionals on \mathcal{L}_p is also a Banach space. The object of the present section is to identify these conjugate spaces—at least up to an isomorphism.

Theorem 8.7. Suppose $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space and \mathcal{L}_p , $1 \leq p < \infty$ is the linear space of \mathcal{F} -measurable functions $f: \Omega \rightarrow \mathbf{R}^*$ whose p th power is integrable, with the usual norm

$$\|f\| = \left\{ \int |f|^p d\mu \right\}^{1/p}.$$

Let $1/p + 1/q = 1$ (if $p = 1, q = \infty$). Then

$$(i) \quad \text{for each } g \in \mathcal{L}_q, \quad F(f) = \int fg d\mu$$

defines a linear functional on \mathcal{L}_p ;

(ii) given any bounded linear functional F on \mathcal{L}_p , there is a $g \in \mathcal{L}_q$ such that

$$F(f) = \int fg d\mu,$$

$$\text{and in this case} \quad \|F\| = \left\{ \int |g|^q d\mu \right\}^{1/q} \quad \text{if } p > 1,$$

$$= \text{ess sup } |g| \quad \text{if } p = 1.$$

Proof. (i) This follows immediately from theorem 7.7 and the linear properties of the integral; for

$$|F(f)| \leq \left\{ \int |f|^p d\mu \right\}^{1/p} \left\{ \int |g|^q d\mu \right\}^{1/q}.$$

(ii) Suppose first that $\mu(\Omega) < \infty$ and F is a linear functional on \mathcal{L}_p . For any measurable $E \subset \Omega$ put

$$\sigma(E) = F(\chi_E),$$

where χ_E is the indicator function of E . The linearity of F implies immediately that σ is finitely additive. Now suppose $E = \bigcup_{i=1}^{\infty} E_i$,

E_i disjoint. Then $\mu\left(\bigcup_{i=1}^N E_i\right) \rightarrow \mu(E)$ as $N \rightarrow \infty$, so that in \mathcal{L}_p ,

$$\|\chi_{Q_N} - \chi_E\| \rightarrow 0, \quad \text{where } Q_N = \bigcup_{i=1}^N E_i.$$

Since F is continuous, we must have

$$\sum_{i=1}^{\infty} \sigma(E_i) = \lim_{N \rightarrow \infty} \sum_{i=1}^N \sigma(E_i) = \sigma(E)$$

so that σ is completely additive on \mathcal{F} . Further $\mu(E) = 0 \Rightarrow \sigma(E) = 0$, so that σ is absolutely continuous with respect to μ . By theorem 6.7 it now follows that there exists a function g which is integrable, such that

$$\begin{aligned} F(\chi_E) = \sigma(E) &= \int_E g d\mu \quad \text{for all } E \in \mathcal{F}, \\ &= \int \chi_E g d\mu. \end{aligned}$$

This gives the required representation for F on the class of indicator functions of measurable sets. We must prove that $g \in \mathcal{L}_q$, and that the representation is valid on the whole of \mathcal{L}_p .

It is clear from linearity that the representation is valid for \mathcal{F} -simple functions. If $f_0 \in \mathcal{L}_p$, $f_0 \geq 0$, we can find a sequence f_n of simple functions which increases monotonely to f_0 . Then by theorem 7.6, $f_n \rightarrow f_0$ in p th mean, and by the continuity of F

$$F(f_0) = \lim_{n \rightarrow \infty} F(f_n) = \lim_{n \rightarrow \infty} \int f_n g d\mu = \int f_0 g d\mu$$

on applying the monotone convergence theorem to $\{f_n g_+\}$ and $\{f_n g_-\}$ separately. The restriction $f_0 \geq 0$ can be removed by considering f_+ and f_- separately, so that

$$F(f) = \int f g d\mu \quad \text{for all } f \in \mathcal{L}_p.$$

Now suppose $p > 1$, and $g(t)$ has been obtained by the above process. Put

$$g_n(t) = \begin{cases} |g(t)|^{q-1} \operatorname{sign} g(t) & \text{if } |g(t)|^{q-1} \leq n, \\ n \operatorname{sign} g(t) & \text{if } |g(t)|^{q-1} > n. \end{cases}$$

Then each g_n is bounded and measurable, and is therefore in \mathcal{L}_p . Hence

$$|F(g_n)| = \left| \int g_n g d\mu \right| \leq \|F\| \|g_n\| = \|F\| \left(\int |g_n|^p d\mu \right)^{1/p}.$$

But
$$g_n g = |g_n| |g| \geq |g_n| |g_n|^{1/q-1} = |g_n|^p,$$

so that
$$\int |g_n|^p d\mu \leq \|F\| \left(\int |g_n|^p d\mu \right)^{1/p},$$

and
$$\left(\int |g_n|^p d\mu \right)^{1/q} \leq \|F\|.$$

But
$$|g_n|^p \rightarrow |g|^q \text{ a.e.}$$

so that, by theorem 5.5
$$\left(\int |g|^q d\mu \right)^{1/q} \leq \|F\|. \quad (8.5.1)$$

Before going on to prove equality in (8.5.1), let us now remove the restriction $\mu(\Omega) < \infty$. Suppose $\mu(\Omega) = \infty$, so that there is a sequence $\{Q_n\}$ of disjoint measurable sets with

$$\Omega = \bigcup_{i=1}^{\infty} Q_i, \quad \mu(Q_n) < \infty \quad \text{all } n.$$

We can apply the above argument to each of the spaces $(Q_n, \mathcal{F}_n, \mu)$ where $\mathcal{F}_n = \mathcal{F} \cap Q_n$. By the uniqueness of the derivative g in the Radon-Nikodym theorem, if $f \in \mathcal{L}_p(\Omega, \mu)$ and vanishes outside $R_N = \bigcup_{i=1}^N Q_i$,
$$F(f) = \int f g d\mu.$$

But if $f \geq 0$, we can apply the monotone convergence theorem to each of $\{f_n g_+\}$, $\{f_n g_-\}$ where $f_n = f \chi_{R_n}$ to obtain this representation by using the continuity of F . The final step is to use $f = f_+ - f_-$ so that the representation is valid on all of \mathcal{L}_p . Further (8.5.1) follows since it is true for the integral over each R_n .

Now by Hölder's inequality (theorem 7.7) we have

$$|F(f)| \leq \|f\| \left\{ \int |g|^q d\mu \right\}^{1/q},$$

so that
$$\|F\| = \left\{ \int |g|^q d\mu \right\}^{1/q} \quad \text{using (8.5.1).}$$

We need to modify the argument in the case $p = 1$, assuming that g has been defined as before as the Radon-Nikodym derivative of σ . For any $t > 0$, let E be a set such that $0 < \mu(E) < \infty$ and $|g(x)| > t$ for $x \in E$. Put $f(x) = \chi_E \operatorname{sign} g(x)$ and we have

$$F(f) \geq t\mu(E), \quad \|f\| = \mu(E)$$

so $\|F\| \geq t$. Since such a set E can be found for any $t < \text{ess sup } |g|$ we must have

$$\|F\| \geq \text{ess sup } |g|.$$

But
$$|F(f)| = \left| \int fg d\mu \right| \leq (\text{ess sup } |g|) \|f\|,$$

so that $\|F\| \leq \text{ess sup } |g|$.]

Corollary. *If H is a separable Hilbert space:*

(i) *for any fixed $h \in H$, the inner product $F(f) = (f, h)$ defines a bounded linear functional;*

(ii) *for any bounded linear functional F on H , there is an $h \in H$ such that $F(f) = (f, h)$ for all $f \in H$: further $\|F\| = \|h\|$.*

Proof. Choose a measure space $(\Omega, \mathcal{F}, \mu)$ such that $\mathcal{L}_2(\Omega, \mu)$ is a separable Hilbert space, and so is isomorphic to H . Now apply the theorem in the case $p = q = 2$.]

Note. One can also construct a direct proof of the Corollary without the restriction that H be separable; the case $p = q = 2$ of theorem 8.7 could then be deduced from this.

Reflexive Banach space

In exercise 8.4 (3) we proved that, for any Banach space H , $H^{**} \supset H$ in the sense that H is isomorphic to a Banach subspace of H^{**} . Those Banach spaces H for which $H = H^{**}$ are called *reflexive*. By our representation theorem 8.7, \mathcal{L}_p is reflexive for $1 < p < \infty$. In general, \mathcal{L}_1 is not reflexive because \mathcal{L}_∞^* is bigger than \mathcal{L}_1 : this will follow from exercises 8.5 (3,4). In fact very little is known about the structure of \mathcal{L}_∞^* : the difficulty is that the axiom of choice, or something equivalent, is needed to construct \mathcal{L}_∞^* and this makes it impossible to get a hold on it.

Exercises 8.5

1. Suppose $1 \leq p \leq +\infty$, $1/p + 1/q = 1$ and $f_n \rightarrow f$ in \mathcal{L}_p norm, $g_n \rightarrow g$ in \mathcal{L}_q norm. Deduce that

$$\int f_n g_n d\mu \rightarrow \int fg d\mu.$$

2. If Ω is the set of positive integers and μ is counting measure, then \mathcal{L}_p ($1 \leq p < \infty$) reduces to the set of sequences $\{x_i\}$ of real numbers such that $\sum_{i=1}^{\infty} |x_i|^p < \infty$; \mathcal{L}_∞ reduces to the set m of bounded sequences.

3. Let $X = [-1, 1]$, μ Lebesgue measure. Show that the collection \mathcal{C} of continuous functions $f: X \rightarrow \mathbb{R}$ is a closed linear subspace of \mathcal{L}_∞ (pro-

vided any function f which is equal a.e. to a continuous function is identified with it). Hence, by theorem 8.6, extend the bounded linear functional $F(f) = f(0)$ from \mathcal{C} to \mathcal{L}_∞ without changing its norm. If possible, suppose there is an f_0 which is integrable and such that

$$F(f) = \int f f_0 d\mu \quad \text{for } f \in \mathcal{L}_\infty.$$

Then, for the special sequence

$$f_n(x) = (1 - |x|^n),$$

we have $F(f_n) = 1$ for all n . Show that, for any $f_0 \in \mathcal{L}_\infty$,

$$\int f_n f_0 d\mu \rightarrow 0.$$

4. Extend example (3) to show that if Ω contains a disjoint sequence of measurable sets of finite positive measure, then $\mathcal{L}_1(\Omega, \mu)$ is a proper subspace of \mathcal{L}_∞^* . Deduce that \mathcal{L}_1 is not reflexive.

8.6* Mean ergodic theorem

In §7.6 we obtained the point-wise ergodic theorem for functions in \mathcal{L}_1 . If the function is in \mathcal{L}_2 there is an alternative form of this theorem in which point-wise convergence is replaced by convergence in second mean. We saw that any \mathcal{L}_2 is a Hilbert space. A measure preserving transformation T on the underlying measure space then leads naturally to a mapping on the Hilbert space to itself which preserves the inner product (and norm). It is therefore possible to state the mean ergodic theorem in terms of the properties of such a mapping in Hilbert space, and deduce the \mathcal{L}_2 theorem by considering this as a realization of Hilbert space. However, we choose instead to state and prove it directly as a theorem about the structure of $\mathcal{L}_2(\Omega, \mu)$. It helps if we first show that bounded linear functionals on a Banach space can be used to separate a closed linear subspace K from a point not in K (see exercise 8.4 (4)).

Theorem 8.8. *Suppose K is a linear subspace of a Banach space H , and $y \in H$ with $d(y, K) = \eta > 0$. Then there is a bounded linear functional F on H such that $\|F\| = 1$, $F(y) = \eta$, $F(x) = 0$ for all $x \in K$.*

Proof. Let J be the set of points of H of the form

$$x = z + ay, \quad z \in K, \quad a \in \mathbf{R}.$$

Then J is a linear subspace of H and the representation of points of J in this form is unique. Define a linear functional f on J by

$$f(z + ay) = a\eta.$$

Then f vanishes on K and, for $a \neq 0$,

$$\|z + ay\| = |a| \left\| \frac{z}{a} + y \right\| \geq |a| \eta = |f(z + ay)|,$$

so that $\|f\| \leq 1$. But if $\{z_n\}$ is a sequence in K for which $\|z_n - y\| \rightarrow \eta$ we have

$$\|f\| \|z_n - y\| \geq |f(z_n - y)| = |f(z_n) - f(y)| = |f(y)| = \eta$$

so that $\|f\| \geq 1$, on letting $n \rightarrow \infty$. Hence $\|f\| = 1$, and f has all the desired properties except that it is only defined on J , a linear subspace of H . Use theorem 8.6 to extend it to a linear functional F on the whole of H with $\|F\| = \|f\| = 1$.]

Corollary. If $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space, and K is a closed linear subspace of $\mathcal{L}_2(\Omega, \mu)$, and $y \in \mathcal{L}_2 - K$, then $y = z + x$ where $z \in K$ and $(x, w) = 0$ for all $w \in K$.

Proof. $\mathcal{L}_2(\Omega, \mu)$ is a Banach space, and K is closed (in the metric ρ_2) so that $d(y, K) = \eta > 0$. Find the functional F satisfying the conditions of theorem 8.8 and represent it, by theorem 8.7, as

$$F(\mu) = (\mu, g) \quad \text{where} \quad g \in \mathcal{L}_2.$$

Now put $x = \eta g$, $z = y - x$ so that

$$(x, w) = \eta F(w) = 0 \quad \text{for all} \quad w \in K.$$

It only remains to show that $z \in K$. For $\epsilon > 0$ choose $k \in K$ such that

$$\|k - y\|^2 = (k - y, k - y) < \eta^2 + \epsilon.$$

$$\begin{aligned} \text{Then} \quad \|k - z\|^2 &= (k - y, k - y) + 2(x, k - y) + (x, x) \\ &= \|k - y\|^2 + 2\eta(g, k - y) + \eta^2\|g\|^2 \\ &= \|k - y\|^2 - 2\eta F(y) + \eta^2\|F\|^2 \\ &= \|k - y\|^2 - \eta^2 < \epsilon, \end{aligned}$$

so that there are points of K arbitrarily close to z , and we must have $z \in K$, since K is closed.]

Let us remind ourselves of the conditions under which we established theorem 7.9. $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space, and T is a measure preserving transformation from Ω to itself. T^k is the result of repeating the transformation k times (T^0 is the identity map). For an \mathcal{F} -measurable function f which is finite a.e. we consider the sequence of means

$$g_n = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x). \quad (8.6.1)$$

Theorem 8.9. If $(\Omega, \mathcal{F}, \mu)$ and T satisfy the conditions in theorem 7.9, $f \in \mathcal{L}_2(\Omega, \mu)$, and g_n is defined by (8.6.1) then $\{g_n\}$ is a Cauchy sequence in second mean. Its limit (in second mean) f^* satisfies

(i) f^* is invariant under T , that is

$$f^*(Tx) = f^*(x) \text{ a.e.};$$

(ii) $\|f^*\| \leq \|f\|$;

(iii) for any function g in \mathcal{L}_2 which is invariant under T , $(g, f^*) = (g, f)$.

Proof. (a) Suppose first that f is such that there is an $h \in \mathcal{L}_2$ such that

$$f(x) = h(Tx) - h(x) \text{ a.e.}$$

Then

$$g_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \frac{1}{n} [h(T^n x) - h(x)]$$

so that $\|g\| \leq 2\|h\|/n \rightarrow 0$ as $n \rightarrow \infty$.

(b) Now suppose f is the limit (in second mean) of a sequence $\{f_k\}$ such that, for each k , $f_k(x) = h_k(Tx) - h_k(x)$ with $h_k \in \mathcal{L}_2$. Then

$$\begin{aligned} \|g_n\| &\leq \frac{1}{n} \left\| \sum_{i=0}^{n-1} \{f(T^i x) - f_k(T^i x)\} \right\| + \frac{1}{n} \left\| \sum_{i=0}^{n-1} f_k(T^i x) \right\| \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \|f(T^i x) - f_k(T^i x)\| + \frac{1}{n} \left\| \sum_{i=0}^{n-1} f_k(T^i x) \right\| \\ &\leq \|f - f_k\| + \frac{2}{n} \|h_k\|; \end{aligned}$$

so that we can make $\|g_n\| < \epsilon$ by first choosing f_k with $\|f - f_k\| < \frac{1}{2}\epsilon$ and then making n large.

The class of $f \in \mathcal{L}_2$ which satisfy either (a) or (b) is clearly a closed linear subspace K of \mathcal{L}_2 . By the corollary to theorem 8.8, any $f \in \mathcal{L}_2$ can be written uniquely as

$$f = f_1 + f_2 \quad \text{where} \quad f_1 \in K,$$

and

$$(f_2, Tf - f) = 0 \quad \text{for all} \quad f \in \mathcal{L}_2.$$

Now

$$\begin{aligned} 0 &= (f_2, Tf - f) = (f_2, Tf) - (f_2, f) \\ &= (T^{-1}f_2, f) - (f_2, f) = (T^{-1}f_2 - f_2, f) \quad \text{for all} \quad f \in \mathcal{L}_2, \end{aligned}$$

and in particular when f is the indicator function of a measurable set E of finite measure. Hence $T^{-1}f_2 = f_2$ a.e. so that f_2 is invariant under T .

Hence

$$\frac{1}{n} \sum_{i=0}^{n-1} f_2(T^i x) = f_2(x) \text{ a.e.} \quad \text{for all} \quad n,$$

so that $f^* = f_2$ is the limit in second mean of $\{g_n\}$. Thus (i) and (ii) are proved. To prove (iii), suppose g is invariant under T ; then

$$(T^i f, g) = (f, T^{-i} g) = (f, g)$$

so that $(g_n, g) = (f, g)$ for each n and the result follows on letting $n \rightarrow \infty$ since the inner product is continuous in the norm topology. \blacksquare

Corollary. *Under the conditions of theorem 8.9, if T is ergodic, then the limit (in second mean) $f^* = c$ a.e. Also*

(i) *if $\mu(\Omega) = \infty$, then $c = 0$,*

(ii) *if $\mu(\Omega) < \infty$, then $\int f^* d\mu = \int f d\mu$.*

Proof. The only invariant functions are constants so (ii) of the theorem implies that $f^* = c$ a.e. Now if $\mu(\Omega) = \infty$, we have $\|f^*\|$ finite, so $c = 0$. If $\mu(\Omega) < \infty$, then the function $g(x) \equiv 1$ is in \mathcal{L}_2 and is invariant so that

$$(1, f^*) = \int f^* d\mu = (1, f) = \int f d\mu. \blacksquare$$

Exercises 8.6

1. If $\mu(\Omega) < \infty$, $f \in \mathcal{L}_p(\Omega, \mu)$, $1 < p < \infty$ and T is a measure preserving transformation, show that g_n , defined by (8.6.1), converges a.e. to a limit function $f^* \in \mathcal{L}_p$ such that $\rho_p(g_n, f^*) \rightarrow 0$. (This gives a simpler proof of theorem 8.9 for the case $\mu(\Omega) < \infty$).

Prove that (ii) of the corollary to theorem 8.9 is valid without the condition that T be ergodic.

2. Suppose X is an open subset of \mathbf{R}^k of finite Lebesgue measure and $T: X \rightarrow X$ preserves Lebesgue measure and is ergodic. Show that, for almost all $x \in X$, the sequence $\{T^k x\}$ is dense in X .

9

STRUCTURE OF MEASURES IN SPECIAL SPACES

In the present book most of the theory of measure and integration has been developed in abstract spaces, and we have used the properties of special spaces only to illustrate the general theory. The present chapter, apart from §9.4, is devoted to a discussion of properties which depend essentially on the structure of the space.

The first question considered is that of point-wise differentiation. In the Radon–Nikodym theorem 6.7 we defined the derivative $d\mu/d\nu$ of one measure with respect to another for suitable measures μ, ν : but the point function $d\mu/d\nu$ obtained is only determined in the sense that the equivalence class of functions equal almost everywhere is uniquely defined. This means that at *no* single point (except for those points which form sets of positive measure) is the derivative defined by the Radon–Nikodym theorem. In order to define $d\mu/d\nu$ at a point x , the local topological structure of the space near x has to be considered. It is possible to develop this local differentiation theory in fairly general spaces, but only at the cost of complicated and rather unnatural additional conditions: we have decided instead to give the detailed theory only in the space \mathbf{R} of real numbers where the term derivative has a clear elementary meaning.

There are several ways of defining an integral with properties similar to those obtained in Chapter 5. So far in this book we have considered definitions which start from a given measure defined on a suitable class of sets. In §9.4 we describe the Daniell integral and show that, under suitable conditions this can be obtained in terms of a measure. Then, for locally compact spaces, we discuss positive linear functionals on the space C_K of real-valued continuous functions which vanish outside a compact set, and show that these also correspond to integrals with respect to a suitable measure.

The final section of the chapter is devoted to the definition of Haar measure in topological spaces which have the algebraic structure of a group and in which the group operation is continuous. The details are given only for locally compact metric groups.

9.1 Differentiating a monotone function

We say that $f: I \rightarrow \mathbf{R}$ where I is an open interval in \mathbf{R} (that is, a set of the form (a, b) with $a, b \in \mathbf{R}^*$), is *monotone increasing*, if

$$x_1, x_2 \in I, \quad x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2).$$

At a given point x in I the function $f: I \rightarrow \mathbf{R}$ may not be differentiable but

$$\begin{aligned} D^+f(x) &= \limsup_{h \rightarrow 0+} \frac{f(x+h) - f(x)}{h}, & D^-f(x) &= \limsup_{h \rightarrow 0+} \frac{f(x) - f(x-h)}{h}, \\ D_+f(x) &= \liminf_{h \rightarrow 0+} \frac{f(x+h) - f(x)}{h}, & D_-f(x) &= \liminf_{h \rightarrow 0+} \frac{f(x) - f(x-h)}{h} \end{aligned}$$

are always uniquely determined in the extended real number system \mathbf{R}^* . These numbers are called the *derivates* of f at x . We say that f is differentiable at x if

$$D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x) = Df(x) \neq \pm \infty.$$

It is clear that f is differentiable at x if and only if there is a real number $Df(x)$ such that, given $\epsilon > 0$ there is a $\delta > 0$ for which

$$\left| \frac{f(x+h) - f(x)}{h} - Df(x) \right| < \epsilon \quad \text{if} \quad 0 < |h| < \delta;$$

so that our definition is equivalent to that usually adopted in elementary texts on real analysis. When f is differentiable at x , we call $Df(x)$ the *derivative* of f at x .

If $f: I \rightarrow \mathbf{R}$ is continuous, but not monotone, it is possible that it is differentiable at no point x . However, a monotone $f: I \rightarrow \mathbf{R}$ must be continuous except at the points in a countable set, and the monotonicity further implies that there are some points x where the derivative exists. In fact we prove much more: the set of points x in I where f is not differentiable turns out to have zero measure. In order to prove this it is convenient first to obtain a new type of covering theorem. When in §2.2 we showed that a bounded closed interval K in \mathbf{R} is compact we started with a covering of K by a family of open sets and we demanded that *all* of K be covered by a finite subfamily. However, in proving compactness we were not interested in economical covering, and the covering sets finally chosen could overlap. Clearly if we require that the covering sets must not overlap we can no longer require that all of K be covered. However, even if we are satisfied with a countable subcovering by disjoint sets of almost all of K (see exercise 9.1(8)) additional conditions on the nature of the original covering are essential. A suitable form of these conditions now follows.

Vitali covering

For a subset $E \subset \mathbf{R}$, a class \mathcal{J} of intervals is said to cover E in the Vitali sense if, given $x \in E$, $\epsilon > 0$ there is an interval $J \in \mathcal{J}$ with $x \in J$ and $0 < |J| < \epsilon$.

Theorem 9.1. *Suppose $E \subset \mathbf{R}$ has finite Lebesgue outer measure and is covered in the Vitali sense by a class \mathcal{J} of intervals. Then there is a countable disjoint subclass $\mathcal{J}_1 \subset \mathcal{J}$ such that*

$$|E - \bigcup \{J : J \in \mathcal{J}_1\}| = 0.$$

Proof. We use $|A|$ to denote the Lebesgue outer measure of A whether or not A is measurable. There is no harm in assuming that all the intervals J in \mathcal{J} are closed since $|\bar{I}| = |I|$ for any interval I . We may further assume without loss of generality that there is an open set $G \supset E$ with $|G| < \infty$, and that all the intervals of \mathcal{J} are contained in G .

We choose \mathcal{J}_1 by induction as follows. Let J_1 be any interval of \mathcal{J} . Suppose we have already chosen disjoint intervals J_1, J_2, \dots, J_m and let s_m be the supremum of the lengths of the intervals in \mathcal{J} which do not intersect any of J_1, J_2, \dots, J_m . Now $s_m < |G| < \infty$, and if E is not contained in $\bigcup_{i=1}^m J_i$, we must have $s_m > 0$. Thus if E is not already covered, we can choose J_{m+1} disjoint from $\bigcup_{i=1}^m J_i$ with $|J_{m+1}| > \frac{1}{2}s_m$. Now the theorem is proved if $E \subset \bigcup_{i=1}^m J_i$ for any finite m . Otherwise we obtain a sequence $\{J_n\}$ of disjoint sets so that

$$\sum_{i=1}^{\infty} |J_i| \leq |G| < \infty.$$

Now suppose, if possible, that

$$|E - \bigcup_{i=1}^{\infty} J_i| = \delta > 0.$$

We can choose N so that $\sum_{i=N+1}^{\infty} |J_i| < \frac{1}{5}\delta$,

and put $F = E - \bigcup_{i=1}^N J_i$.

F must be non-void and $\bigcup_{i=1}^N J_i$ is closed so we can find a point x in E

and an interval J of \mathcal{J} containing x and short enough to be disjoint from $\bigcup_{i=1}^N J_i$. This implies $|J| \leq s_n < 2|J_{n+1}|$. Since

$$\lim_{n \rightarrow \infty} |J_n| = 0,$$

this J must meet at least one of the J_i for $i > n$. Let k be the smallest integer for which $J \cap J_k \neq \emptyset$. Then $|J| \leq s_{k-1} < 2|J_k|$, so the distance from x to the mid-point of J_k is at most $|J| + \frac{1}{2}|J_k| \leq \frac{5}{2}|J_k|$, and x must belong to the interval H_k which has the same centre as J_k and 5 times the length. Thus

$$F \subset \bigcup_{i=N+1}^{\infty} H_i$$

and
$$\delta = |F| \leq \sum_{i=N+1}^{\infty} |H_i| = 5 \sum_{i=N+1}^{\infty} |J_i| < \delta,$$

which establishes a contradiction. \blacksquare

Corollary. Under the conditions of theorem 9.1, for each $\epsilon > 0$ there is a finite set J_1, J_2, \dots, J_p of disjoint intervals of \mathcal{J} such that

$$\left| E - \bigcup_{i=1}^p J_i \right| < \epsilon.$$

Theorem 9.2. Suppose $f: I \rightarrow \mathbf{R}$ is monotone increasing. Then the set E of points x in I for which f is differentiable at x satisfies $|I - E| = 0$. The derivative f' is Lebesgue measurable, and if $[a, b] \subset I$,

$$\int_a^b f'(x) dx \leq f(b) - f(a).$$

Proof. It is clearly sufficient to prove the theorem for a finite closed interval $I = [c, d]$. The first step is to show that each of the subsets of I :

$$\{x: D^+f(x) > D_-f(x)\}, \quad \{x: D^-f(x) > D_+f(x)\},$$

$$\{x: D^+f(x) > D_+f(x)\}, \quad \{x: D^-f(x) > D_-f(x)\},$$

has zero Lebesgue measure. We give the details for the set

$$E = \{x: D^+f(x) > D_-f(x)\};$$

the proof for the others is similar. Now E is the (countable) union of sets

$$E_{u,v} = \{x: D^+f(x) > u > v > D_-f(x)\}$$

over rational pairs u, v . It is therefore sufficient to show that $|E_{u,v}| = 0$ for all pairs u, v with $u > v$.

Let $t = |E_{u,v}|$ and $\epsilon > 0$. Find an open set $G \supset E_{u,v}$ with

$$|G| < t + \epsilon.$$

For each $x \in E_{u,v}$, there is an arbitrarily small closed interval

$$[x-h, x] \subset G$$

with

$$f(x) - f(x-h) < v h.$$

By theorem 9.1, corollary we can find a finite disjoint collection J_1, J_2, \dots, J_N of such intervals whose interiors cover a subset F of $E_{u,v}$ with $|E_{u,v} - F| < \epsilon$. If we sum over these intervals

$$\begin{aligned} \sum_{n=1}^N \{f(x_n) - f(x_n - h_n)\} &< v \sum_{n=1}^N h_n < v|G| \\ &< v(t + \epsilon). \end{aligned}$$

But each $y \in F$ is the left-hand end-point of an arbitrarily small interval $[y, y+k]$ which is contained in one of the J_i ($i = 1, 2, \dots, N$) and such that

$$f(y+k) - f(y) > u k.$$

Use theorem 9.1 again to find a disjoint collection K_1, K_2, \dots, K_P of such intervals which covers a subset H of F with $|H| > t - 2\epsilon$. Summing over these intervals, since each K_i is contained in a J_n ,

$$\begin{aligned} \sum_{i=1}^N \{f(x_n) - f(x_n - h_n)\} &\geq \sum_{i=1}^P f(y_i + k_i) - f(y_i) \\ &> u \sum_{i=1}^P k_i > u(t - 2\epsilon) \end{aligned}$$

so that

$$v(t + \epsilon) > u(t - 2\epsilon).$$

Since $u > v$ and ϵ is arbitrary, we must have $t = 0$. Thus for almost all x in I ,

$$g(x) = Df(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

exists as an element in \mathbf{R}^* (we are thus allowing the value $\pm \infty$ for a limit). If we put

$$g_n(x) = n \left[f\left(x + \frac{1}{n}\right) - f(x) \right] \quad \text{for } x \in [a, b]$$

where we re-define $f(x) = f(b)$ for $x \geq b$, then $g_n(x)$ is defined and measurable and $g_n(x) \rightarrow g(x)$ for almost all x in $[a, b]$ as $n \rightarrow \infty$ so that $g: I \rightarrow \mathbf{R}^*$ is Lebesgue measurable if we define it arbitrarily to be zero

on the exceptional set where $Df(x)$ is not defined. By Fatou (theorem 5.7)

$$\begin{aligned} \int_a^b g(x) dx &\leq \liminf_{n \rightarrow \infty} \int_a^b g_n(x) dx \\ &= \liminf_n \int_a^b \left\{ f\left(x + \frac{1}{n}\right) - f(x) \right\} dx \\ &= \liminf \left[n \int_b^{b+(1/n)} f(x) dx - n \int_a^{a+(1/n)} f(x) dx \right] \\ &\leq f(b) - f(a). \end{aligned}$$

This shows that the function g is integrable and so finite almost everywhere. Thus f is differentiable a.e. in $[a, b]$. Since $[a, b]$ is an arbitrary subinterval of I , f is differentiable almost everywhere in I . ■

Functions of bounded variation

A function $f: I \rightarrow \mathbf{R}$ is said to be of bounded variation on I if

$$\sum_{i=1}^n |f(x_i) - f(x_{i-1})|$$

is bounded above for all ordered finite sequences $x_0 < x_1 < \dots < x_n$ in I . Clearly if $f: I \rightarrow \mathbf{R}$ is of bounded variation on I , it is also of bounded variation on each interval $J \subset I$. For an ordered sequence $\alpha = \{x_i\}$, $i = 0, 1, \dots, n$ put

$$\begin{aligned} p(\alpha) &= \sum_{i=1}^n \max [0, f(x_i) - f(x_{i-1})], \\ n(\alpha) &= - \sum_{i=1}^n \min [0, f(x_i) - f(x_{i-1})], \\ t(\alpha) &= p(\alpha) + n(\alpha) = \sum_{i=1}^n |f(x_i) - f(x_{i-1})|. \end{aligned}$$

If $f: [a, b] \rightarrow \mathbf{R}$ is of bounded variation on $[a, b]$, put

$$T_a^b = \sup_{\alpha} t(\alpha), \quad P_a^b = \sup_{\alpha} p(\alpha), \quad N_a^b = \sup_{\alpha} n(\alpha),$$

where each of the suprema is taken over all ordered finite sequences α in $[a, b]$. It is easy to check that, in this case

$$T_a^b = P_a^b + N_a^b, \quad f(b) - f(a) = P_a^b - N_a^b.$$

Now if $f: [a, b] \rightarrow \mathbf{R}$ is of bounded variation on $[a, b]$ we can put

$$g(x) = N_a^x, \quad h(x) = P_a^x \quad \text{for all } x \in [a, b]$$

so that $f(x)$ can be expressed as the difference of two non-decreasing functions of bounded variation.

Corollary (Lebesgue). *A function $f: I \rightarrow \mathbf{R}$ which is of bounded variation on each finite interval $[a, b] \subset I$ must be differentiable at x for almost all x in I .*

Proof. In each finite $[a, b]$ we can express f as the difference of two monotone increasing functions g and h . Each of these is differentiable almost everywhere in $[a, b]$ by theorem 9.2. Hence the difference f is differentiable almost everywhere in $[a, b]$. **■**

Exercises 9.1

1. Show that, if $g: I \rightarrow \mathbf{R}$, $h: I \rightarrow \mathbf{R}$ are each monotone increasing, then $f = g - h$ is of bounded variation on each $[a, b] \subset I$.

2. If $f: I \rightarrow \mathbf{R}$ is of bounded variation on each $[a, b] \subset I$, show that the limits $f(x+0)$, $f(x-0)$ exist at each interior point of I .

3. If c is an interior point of I and $f: I \rightarrow \mathbf{R}$ has a (local) maximum at c , show that $D^+f(c) \leq 0$, $D_-f(c) \geq 0$.

4. If $f: [a, b] \rightarrow \mathbf{R}$ is continuous and $D^+f(x) \geq 0$ for all x in $[a, b]$, show that $f(b) \geq f(a)$.

5. Define

$$\begin{aligned} f(0) &= 0, \\ f(x) &= x^2 \sin x^{-2} \quad \text{for } x \neq 0; \\ g(0) &= 0, \\ g(x) &= x^2 \sin x^{-1} \quad \text{for } x \neq 0. \end{aligned}$$

Which of the functions f , g is of bounded variation on $[-1, 1]$?

6. Give an example of a function for which all the four derivatives are different at $x = 0$.

7. For any Lebesgue measurable $f: I \rightarrow \mathbf{R}$, prove that $D^+f(x)$ is Lebesgue measurable.

8. Show that theorem 9.1 as stated in \mathbf{R} is false in \mathbf{R}^n for $n \geq 2$.

Hint. Take a Vitali covering of $[0, 1]$ and for each J of covering consider $J \times [0, \frac{2}{3}]$ and $J \times [\frac{1}{3}, 1]$. This will give a covering in the sense of our definition of the unit square $[0, 1] \times [0, 1]$. Show theorem 9.1 is not satisfied.

(In fact a more complicated construction shows that theorem 9.1 fails even if we require each point of the set to be covered by an interval J of arbitrarily small diameter.)

9. Show that theorem 9.1 is true in \mathbf{R}^n for all n if we restrict the covering to cubes. (In fact it can be shown that it is true if there is a constant K such that the ratio of the lengths of longest and shortest sides is bounded for the intervals in \mathcal{J} .)

10. For the Cantor ternary function $g: [0, 1] \rightarrow [0, 1]$ show that $g'(x) = 0$ for all $x \in [0, 1] - C$.

(This shows that we cannot hope, in general, for equality in

$$\int_a^b f'(x) dx \leq f(b) - f(a).)$$

11. Prove that a convergent series of non-decreasing real functions can be differentiated term by term a.e.

9.2 Differentiating the indefinite integral

The ‘fundamental theorem of the integral calculus’ states that, if $f: [a, b] \rightarrow \mathbf{R}$ is a continuous function and

$$F(x) = \int_a^x f(t) dt$$

then $F: [a, b] \rightarrow \mathbf{R}$ is differentiable in (a, b) with $F'(x) = f(x)$. The object of this section is to obtain the analogous theorem for the Lebesgue integral, where it is not appropriate to assume that f is continuous. (Of course, if $f: [a, b] \rightarrow \mathbf{R}$ is continuous on $[a, b]$, we know that $F'(x) = f(x)$ for all x in (a, b) since the Lebesgue integral coincides with the Riemann integral in this case.) The first thing to note is that, even for a monotonic function F , we cannot claim that, in general,

$$\int_a^b F'(x) dx = F(b) - F(a), \quad (9.2.1)$$

see exercise 9.1 (10). We will, however, obtain necessary and sufficient conditions for the truth of (9.2.1).

Lemma. If $f: [a, b] \rightarrow \mathbf{R}^*$ is Lebesgue integrable on $[a, b]$ and

$$\int_a^x f(t) dt = 0 \quad \text{for all } x \text{ in } [a, b],$$

then $f(t) = 0$ for almost all t in $[a, b]$.

Note. This strengthens the result of theorem 5.5 (vii).

Proof. If the lemma is false then at least one of the sets

$$\{t: f(t) < 0\}, \quad \{t: f(t) > 0\}$$

has positive measure. If $|\{t: f(t) > 0\}| > 0$ then we can find a $\delta > 0$ for which $|E| > 0$, where $E = \{t: f(t) > \delta\}$. Now choose a closed set $F \subset E$ with $|F| > 0$, and consider the open set $G = (a, b) - F$. Then

$$0 = \int_a^b f dm = \int_F f dm + \int_G f dm.$$

But G is the disjoint union of a countable collection of open intervals

$$(a_n, b_n) \quad \text{and} \quad \int_{a_n}^{b_n} f dm = 0$$

for each n . Since the integral defines a σ -additive set function we must have

$$\int_G f dm = 0 \quad \text{so that} \quad \int_F f dm = 0$$

and this contradicts $\int_F f dm \geq \delta |F| > 0$. ■

Let us now consider the properties of any function F which is an indefinite integral, that is

$$F(x) = \int_a^x f(t) dt$$

for a function $f: [a, b] \rightarrow \mathbf{R}^*$ which is Lebesgue integrable. It is immediate from theorem 5.6 that F is continuous on $[a, b]$, but more can be said: since it is the difference of the indefinite integrals of f_+ and f_- it must be the difference of two monotone functions and therefore it is of bounded variation. In fact, we saw in theorem 5.6 that the set function

$$\nu(E) = \int_E f dm: E \text{ measurable, } E \subset [a, b]$$

is absolutely continuous; that is that $\nu(E) \rightarrow 0$ as $m(E) \rightarrow 0$. This means in particular that given $\epsilon > 0$, there is a $\delta > 0$ such that if

$E = \bigcup_{k=1}^n I_k$ is a finite disjoint union of intervals in $[a, b]$ for which

$$\sum_{k=1}^n m(I_k) < \delta, \quad \text{then} \quad |\nu(E)| = \left| \sum_{k=1}^n \nu(I_k) \right| < \epsilon.$$

In fact, by considering separately the intervals I_k for which ν is positive and negative we can find $\delta > 0$ such that

$$\sum_{k=1}^n m(I_k) < \delta \Rightarrow \sum_{k=1}^n |\nu(I_k)| < \epsilon.$$

In terms of the indefinite integral F this means that the function $F: [a, b] \rightarrow \mathbf{R}$ is such that, for each $\epsilon > 0$ there is a $\delta > 0$ for which

$$\sum_{i=1}^n (b_i - a_i) < \delta \Rightarrow \sum_{i=1}^n |F(b_i) - F(a_i)| < \epsilon \quad (9.2.2)$$

for any finite class of disjoint intervals $(a_i, b_i) \subset (a, b)$. Any function $F: I \rightarrow \mathbf{R}$ which satisfies this condition on every finite interval $(a, b) \subset I$ is said to be *absolutely continuous* on I .

It is immediate that any function $F: I \rightarrow \mathbf{R}$ which is absolutely continuous is of bounded variation on each finite interval $[a, b] \subset I$. For if we put $\epsilon = 1$ in (9.2.2) and choose $\delta > 0$, then any finite dissec-

tion of $[a, b]$ can be split into K sets of intervals (by inserting extra division points if necessary) each of total length less than δ , where $K = [(b-a)/\delta] + 1$; and it follows that, for any dissection of $[a, b]$

$$\sum_{r=1}^n |F(x_r) - F(x_{r-1})| \leq K.$$

By the corollary to theorem 9.2 we now see that any function F which is absolutely continuous is differentiable except on a set of zero measure.

Theorem 9.3. Suppose $f: [a, b] \rightarrow \mathbf{R}^*$ is Lebesgue integrable on $[a, b]$ and $F: [a, b] \rightarrow \mathbf{R}$ satisfies

$$F(x) = F(a) + \int_a^x f(t) dt.$$

Then F is differentiable with $F'(x) = f(x)$ for almost all x in $[a, b]$.

Proof. Assume first that f is bounded on $[a, b]$ so that for a suitable M in \mathbf{R} , $|f(x)| \leq M$, for all x in $[a, b]$. Now we know that F is absolutely continuous and therefore differentiable almost everywhere. Put

$$f_n(x) = n \left[F\left(x + \frac{1}{n}\right) - F(x) \right].$$

Then $|f_n| \leq M$ and $f_n(x) \rightarrow F'(x)$ almost everywhere; so, by theorem 5.8 for $a \leq c \leq b$,

$$\begin{aligned} \int_a^c F'(x) dx &= \lim \int_a^c f_n(x) dx = \lim n \int_a^c \left[F\left(x + \frac{1}{n}\right) - F(x) \right] dx \\ &= \lim \left[n \int_c^{c+(1/n)} F(x) dx - n \int_a^{a+(1/n)} F(x) dx \right] \\ &= F(c) - F(a) = \int_a^c f(x) dx \end{aligned}$$

since F is continuous. Hence

$$\int_a^c \{F'(x) - f(x)\} dx = 0$$

for all c in $[a, b]$ so that $F'(x) = f(x)$ almost everywhere.

Now suppose that $f: [a, b] \rightarrow \mathbf{R}^*$ is integrable but not bounded. From the definition of the integral it is sufficient to prove the theorem when $f \geq 0$. Put

$$g_n(x) = \min [n, f(x)]$$

and

$$G_n(x) = \int_a^x [f(t) - g_n(t)] dt.$$

Since $f - f_n \geq 0$, G_n is monotone increasing and so has a non-negative derivative almost everywhere. Since f_n is bounded (by n) we know that

$$\frac{d}{dx} \left\{ \int_a^x f_n(t) dt \right\} = f_n(x) \text{ a.e.,}$$

so that the derivative

$$F'(x) = G'_n(x) + \frac{d}{dx} \left\{ \int_a^x f_n(t) dt \right\} \geq f_n(x),$$

and exists almost everywhere. Since this is true for each integer n ,

$$F'(x) \geq f(x) \text{ a.e.} \quad (9.2.3)$$

Hence
$$\int_a^b F'(x) dx \geq \int_a^b f(x) dx = F(b) - F(a),$$

and by theorem 9.2 we must have

$$\int_a^b F'(x) dx = F(b) - F(a) = \int_a^b f(x) dx,$$

and

$$\int_a^b \{F'(x) - f(x)\} dx = 0.$$

This with (9.2.3) implies that $F'(x) = f(x)$ a.e. \blacksquare

Lemma. If $F: [a, b] \rightarrow \mathbf{R}$ is absolutely continuous on $[a, b]$ and

$$F'(x) = 0 \text{ a.e.,}$$

then F is constant.

Proof. Suppose $a \leq c \leq b$, and $E = \{x \in [a, c]; F'(x) = 0\}$. For a fixed $\epsilon > 0$, there are arbitrarily small intervals $[x, x+h]$ for each $x \in E$ such that

$$|F(x+h) - F(x)| < \epsilon h.$$

Choose $\delta > 0$ to satisfy (9.2.2) in the definition of absolute continuity and use theorem 9.1 to obtain a finite collection $[x_k, y_k]$ of intervals with

$$|F(y_k) - F(x_k)| < \epsilon(y_k - x_k)$$

which cover all of E except for a subset of measure less than δ . Order these intervals so that

$$y_0 = a \leq x_1 < y_1 \leq x_2 < \dots \leq y_n \leq c = x_{n+1},$$

and

$$\sum_{i=0}^n |x_{i+1} - y_i| < \delta.$$

By (9.2.2) this implies
$$\sum_{i=0}^n |F(x_{i+1}) - F(y_i)| < \epsilon$$

and, from the choice of the covering family

$$\sum_{i=0}^n |F(y_i) - F(x_i)| < \epsilon(c-a)$$

so that

$$\begin{aligned} |F(c) - F(a)| &= \left| \sum_{i=0}^n \{F(x_{i+1}) - F(y_i)\} + \sum_{i=0}^n \{F(y_i) - F(x_i)\} \right| \\ &< \epsilon(c-a+1). \end{aligned}$$

Since ϵ is arbitrary, we have $F(c) = F(a)$. **■**

Theorem 9.4. *A function $F: I \rightarrow \mathbf{R}$ is an indefinite integral, that is there is a measurable $f: I \rightarrow \mathbf{R}^*$ such that*

$$F(b) - F(a) = \int_a^b f(x) dx$$

for all $[a, b] \subset I$, if and only if F is absolutely continuous on I .

Proof. We have already seen that any indefinite integral is absolutely continuous. Conversely suppose $F: I \rightarrow \mathbf{R}$ is absolutely continuous. Then F is differentiable almost everywhere in $[a, b]$ and

$$|F'(x)| \leq F'_1(x) + F'_2(x) \text{ a.e.,}$$

where $F = F_1 - F_2$ expresses F as the difference of two monotone functions. By theorem 9.2, F' is integrable on $[a, b]$. Put

$$G(x) = \int_a^x F'(t) dt.$$

Then G is absolutely continuous and so is $H = F - G$. But, by theorem 9.3,

$$H' = F' - G' = F' - F' = 0 \text{ a.e.}$$

so that H is constant by the lemma. Hence

$$F(x) = \int_a^x F'(t) dt + F(a). \text{ **■**}$$

Corollary. *Every absolutely continuous function $F: I \rightarrow \mathbf{R}$ is the indefinite integral of its derivative.*

Density

Given a set $A \subset \mathbf{R}$, $x \in \mathbf{R}$ consider the ratio

$$\frac{|I \cap A|}{|I|}$$

for all intervals I containing x where $|E|$ denotes the Lebesgue outer measure of E . If this ratio converges to a limit as $|I| \rightarrow 0$, then

this limit is called the density of A at x and denoted $\tau(x, A)$. The point x is called a point of density for A if $\tau(x, A) = 1$, and a point of dispersion for A if $\tau(x, A) = 0$. We can obtain the following as a corollary of theorem 9.4.

Lemma (Lebesgue). *If $A \subset \mathbf{R}$, A is Lebesgue measurable, then*

$$\tau(x, A) = 1 \quad \text{for almost all } x \in A,$$

$$\tau(x, A) = 0 \quad \text{for almost all } x \in \mathbf{R} - A.$$

Proof. Suppose $a < x < b$. Then the indicator function χ_A is Lebesgue integrable over $[a, b]$. Hence

$$F(x) = \int_a^x \chi_A dx$$

is differentiable almost everywhere and

$$F'(x) = 1 \quad \text{for almost all } x \text{ in } [a, b] \cap A,$$

$$F'(x) = 0 \quad \text{for almost all } x \text{ in } [a, b] \cap (\mathbf{R} - A).$$

But if x is such that $F'(x) = 1$, there is for each $\epsilon > 0$ a $\delta > 0$ such that

$$(i) \quad 1 \geq \frac{|[x, x+h] \cap A|}{h} > 1 - \epsilon \quad \text{for } 0 < h < \delta,$$

$$(ii) \quad 1 \geq \frac{|[x-k, x] \cap A|}{k} > 1 - \epsilon \quad \text{for } 0 < k < \delta;$$

$$\text{and so} \quad 1 \geq \frac{|[x-k, x+h] \cap A|}{h+k} > 1 - \epsilon \quad \text{for } 0 < h, k < \delta,$$

which is precisely the condition for $\tau(x, A) = 1$. A similar proof shows that, at points x where $F'(x) = 0$ we have $\tau(x, A) = 0$. \blacksquare

Exercises 9.2

1. If $F: I \rightarrow \mathbf{R}$ is absolutely continuous, show that F^p is absolutely continuous for each $p \geq 1$, but not, in general, for $p < 1$.

2. If $F: [a, b] \rightarrow \mathbf{R}$ is such that F' exists everywhere in (a, b) and is bounded show that

$$\int_a^b F'(x) dx = F(b) - F(a).$$

For $F(x) = x^2 \sin 1/x^2$ ($x \neq 0$), $F(0) = 0$ show that $F'(x)$ exists for all x but is not Lebesgue integrable over $[-1, 1]$. (This shows that even the Lebesgue integral is not strong enough to integrate all derivatives.)

3. Construct a subset $A \subset \mathbf{R}$ for which $\tau(0, A) = \frac{1}{2}$.

4. Extend the density result to non-measurable sets A by showing that for any $A \subset \mathbf{R}$, $\tau(x, A) = 1$ for all x in A except a subset of zero measure.

Hint. Assume A is contained in a finite interval, and take a measurable set $B \supset A$ with $|B| = |A|$.

Deduce that a set $A \subset \mathbf{R}$ is measurable if and only if $\tau(x, A) = 0$ for almost all x in $(\mathbf{R} - A)$.

5. Prove that the Cantor function $g: [0, 1] \rightarrow [0, 1]$ defined in §2.7 is monotone increasing and continuous but not absolutely continuous.

6. The function $f: [0, 1] \rightarrow \mathbf{R}$ is absolutely continuous on $[\epsilon, 1]$ for each $\epsilon > 0$. Can one deduce that f is absolutely continuous on $[0, 1]$? Does the additional condition that f is of bounded variation on $[0, 1]$ help?

9.3 Point-wise differentiation of measures

In theorem 4.8 we proved that all measures μ in \mathbf{R} defined for Borel sets and finite on bounded sets are Lebesgue–Stieltjes measures: that is, there is a monotone increasing function $F: \mathbf{R} \rightarrow \mathbf{R}$ which is continuous on the right such that $\mu = \mu_F$ on \mathcal{B} . Because of this correspondence we can obtain properties of such Borel measures in terms of the corresponding properties of F .

Lemma 1. *Suppose μ_F is the Lebesgue–Stieltjes measure with respect to the function $F: \mathbf{R} \rightarrow \mathbf{R}$ which is continuous on the right. Then μ_F is absolutely continuous with respect to Lebesgue measure m if and only if F is absolutely continuous.*

Proof. Suppose first that F is absolutely continuous. Then, by theorem 9.4

$$\mu_F(a, b] = F(b) - F(a) = \int_a^b F'(t) dt$$

so that, for $E \in \mathcal{P}$, μ_F coincides with the set function

$$\nu(E) = \int_E F' dm.$$

But the extension of a measure from \mathcal{P} to \mathcal{B} is unique, so that $\mu_F = \nu$ on \mathcal{B} , and μ_F must therefore be absolutely continuous with respect to m .

Conversely, if μ_F is absolutely continuous with respect to m , by the Radon–Nikodym theorem there is an $f \geq 0$ such that

$$\mu_F = \int_E f dm \quad \text{for } E \in \mathcal{B}.$$

Hence $\mu_F(0, x] = F(x) - F(0) = \int_0^x f(t) dt$ for $x \geq 0$,

$$\mu_F(x, 0] = F(0) - F(x) = -\int_0^x f(t) dt \quad \text{for } x < 0,$$

so that $F: \mathbf{R} \rightarrow \mathbf{R}$ is an indefinite integral and must therefore be absolutely continuous. **】**

Atom

Given any measure space (X, \mathcal{F}, μ) in which \mathcal{F} contains all single point sets the point $x \in X$ is said to be an atom for the measure μ if $\mu\{x\} > 0$. A measure μ with no atoms is said to be *non-atomic*. Now if μ is σ -finite, the set of atoms of μ is countable. In this case if we put

$$\nu(E) = \sum_{\substack{x \in E \\ \mu\{x\} > 0}} \mu\{x\}$$

we obtain a new measure ν defined on all subsets of X , and ν is a *discrete measure* as defined in §3.1. Further, the set function

$$\tau = \mu - \nu$$

defined on \mathcal{F} is clearly non-atomic and so

$$\mu = \nu + \tau$$

is a decomposition of a σ -finite measure μ into the sum of a discrete measure and a non-atomic measure. This decomposition is clearly unique. Thus we have proved

Lemma 2. *Given a σ -finite measure space (X, \mathcal{F}, μ) in which \mathcal{F} contains all single point sets there is a unique decomposition of μ ,*

$$\mu = \nu + \tau$$

for which ν is a discrete measure on X and τ is a non-atomic measure on \mathcal{F} .

Lemma 3. *A measure μ on \mathcal{B} (the Borel sets of \mathbf{R}) which is finite on bounded intervals is a discrete measure if and only if $\mu = \mu_F$ where F is a jump function, that is,*

$$\left. \begin{aligned} F(x) &= \sum_{0 < x_i \leq x} p_i \quad \text{for } x \geq 0, \\ -F(x) &= \sum_{x < x_i \leq 0} p_i \quad \text{for } x < 0, \end{aligned} \right\} \quad (9.3.1)$$

where the measure μ has atoms x_i of weight (or measure) p_i .

Proof. It is clear that if $F: \mathbf{R} \rightarrow \mathbf{R}$ satisfies (9.3.1) then

$$\mu_F(a, b] = \sum_{a < x_i \leq b} p_i = \sum_{x_i \in (a, b]} p_i$$

so that μ_F coincides with the discrete measure

$$\nu(E) = \sum_{x_i \in E} p_i$$

for $E \in \mathcal{P}$. By uniqueness of extension μ_F must be a discrete measure.

Conversely, if μ is a discrete measure with atoms x_i of weight p_i , an application of the theorem 4.8 shows that $\mu = \mu_F$ with F a jump function. **]**

Lemma 4. *A measure μ defined on \mathcal{B} which is finite on bounded intervals is non-atomic if and only if $\mu = \mu_F$ for a continuous $F: \mathbf{R} \rightarrow \mathbf{R}$.*

Proof. If F is continuous, then

$$0 \leq \mu_F\{x\} < \mu_F(x-h, x] = F(x) - F(x-h)$$

for all $h > 0$, so that $\mu_F\{x\} = 0$.

Conversely if F is not continuous at x_0 , then

$$\mu_F\{x_0\} = \lim_{n \rightarrow \infty} \mu_F\left(x_0 - \frac{1}{n}, x_0\right] = F(x_0) - F(x_0 - 0)$$

so x_0 is an atom. **]**

Singular monotone function

Any function $F: I \rightarrow \mathbf{R}$ which is continuous and monotone increasing, such that $F'(x) = 0$ for all x in I except for a set of zero Lebesgue measure, is said to be *singular*. The function g defined in §2.7 clearly satisfies these conditions without being constant.

Lemma 5. *A function $F: \mathbf{R} \rightarrow \mathbf{R}$ is singular if and only if the Lebesgue-Stieltjes measure μ_F is non-atomic and singular with respect to Lebesgue measure.*

Proof. The continuity of F is equivalent to the condition that μ_F be non-atomic by lemma 4. Now a measure ν is singular with respect to Lebesgue measure if and only if any absolutely continuous τ satisfying $\tau(E) \leq \nu(E)$ for all E in \mathcal{B} must be zero. Now if $F'(x) > 0$ on a set of positive measure, the set function

$$\tau(E) = \int_E F'(x) dx$$

is not always zero and $\tau \leq \mu_F$ by theorem 9.2 so μ_F is not singular with respect to Lebesgue measure. Conversely, if μ_F is not singular a

non-null absolutely continuous measure $\tau \leq \mu_F$ can be found, and this corresponds to a function G , that is

$$G(b) - G(a) = \int_a^b G'(x) dx.$$

But $F'(x) \geq G'(x)$ when both are defined, so $F'(x) > 0$ on a set of positive measure. \blacksquare

Theorem 9.5 (Lebesgue). *Given any function $F: \mathbf{R} \rightarrow \mathbf{R}$ which is monotone increasing and continuous on the right, there is a decomposition of F*

$$F = F_1 + F_2 + F_3$$

where F_1 is a jump function,

F_2 is singular,

F_3 is absolutely continuous.

This decomposition is unique if we insist that $F_1(0) = F_2(0) = 0$.

Proof. Use the function F to define a Lebesgue-Stieltjes measure μ_F on \mathcal{B} . Decompose μ_F with respect to Lebesgue measure m by theorem 6.7 so that

$$\mu_F = \nu_1 + \nu_3$$

with $\nu_3 \leq m$ and ν_1 singular with respect to m . Decompose ν_1 by lemma 2,

$$\nu_1 = \lambda_1 + \lambda_2,$$

where λ_1 is discrete and λ_2 is non-atomic.

Let F_1, F_2 be the monotone functions (with $F_1(0) = F_2(0) = 0$) obtained by theorem 4.8 for which $\lambda_1 = \mu_{F_1}$, $\lambda_2 = \mu_{F_2}$ on \mathcal{B} . Then by lemmas 3 and 5, F_1 is a jump function, and F_2 is a singular function. If one applies theorem 4.8 to ν_3 one obtains an absolutely continuous G_3 for which $\nu_3 = \mu_{G_3}$. Finally, put $F_3(x) = G_3(x) + F(0)$ and we still have F_3 absolutely continuous, and $\nu_3 = \mu_{F_3}$. Now

$$F(x) - F(0) = F_1(x) - F_1(0) + F_2(x) - F_2(0) + F_3(x) - F_3(0)$$

for all x so that $F(x) = F_1(x) + F_2(x) + F_3(x)$.

The uniqueness follows from the uniqueness of the decomposition $\mu_F = \lambda_1 + \lambda_2 + \nu_3$, and theorem 4.8. \blacksquare

In \mathbf{R} we can also use the connexion between μ_F and F to define differentiation. Thus if $F: I \rightarrow \mathbf{R}$ is differentiable at x_0 , this means that

$$\frac{F(x_0 + h) - F(x_0 - k)}{h + k} \rightarrow F'(x_0) \quad \text{as } h, k \rightarrow 0$$

with $h > 0, k > 0$, and

$$\frac{\mu_F(x_0 - k, x_0 + h]}{|(x_0 - k, x_0 + h]|} \rightarrow F'(x_0).$$

This can be written

$$\frac{\mu_F(J)}{|J|} \rightarrow F'(x_0) \quad \text{as } |J| \rightarrow 0$$

for intervals J containing x_0 , and we can write $d\mu_F/dm(x_0)$ for the value of this limit. More generally, if μ, ν are two measures in \mathbf{R} which are finite for bounded sets then

$$\lim_{\substack{|J| \rightarrow 0 \\ x \in J}} \left[\frac{\mu(J)}{\nu(J)} \right],$$

when it exists, is called the derivative of μ with respect to ν at the point x .

In \mathbf{R}^n we can consider the values of the ratio

$$\frac{\mu(J)}{\nu(J)} \tag{9.3.2}$$

for rectangles J (in \mathcal{P}^n) containing a fixed point x and ask whether or not this ratio approaches a limit as $\text{diam}(J) \rightarrow 0$. The existence of this limit for all x except for a set of zero ν measure can be proved when ν is Lebesgue measure: the limit in this case is called the *strong derivate* of μ at x . This result is harder to prove than the result in §9.1 because theorem 9.1 is not valid without some restriction on the ratio of the sides of the covering class \mathcal{J} . Essentially similar methods to those of §9.1 will work if only cubes J are considered. On the other hand if in (9.3.2) one considers rectangles with arbitrary orientation an example can be given for which the limit exists nowhere.

Differentiation point-wise in abstract spaces can also be defined in terms of suitable ‘nets’, and the theorems of this chapter can be obtained if sufficient conditions are imposed. Since the results are not often used in practice, we will not state them in detail.

Exercises 9.3

1. Enumerate the rationals as a sequence $\{r_i\}$. By considering the discrete measure with mass $1/i^2$ at r_i ($i = 1, 2, \dots$) define a jump function which is constant in no interval.
2. Give an example of a singular function which is constant in no interval.
3. If F, G are two monotone real functions differentiable at x_0 with $G'(x_0) \neq 0$, show that

$$\frac{d\mu_F}{d\mu_G}(x_0) = \lim_{\substack{x_0 \in J \\ |J| \rightarrow 0}} \frac{\mu_F(J)}{\mu_G(J)} \quad \text{exists and equals} \quad \frac{F'(x_0)}{G'(x_0)}.$$

9.4* The Daniell integral

Our approach in this book has been to regard measure as the primitive concept, and to define the integration process in terms of a given measure. One important alternative is to start with an 'integral' defined on a suitable class of functions, extend its definition to a larger domain with desirable properties and then obtain measure as a by-product at a later stage. In the present section we describe this alternative approach: it is convenient to use it in the following section to obtain the integral representation of an important class of linear functionals.

For an arbitrary space X , we consider a family L of functions $f: X \rightarrow \mathbf{R}$ satisfying

- (i) L is a linear space over the reals;
- (ii) for each $f \in L$, the function $f_+ \in L$, where

$$f_+(x) = \max(0, f(x)).$$

Now if we define, for each $f, g \in L$, $x \in X$

$$(f \vee g)(x) = \max(f(x), g(x)),$$

$$(f \wedge g)(x) = \min(f(x), g(x)),$$

the relations

$$f_+ = f \vee 0, \quad f \vee g = (f - g) \vee 0 + g, \quad f \wedge g = f + g - (f \vee g);$$

show that

- (iii) if $f, g \in L$, then $f \vee g, f \wedge g \in L$.

Any family L satisfying conditions (i) and (ii) (and therefore (iii)) is called a *vector lattice* of functions. Suppose \mathcal{J} is a linear functional on L (considered as a real linear space), then we say \mathcal{J} is *positive* if

$$f \in L, \quad f \geq 0 \Rightarrow \mathcal{J}(f) \geq 0.$$

A positive linear functional \mathcal{J} on L is said to be a *Daniell functional* if, for every increasing sequence $\{f_n\}$ of functions of L

$$\mathcal{J}(g) \leq \lim_{n \rightarrow \infty} \mathcal{J}(f_n) \tag{9.4.1}$$

for each $g \in L$ satisfying $g(x) \leq \lim_{n \rightarrow \infty} f_n(x)$ for all $x \in X$. (Note that $\lim_{n \rightarrow \infty} f_n(x)$ will be $+\infty$ if the sequence $\{f_n(x)\}$ is unbounded, and even if $\lim f_n$ exists as a function with finite values we do not assume that it is in L .)

In particular, this implies that, if \mathcal{J} is a Daniell functional, $\{f_n\}$ a monotone sequence in L such that $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, $x \in X$ defines

a function in L then $\mathcal{J}(f) = \lim_{n \rightarrow \infty} \mathcal{J}(f_n)$. For if $\{f_n\}$ is increasing then $f \geq f_n$ for all n , so $\mathcal{J}(f) \geq \mathcal{J}(f_n)$ since \mathcal{J} is positive, which with (9.4.1) gives the required equality. Thus a Daniell functional is continuous in the sense that for any sequence $\{f_n\}$ in L which decreases monotonically to the zero function we must have $\mathcal{J}(f_n) \rightarrow 0$. Any Daniell functional is therefore an 'integral' in the sense discussed in §5.1. However, for the integral to be useful we want the domain L to be as large as possible: if $\{f_n\}$ is an increasing sequence in L which is bounded above by an element of L we would certainly want $\lim f_n$ to be in L . The Daniell integral is essentially the result of extending a Daniell functional \mathcal{J} from L to a class $L_1 \supset L$: it turns out that this extension can be carried out in two stages.

Suppose \mathcal{J} is a Daniell functional on a vector lattice L . Denote by L^+ the set of functions $f: X \rightarrow \mathbf{R}^*$ which are limits of monotone increasing functions of L . L^+ is not a linear space but

$$\alpha, \beta \geq 0 \quad f, g \in L^+ \Rightarrow \alpha f + \beta g \in L^+.$$

Then if $\{f_n\}$ is an increasing sequence in L , $\{\mathcal{J}(f_n)\}$ is an increasing sequence in \mathbf{R} which has a unique limit in $\mathbf{R} \cup \{+\infty\}$. We can define \mathcal{J} in L^+ by

$$\mathcal{J}(\lim f_n) = \lim \mathcal{J}(f_n).$$

This definition is proper because if $\{f_n\}, \{g_n\}$ are two monotone sequences each converging to h in L^+ , condition (9.4.1) gives, for fixed n ,

$$f_n \leq h = \lim g_n \Rightarrow \mathcal{J}(f_n) \leq \lim \mathcal{J}(g_n)$$

so that $\lim \mathcal{J}(f_n) \leq \lim \mathcal{J}(g_n)$ and the opposite inequality can be similarly obtained. It is clear that \mathcal{J} is linear on L^+ in the sense that

$$\alpha \geq 0, \quad \beta \geq 0; \quad f, g \in L^+ \Rightarrow \mathcal{J}(\alpha f + \beta g) = \alpha \mathcal{J}(f) + \beta \mathcal{J}(g).$$

For an arbitrary function $f: X \rightarrow \mathbf{R}^*$ we define the *upper integral* $\mathcal{J}^*(f)$ by

$$\mathcal{J}^*(f) = \inf_{\substack{g \geq f \\ g \in L^+}} \mathcal{J}(g),$$

where we adopt the (usual) convention that the infimum of the empty set is $+\infty$. Similarly, the *lower integral* $\mathcal{J}_*(f)$ is defined by

$$\mathcal{J}_*(f) = -\mathcal{J}^*(-f),$$

and we say that a function $f: X \rightarrow \mathbf{R}^*$ is *integrable* (with respect to \mathcal{J}) if $\mathcal{J}^*(f) = \mathcal{J}_*(f)$ and is finite. The class of integrable functions will be denoted by $L_1 = L_1(\mathcal{J}, L)$. For $f \in L_1$ we call the common value of $\mathcal{J}^*(f), \mathcal{J}_*(f)$ the *integral* of f and denote it by $\mathcal{J}(f)$. We now show that

this functional \mathcal{J} on L_1 is a Daniell functional which extends \mathcal{J} , and that L_1 has the closure properties desired. It is convenient to obtain a number of preliminary results before stating the theorem.

Lemma 1. *If $\{g_n\}$ is a sequence of non-negative functions in L^+ , then*

$$g = \sum_{n=1}^{\infty} g_n \quad \text{is in } L^+ \quad \text{and} \quad \mathcal{J}(g) = \sum_{n=1}^{\infty} \mathcal{J}(g_n).$$

Proof. It is clear that a non-negative function $f: X \rightarrow \mathbf{R}^+$ belongs to L^+ if and only if there is a sequence $\{f_n\}$ of non-negative functions

in L with $f = \sum_{n=1}^{\infty} f_n$. By definition, in this case

$$\mathcal{J}(f) = \sum_{n=1}^{\infty} \mathcal{J}(f_n).$$

Hence, each function g_n can be expressed as a sum

$$g_n = \sum_{v=1}^{\infty} f_{n,v} \quad \text{with} \quad f_{n,v}: X \rightarrow \mathbf{R}^+, \quad f_{n,v} \in L.$$

It follows that

$$g = \sum_n \sum_v f_{n,v}$$

is a countable sum of non-negative functions of L and so must be in L^+ . Further since all the terms are non-negative, the order of summation is immaterial and

$$\mathcal{J}(g) = \sum_{n=1}^{\infty} \left(\sum_{v=1}^{\infty} \mathcal{J}(f_{n,v}) \right). \quad \mathbf{1}$$

Lemma 2. *For arbitrary functions $f: X \rightarrow \mathbf{R}^*$, $g: X \rightarrow \mathbf{R}^*$:*

- (i) $\mathcal{J}^*(f+g) \leq \mathcal{J}^*(f) + \mathcal{J}^*(g)$;
- (ii) if $c \geq 0$, $\mathcal{J}^*(cf) = c\mathcal{J}^*(f)$;
- (iii) if $f \leq g$, $\mathcal{J}^*(f) \leq \mathcal{J}^*(g)$, $\mathcal{J}_*(f) \leq \mathcal{J}_*(g)$;
- (iv) $\mathcal{J}_*(f) \leq \mathcal{J}^*(f)$;
- (v) if $f \in L^+$, $\mathcal{J}_*(f) = \mathcal{J}^*(f) = \mathcal{J}(f)$.

Proof. (i), (ii) and (iii) follow immediately from the definitions. It is worth noting in (i), that we can put $(f+g)(x) = +\infty$ at those points x for which one of $f(x)$ is $+\infty$ and the other is $-\infty$ so that (i) is true whatever the value in \mathbf{R}^* chosen for $(f+g)(x)$ at such points x .

(iv) Since $0 = \mathcal{J}(0) = \mathcal{J}(f-f) \leq \mathcal{J}^*(f) + \mathcal{J}^*(-f)$ by (i), it follows that $\mathcal{J}_*(f) = -\mathcal{J}_*(-f) \leq \mathcal{J}^*(f)$.

(v) If $f \in L^+$, then by definition $\mathcal{J}^*(f) = \mathcal{J}(f)$. Now if $g \in L$, then

$-g \in L \subset L^+$ so that $\mathcal{J}_*(g) = \mathcal{J}(g)$. But each f in L^+ is the limit of an increasing sequence $\{g_n\}$ in L . Thus $f \geq g_n$ so $\mathcal{J}_*(f) \geq \mathcal{J}_*(g_n) = \mathcal{J}(g_n)$ and $\mathcal{J}_*(f) \geq \lim \mathcal{J}(g_n) = \mathcal{J}(f)$. \blacksquare

Lemma 3. *If $\{g_n\}$ is a sequence of functions on X to \mathbf{R}^+ , and*

$$g = \sum_{n=1}^{\infty} g_n, \quad \text{then} \quad \mathcal{J}^*(g) \leq \sum_{n=1}^{\infty} \mathcal{J}^*(g_n).$$

Proof. If $\mathcal{J}^*(g_n) = +\infty$ for some n , or if the series $\sum \mathcal{J}^*(g_n)$ diverges there is nothing to prove. Otherwise, given $\epsilon > 0$, for each integer n choose $h_n \geq g_n, h_n \in L^+$ such that $\mathcal{J}^*(g_n) > \mathcal{J}(h_n) - \epsilon 2^{-n}$. Then $h = \sum h_n \in L^+$ by lemma 1, $h \geq g$ and

$$\mathcal{J}^*(g) \leq \mathcal{J}(h) = \sum \mathcal{J}(h_n) < \epsilon + \sum_{n=1}^{\infty} \mathcal{J}^*(g_n).$$

Since ϵ is arbitrary the result is proved. \blacksquare

Theorem 9.6. *Given a Daniell functional \mathcal{J} on a vector lattice L of functions on X to \mathbf{R} , the process defining a functional \mathcal{J} on the set L_1 determines a Daniell functional on a lattice L_1 which extends \mathcal{J} . Further, if $\{f_n\}$ is an increasing sequence of functions in L_1 and $f = \lim f_n$, then $f \in L_1$ if and only if $\lim \mathcal{J}(f_n)$ is finite in which case $\mathcal{J}(f) = \lim \mathcal{J}(f_n)$.*

Proof. Lemma 2(v) shows that $L_1 \supset L$ and that \mathcal{J} is an extension of \mathcal{J} . Now if $g \in L_1$ so does cg for c in \mathbf{R} since

$$c \geq 0 \Rightarrow \mathcal{J}^*(cf) = c\mathcal{J}^*(f) = c\mathcal{J}_*(f) = \mathcal{J}_*(cf),$$

$$c < 0 \Rightarrow \mathcal{J}^*(cf) = c\mathcal{J}_*(f) = c\mathcal{J}^*(f) = \mathcal{J}_*(cf).$$

Further, if f and g are both in L_1 , using lemma 2(i), \dagger

$$-\mathcal{J}_*(f+g) = \mathcal{J}^*(-f-g) \leq -\mathcal{J}(f) - \mathcal{J}(g)$$

so

$$\mathcal{J}_*(f+g) \geq \mathcal{J}(f) + \mathcal{J}(g) \geq \mathcal{J}^*(f+g);$$

and, by lemma 2(iv), $f+g \in L_1$ and

$$\mathcal{J}(f+g) = \mathcal{J}(f) + \mathcal{J}(g).$$

Thus L_1 is a real linear space, and \mathcal{J} is a linear functional on L_1 .

To prove that L_1 is a lattice it is sufficient to prove that

$$f \in L_1 \Rightarrow f_+ \in L_1.$$

\dagger As pointed out in the proof of lemma 2(i) the inequality is valid, whatever value in \mathbf{R}^* is chosen for $(f+g)(x)$ at points x where $f(x) = +\infty, g(x) = -\infty$. The proof given then shows that, for $f, g \in L_1$, $(f+g) \in L_1$ whatever values are assumed at such points.

For a fixed f in L_1 and each $\epsilon > 0$, choose functions g, h in L^+ such that $-h \leq f \leq g$ and

$$\mathcal{J}(g) < \mathcal{J}(f) + \epsilon < \infty, \quad \mathcal{J}(h) \leq -\mathcal{J}(f) + \epsilon < \infty.$$

Now $g = (g \vee 0) + (g \wedge 0)$ and $g \wedge 0 \in L^+$; so $\mathcal{J}(g \vee 0) \leq \mathcal{J}(g) - \mathcal{J}(g \wedge 0) < \infty$. Thus $g_+ = g \vee 0 \in L^+$ and $\mathcal{J}(g_+) < \infty$. Similarly, $-h_- = h \wedge 0 \in L^+$ and $h_- \leq f_+ \leq g_+$. But $(g + h) \geq 0$; and separate consideration of each possible pair of signs for g, h shows that $g_+ - h_- \leq g + h$. Hence

$$\mathcal{J}(g_+) + \mathcal{J}(-h_-) \leq \mathcal{J}(g) + \mathcal{J}(h) < 2\epsilon.$$

But $-\mathcal{J}(-h_-) \leq \mathcal{J}_*(f_+) \leq \mathcal{J}^*(f_+) \leq \mathcal{J}(g_+)$

so that $\mathcal{J}^*(f_+) - \mathcal{J}_*(f_+) < 2\epsilon$. Since ϵ is arbitrary and $\mathcal{J}(g_+)$ is finite we have $f_+ \in L_1$ as required.

Now suppose $\{f_n\}$ is an increasing sequence of functions in L_1 and $f = \lim f_n$. Then if $\lim \mathcal{J}(f_n) = +\infty$, and $g \leq f, g \in L_1$ it is clear that $\mathcal{J}(g) \leq \lim \mathcal{J}(f_n)$ since $\mathcal{J}(g)$ is finite. On the other hand if $\lim \mathcal{J}(f_n)$ is finite, put $h = f - f_1$. Then $h \geq 0$ and

$$h = \sum_{n=1}^{\infty} (f_{n+1} - f_n).$$

By lemma 3,

$$\mathcal{J}^*(h) \leq \sum_{n=1}^{\infty} \{\mathcal{J}(f_{n+1}) - \mathcal{J}(f_n)\}$$

$$= \lim \mathcal{J}(f_n) - \mathcal{J}(f_1)$$

so that $\mathcal{J}^*(f) = \mathcal{J}^*(f_1 + h) \leq \mathcal{J}^*(f_1) + \mathcal{J}^*(h) \leq \lim \mathcal{J}(f_n)$.

But $f_n \leq f$ so that $\mathcal{J}_*(f) \geq \lim \mathcal{J}(f_n)$, and we must have

$$\mathcal{J}^*(f) = \mathcal{J}_*(f) = \lim \mathcal{J}(f_n).$$

This means that, if $\lim \mathcal{J}(f_n)$ is finite, then f is in L_1 , and

$$\mathcal{J}(f) = \lim \mathcal{J}(f_n).$$

The positive functional \mathcal{J} therefore satisfies (9.4.1) and must be a Daniell functional on L_1 . **■**

Remark. There may be some functions $f: X \rightarrow \mathbb{R}^*$ which take the values $\pm \infty$ at some points but are still in L_1 . In the course of the proof we saw that it made no difference to the linear functional \mathcal{J} what value was assigned to $(bf + cg)(x)$ at points x where the usual calculation leads to $+\infty + (-\infty)$. It is in this sense that \mathcal{J} is a linear functional

on the real linear space L_1 . However, we will shortly see that all functions $f: X \rightarrow \mathbf{R}^*$ in L_1 must take finite values at 'almost all' points, so that the set where $(bf + cg)$ is not determined by the laws of algebra is always small (relative to \mathcal{J}).

Now if one starts with a Daniell functional \mathcal{J} on a vector lattice L which is already closed for monotone limits, i.e. if $\{f_n\}$ is a monotone sequence in L and $\lim \mathcal{J}(f_n)$ is finite, then $f = \lim f_n$ is in L ; the extension process defined will lead to nothing new as the part of L^+ on which \mathcal{J} is finite is in L and this will give $L = L_1$.

Daniell integral

Any Daniell functional \mathcal{J} on a vector lattice L_1 of functions on X to \mathbf{R}^* such that the limit f of a monotone sequence $\{f_n\}$ of functions in L_1 is in L_1 provided $\lim \mathcal{J}(f_n)$ is finite is called a Daniell integral.

We now see how one can obtain a theory of measure if one starts with a linear operator \mathcal{J} satisfying these conditions. The definitions are made so that the integral (in the sense of Chapter 5) with respect to the measure recovers the operator \mathcal{J} . Starting with a Daniell integral \mathcal{J} we say that a non-negative function $f: X \rightarrow \mathbf{R}^+$ is *measurable* (with respect to \mathcal{J}) if $g \in L_1 \Rightarrow f \wedge g \in L_1$. We say that a set $A \subset X$ is *measurable* (with respect to \mathcal{J}) if the indicator function χ_A is measurable; while the set A is *integrable* if $\chi_A \in L_1$. In order to ensure that the class of measurable functions and sets has useful properties we will further assume that the space X is measurable, that is, that the constant function $f(x) \equiv 1$ is measurable.

Lemma 4. *If X is measurable, then the class \mathcal{A} of sets measurable with respect to \mathcal{J} is a σ -field. If $f: X \rightarrow \mathbf{R}^+$ is any non-negative integrable function, the set $E_\alpha = \{x: f(x) > \alpha\} \in \mathcal{A}$ for all $\alpha \in \mathbf{R}$.*

Proof. Given f, g non-negative measurable functions, the lattice properties of L_1 immediately give that $f \vee g$ and $f \wedge g$ are measurable. But

$$\chi_{A \cap B} = \chi_A \wedge \chi_B, \quad \chi_{A \cup B} = \chi_A \vee \chi_B$$

so that $A, B \in \mathcal{A} \Rightarrow A \cap B$ and $A \cup B \in \mathcal{A}$. Further for any set E ,

$$g \wedge \chi_E = (g \vee 0 + g \wedge 0) \wedge \chi_E = (g \vee 0) \wedge \chi_E + g \wedge 0$$

so that if $g \in L_1$, $g \vee 0$ and $g \wedge 0 \in L_1$ and

$$(g \vee 0) \wedge \chi_{A-B} = (g \vee 0) \wedge \chi_A - (g \vee 0) \wedge \chi_{A \cap B} \in L_1,$$

$$(g \wedge 0) \wedge \chi_{A-B} = g \wedge 0 \in L_1,$$

so that $g \wedge \chi_{A-B} \in L_1$. Thus \mathcal{A} is a ring, and since $X \in \mathcal{A}$, we have proved that \mathcal{A} is a field. To show that \mathcal{A} is a σ -field one need only use

the fact that L_1 is closed for monotone limits which are bounded, since $E_n = \bigcup_{i=1}^n A_i$ is monotone and so is χ_{E_n} .

Now if $f: X \rightarrow \mathbb{R}^+$ is non-negative and is in L_1 , $E_\alpha = X$ for $\alpha \leq 0$. If $\alpha = 0$ put $h = f$; while if $\alpha > 0$ put $h = [\alpha^{-1}f - (\alpha^{-1}f) \wedge 1]$. Then $h \in L_1$, and in either case $h(x) > 0$ for $x \in E_\alpha$ and $h(x) = 0$ for $x \in X - E_\alpha$. For each integer n , put $f_n = 1 \wedge (nh)$. Then $f_n \in L_1$ and the sequence $\{f_n\}$ increases monotonely to χ_{E_α} . Hence χ_{E_α} is measurable, so E_α is measurable. \blacksquare

Theorem 9.7 (Stone). Suppose \mathcal{J} is a Daniell integral on the class L_1 of functions $f: X \rightarrow \mathbb{R}^*$, and X is a measurable set with respect to \mathcal{J} . Then

$$\mu(E) = \mathcal{J}(\chi_E) \text{ when } E \text{ is integrable,}$$

$$\mu(E) = +\infty \text{ otherwise}$$

defines a measure μ on the σ -field \mathcal{A} of measurable sets. A function $f: X \rightarrow \mathbb{R}^*$ is in L_1 if and only if it is integrable with respect to this measure μ , and

$$\mathcal{J}(f) = \int f d\mu \quad \text{for all } f \in L_1.$$

Proof. It is immediate that $\mu(\emptyset) = 0$. If B is integrable and A is measurable with $A \subset B$, the definitions ensure that A is integrable and $0 \leq \mu(A) \leq \mu(B)$. This inequality is trivially satisfied when B is measurable but not integrable, so μ is monotone on \mathcal{A} .

Now let $\{E_n\}$ be a disjoint sequence in \mathcal{A} and $E = \bigcup_{n=1}^{\infty} E_n$. If at least one of the E_n fails to be integrable, then E is not integrable and

$$\mu(E) = +\infty = \Sigma \mu(E_n). \quad (9.4.2)$$

If each of the sets E_n is integrable, then E will be integrable if and only if $\Sigma \mu(E_n) < \infty$ by theorem 9.6, since $\chi_E = \Sigma \chi_{E_n}$ and in this case

$$\mu(E) = \Sigma \mu(E_n) < \infty.$$

It is clear from the statement of theorem 9.6 that (9.4.2) will be satisfied if $\Sigma \mu(E_n) = +\infty$. Thus in all cases, μ is σ -additive on \mathcal{A} .

Now lemma 4 ensures that \mathcal{A} is a σ -field, and that any non-negative g -integrable functions is \mathcal{A} -measurable. Since each g -integrable function is the difference of two non-negative g -integrable functions it follows that any f in L_1 is \mathcal{A} -measurable.

Consider a non-negative $f: X \rightarrow \mathbf{R}^+$ in L_1 . For each pair (r, s) of positive integers put

$$E_{r,s} = \{x: f(x) > r/s\}.$$

Now $E_{r,s} \in \mathcal{A}$ and $\chi_{E_{r,s}} \in L_1$ (that is, $\mu(E_{r,s}) < \infty$) since

$$\chi_{E_{r,s}} = \chi_{E_{r,s}} \wedge \left(\frac{r}{s}f\right).$$

Put

$$f_n = \frac{1}{s} \sum_{r=1}^{s^2} \chi_{E_{r,s}}, \quad s = 2^n,$$

and note that $\{f_n\}$ is a monotone sequence in L_1 which converges to f . Hence $\mathcal{J}(f) = \lim \mathcal{J}(f_n)$. But

$$\mathcal{J}(f_n) = \frac{1}{s} \sum_{r=1}^{s^2} \mathcal{J}(\chi_{E_{r,s}}) = \frac{1}{s} \sum_{r=1}^{s^2} \mu(E_{r,s}) = \int f_n d\mu,$$

and from the definition of the integral of a non-negative \mathcal{A} -measurable function we have

$$\mathcal{J}(f) = \lim \int f_n d\mu = \int f d\mu.$$

Conversely, if $f: X \rightarrow \mathbf{R}^+$ is non-negative and integrable with respect to μ , then each of the sets $E_{r,s}$ is in \mathcal{A} and has finite μ -measure. Hence $\chi_{E_{r,s}}$ and therefore f_n are in L_1 . Since

$$\int f d\mu = \lim \int f_n d\mu = \lim \mathcal{J}(f_n) < \infty,$$

by theorem 9.6, $f = \lim f_n$ is in L_1 . This completes the representation theorem for non-negative functions. But for both the functional \mathcal{J} , and the integral with respect to μ we have a decomposition $f = f_+ - f_-$ of any integrable $f: X \rightarrow \mathbf{R}^*$ as the difference of two non-negative integrable functions, so we can deduce the representation for arbitrary integrable functions. \blacksquare

An obvious question arising is that of uniqueness for the measure μ in theorem 9.7. This cannot always be obtained, but we give an outline of the uniqueness proof under suitable conditions in exercises 9.4 (8, 9).

Exercises 9.4

1. Show that the condition (9.4.1) for a positive linear functional is equivalent to saying that, if $\{u_n\}$ is a sequence of non-negative functions in L and $\phi \in L$ satisfies $\phi \leq \sum u_n$, then $\mathcal{J}(\phi) \leq \sum \mathcal{J}(u_n)$.

2. If $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space, L is the class of μ -integrable functions and $\mathcal{J}(f) = \int f d\mu$, show that \mathcal{J} is a Daniell functional on L .

3. Let J be the class of continuous functions on \mathbf{R} to \mathbf{R} which are zero outside $[-K, K]$ for some K and put

$$\mathcal{J}(f) = \int_{-\infty}^{\infty} f(x) dx \text{ in the Riemann sense.}$$

Show that \mathcal{J} is a Daniell functional on J .

4. If \mathcal{J} is a Daniell integral defined on the class L_1 prove that

$$f \in L_1 \Rightarrow |f| \in L_1.$$

5. (Fatou for Daniell integral.) Suppose $\{f_n\}$ is a sequence of non-negative functions in L_1 . Prove that $\liminf f_n$ is in L_1 if $\liminf \mathcal{J}(f_n) < \infty$ and in this case

$$\mathcal{J}(\liminf f_n) \leq \liminf \mathcal{J}(f_n).$$

6. (Dominated convergence.) Suppose $\{f_n\}$ is a convergent sequence in L_1 such that $|f_n| \leq g$ for all n where $g \in L_1$. Then if $f = \lim f_n$, $f \in L_1$ and

$$\mathcal{J}(f) = \lim \mathcal{J}(f_n).$$

7. Suppose μ is a measure on a field \mathcal{A} of subsets of X , and L is the family of finite linear combinations of indicator functions of sets of \mathcal{A} with finite measure. Show that L is a vector lattice and if \mathcal{J} is defined on L to be integration with respect to μ , then \mathcal{J} is a Daniell functional. Discuss its extension \mathcal{J} to a Daniell integral.

8. Suppose \mathcal{J} is a Daniell functional on a vector lattice L , and \mathcal{J}' is an extension of \mathcal{J} to a Daniell functional on a vector lattice $L' \supset L$. If \mathcal{J} and \mathcal{J}' are extended to give Daniell integrals over L_1 and L'_1 show that $L'_1 \supset L_1$ and \mathcal{J}' is an extension of \mathcal{J} .

9. Suppose L is a fixed vector lattice containing the constant function 1 and \mathcal{B} is the smallest σ -field of subsets of X such that each function in L is measurable \mathcal{B} . Prove that for each Daniell integral \mathcal{J} on L_1 there is a unique measure μ on \mathcal{B} such that

$$\mathcal{J}(f) = \int f d\mu \quad \text{for all } f \in L.$$

Hint. If \mathcal{A} is σ -field of sets measurable w.r.t. \mathcal{J} , $\mathcal{A} \supset \mathcal{B}$. Existence of μ follows from theorem 9.7. To prove uniqueness it is sufficient to show that for any such μ ,

$$\mu(B) = \mathcal{J}(\chi_B) \quad \text{for all } B \in \mathcal{B}.$$

Use questions 8 and 7 above to extend the two Daniell functionals—one given and the other defined in terms of the integrals with respect to μ .

9.5* Representation of linear functionals

In this section we restrict our attention to topological spaces X which are locally compact and Hausdorff. A topological space is Hausdorff if given two distinct points $x, y \in X$, there are open sets G, H with $x \in G, y \in H, G \cap H = \emptyset$. The family of functions $f: X \rightarrow \mathbf{R}$ which are continuous on X and vanish outside a compact subset of X is called $C_0(X)$. If we define the *support* of a function $f: X \rightarrow \mathbf{R}$ to be the closure of the set $\{x: f(x) \neq 0\}$, then $C_0(X)$ is the family of those continuous functions $f: X \rightarrow \mathbf{R}$ which have compact support.

Baire sets and measure

The class of Baire sets is the smallest σ -field \mathcal{C} of subsets of X such that each function f in $C_0(X)$ is \mathcal{C} -measurable. Thus \mathcal{C} is the σ -field generated by the sets of the form $\{x: f(x) > \alpha\}$, $f \in C_0(X)$, $\alpha \in \mathbf{R}$. A measure μ is called a Baire measure on X if μ is defined on the σ -field \mathcal{C} of Baire subsets, and $\mu(K)$ is finite for each compact set K in \mathcal{C} .

Clearly $C_0(X)$ is a normed linear space if we put

$$\|f\| = \sup_{x \in X} |f(x)|,$$

and we will also use the fact that $C_0(X)$ is a vector lattice. This allows us to identify the positive linear functionals on $C_0(X)$.

Theorem 9.8 (Riesz). *Suppose X is locally compact Hausdorff, and \mathcal{J} is a positive linear functional on the space $C_0(X)$ of continuous functions $f: X \rightarrow \mathbf{R}$ with compact support. Then there is a Baire measure μ on X such that*

$$\mathcal{J}(f) = \int f d\mu \quad \text{for all } f \in C_0(X).$$

Proof. The first step is to show that \mathcal{J} must be a Daniell functional on $C_0(X)$. Suppose $f \in C_0(X)$, $\{f_n\}$ is an increasing sequence in $C_0(X)$ and $f \leq \lim f_n$. In order to prove that $\mathcal{J}(f) \leq \lim \mathcal{J}(f_n)$ it is sufficient to show that $\mathcal{J}(f) = \lim \mathcal{J}(g_n)$ where $g_n = f \wedge f_n$ so that

$$f = \lim g_n \leq \lim f_n.$$

But then, if we put $h_n = f - g_n$ we obtain a decreasing sequence of functions of $C_0(X)$ whose limit is zero. Let K be the support of h_1 , then there is a function ϕ in $C_0(X)$ which is non-negative and satisfies $\phi(x) = 1$ for $x \in K$.† For each $x \in K$, $\epsilon > 0$ there is an n_x such that $h_{n_x}(x) < \frac{1}{2}\epsilon$ and, since h_{n_x} is continuous, there is an open set G_x for which $x \in G_x$ and $h_{n_x}(t) < \epsilon$ for $t \in G_x$.

† This uses a separation property of X ; see, for example page 146 of J. L. Kelley *General Topology*, Van Nostrand (1955).

Since K is compact there is a finite subcovering G_{x_1}, \dots, G_{x_s} of K . If $N = \max [n_{x_1}, \dots, n_{x_s}]$ we have $h_n(x) < \epsilon$ for all x in K , $n \geq N$. Thus

$$0 \leq h_n < \epsilon \phi$$

so that

$$0 \leq \mathcal{I}(h_n) < \epsilon \mathcal{I}(\phi).$$

Since ϵ is arbitrary we must have $\lim \mathcal{I}(h_n) = 0$ which implies condition (9.4.1).

We can now apply theorem 9.7 to the extension \mathcal{J} of \mathcal{I} to

$$L_1 \supset C_0(X)$$

to obtain a measure μ on the σ -field \mathcal{A} which contains the Baire sets and such that, for $f \in C_0(X)$,

$$\mathcal{J}(f) = \mathcal{I}(f) = \int f d\mu.$$

By considering the above function ϕ which is in $C_0(X)$ and takes the value 1 on the compact K , we see that

$$\mu(K) = \mathcal{J}(\chi_K) \leq \mathcal{J}(\phi) = \int \phi d\mu < \infty,$$

so that the measure μ we have obtained is finite on compact sets. **■**

When X is compact, $C_0(X)$ is the same as $C(X)$ the space of continuous $f: X \rightarrow \mathbf{R}$, so that in this case the positive linear functionals on $C(X)$ correspond to finite Baire measures. Further, because of exercise 9.5 (9) there is uniqueness. This gives

Corollary. *If X is a compact topological space and $C(X)$ is the set of continuous functions $f: X \rightarrow \mathbf{R}$, then there is a (1, 1) correspondence between positive linear functionals \mathcal{J} on $C(X)$ and finite Baire measures μ on X given by*

$$\mathcal{J}(f) = \int f d\mu.$$

If we want to consider more general linear functionals on $C(X)$, it is convenient to express these as the difference of two positive linear functionals so that theorem 9.8 can be applied. This can be done for bounded linear functionals.

If L is a vector lattice of bounded functions $f: X \rightarrow \mathbf{R}$, then L is a normed linear space with $\|f\| = \sup |f(x)|$. A bounded linear functional F has a norm

$$\|F\| = \sup_{\|f\| \leq 1} |F(f)|.$$

Theorem 9.9. *Suppose L is a vector lattice of bounded functions $f: X \rightarrow \mathbf{R}$ which contains the constant function 1. Then for each bounded linear functional F on L , there are two positive linear functionals F^+ and F^- such that $F = F^+ - F^-$ and $\|F\| = F^+(1) + F^-(1)$.*

Proof. For each $f \geq 0$ in L put

$$F^+(f) = \sup_{\substack{0 \leq g \leq f \\ g \in L}} F(g).$$

Since $F(0) = 0$, $F^+(f) \geq 0$ and $F^+(f) \geq F(f)$. Further

$$F^+(cf) = cF^+(f) \quad \text{for } c \geq 0.$$

If f, g are two non-negative functions in L , such that $0 \leq \phi \leq f$, $0 \leq \chi \leq g$, then $0 \leq \phi + \chi \leq f + g$, so that $F^+(f + g) \geq F(\phi) + F(\chi)$. Taking suprema over all such ϕ, χ in L gives

$$F^+(f + g) \geq F^+(f) + F^+(g).$$

To obtain the reverse inequality consider $\chi \in L$ such that

$$0 \leq \chi \leq f + g: \quad \text{then } 0 \leq \chi \wedge f \leq f \quad \text{and} \quad 0 \leq \chi - (\chi \wedge f) \leq g$$

so that

$$\begin{aligned} F(\chi) &= F(\chi \wedge f) + F[\chi - (\chi \wedge f)] \\ &\leq F^+(f) + F^+(g) \end{aligned}$$

and taking the supremum over such χ gives

$$F^+(f + g) \leq F^+(f) + F^+(g).$$

For an arbitrary $f \in L$, let p, q be two constants such that $(f + p)$ and $(f + q)$ are both non-negative. Then

$$F^+(f + p + q) = F^+(f + p) + F^+(q) = F^+(f + q) + F^+(p)$$

so that

$$F^+(f + p) - F^+(p) = F^+(f + q) - F^+(q).$$

This means that the value of $[F^+(f + p) - F^+(p)]$ is independent of p and we can define $F^+(f)$ to be this value. Thus F^+ is now defined on L ,

$$F^+(f + g) = F^+(f) + F^+(g) \quad \text{for all } f, g \in L,$$

and

$$F^+(cf) = cF^+(f) \quad \text{for } c \geq 0, \quad f \in L.$$

But $F^+(-f) + F^+(f) = F^+(0) = 0$ so we have $F^+(-f) = -F^+(f)$ and F^+ is a positive linear functional on L .

But $F^+(f) \geq F(f)$ so that $F^- = F^+ - F$ is also a positive linear functional on L .

Now

$$\|F\| \leq \|F^+\| + \|F^-\| = F^+(1) + F^-(1).$$

To establish the opposite inequality consider functions $f \in L$ for which $0 \leq f \leq 1$. Since $|2f - 1| \leq 1$

$$\|F\| \geq F(2f - 1) = 2F(f) - F(1).$$

Taking the supremum over such f gives

$$\|F\| \geq 2F^+(1) - F(1) = F^+(1) + F^-(1). \quad \blacksquare$$

Corollary. *Let X be a compact Hausdorff space and $C(X)$ the set of continuous functions $f: X \rightarrow \mathbf{R}$. Then there is a $(1, 1)$ correspondence between finite signed Baire measures ν on X and the dual space to $C(X)$ given by*

$$F(f) = \int f d\nu.$$

Moreover, $\|F\| = |\nu|(X)$.

Proof. If one starts with a finite signed Baire measure ν , then by theorem 3.3, there is a decomposition $\nu = \nu_+ - \nu_-$ into the difference of two finite Baire measures. Clearly

$$F(f) = \int f d\nu_+ - \int f d\nu_-$$

then defines a bounded linear functional on $C(X)$ since each function f in $C(X)$ is bounded and measurable with respect to the class of Baire sets.

Conversely given a bounded linear functional F on $C(X)$, this can be decomposed by theorem 9.9 into the difference $F = F^+ - F^-$ of two positive linear functionals. Apply theorem 9.8 and corollary to find finite Baire measures μ_1, μ_2 with

$$F^+(f) = \int f d\mu_1, \quad F^-(f) = \int f d\mu_2.$$

If we put $\nu = \mu_1 - \mu_2$, then ν is a finite Baire measure and

$$F(f) = \int f d\nu.$$

Now

$$\begin{aligned} |F(f)| &\leq \int |f| d|\nu| \\ &\leq \|f\| \cdot |\nu|(X) \end{aligned}$$

so that $\|F\| \leq |\nu|(X)$. Further

$$|\nu|(X) \leq \mu_1(X) + \mu_2(X) = F^+(1) + F^-(1) = \|F\|$$

so we have $\|F\| = |\nu|(X)$.

To prove that ν is uniquely determined by F , suppose there are two signed measures ν_1, ν_2 with

$$\int f d\nu_1 = \int f d\nu_2 \quad \text{for each } f \in C(X).$$

Decompose $\lambda = \nu_1 - \nu_2$ by theorem 3.2 to give $\lambda = \lambda_+ - \lambda_-$. Then

$$\int f d\lambda_+ = \int f d\lambda_- \quad \text{for all } f \in C(X),$$

so that by the uniqueness proved in exercise 9.4 (9), $\lambda_+ = \lambda_-$ on the Baire sets. Hence $\nu_1 = \nu_2$. **■**

Exercises 9.5

1. Show that in a locally compact separable metric space the class of Baire sets is the same as the class of Borel sets.

2. Suppose μ is a Baire measure on a locally compact space X . Let H be the union of all open Baire sets G for which $\mu(G) = 0$. The complement $F = X - H$ is closed and called the *support* of μ . Prove

- (i) if G is an open Baire set and $G \cap F \neq \emptyset$ then $\mu(G) > 0$;
- (ii) if K is a compact Baire set with $K \cap F = \emptyset$, then $\mu(K) = 0$;
- (iii) if $f \in C_0(X)$ and $f \geq 0$, $\int f d\mu = 0$ if and only if $f \equiv 0$.

3. The corollary to theorem 9.9 is not valid on $C_0(X)$ for X locally compact Hausdorff. A *Radon measure* ϕ on a locally compact space is defined to be a linear functional on $C_0(X)$ which is continuous in the sense that, for each compact K , $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x)| < \delta$ for all x , with the support of f contained in K , implies that $|\phi(f)| < \epsilon$. Prove every positive linear functional is a Radon measure.

For \mathbf{R} and the usual topology define

$$\phi(f) = \sum_{r=-\infty}^{\infty} (-1)^r f(r) \quad \text{for } f \in C_0(\mathbf{R}).$$

Show that ϕ is a Radon measure, but that ϕ does not correspond to any signed Baire measure.

9.6* Haar measure

There is a general method of defining a measure on an important class of topological spaces which have the algebraic structure of a group. For notational purposes we will represent the group operation in the set X by multiplication. We do not assume that the group operation is commutative. For subsets A, B of X and an element $x \in X$ we define

$$xA = \{xy: y \in A\},$$

$$AB = \{xy: x \in A, y \in B\},$$

$$A^{-1} = \{x: x^{-1} \in A\},$$

and call xA and Ax respectively the left translation and right translation of A by x . We also require the algebraic operations to be con-

tinuous in the topology of X . The theory of Haar measure can be developed for any such topological group which is locally compact and Hausdorff, but in this section we will make the additional (unnecessary) assumption that the topology is determined by a metric ρ .

A set X is a *metric group* if X is a group and there is a metric ρ such that in (X, ρ) , the group operation is continuous. In particular

$$\left. \begin{array}{l} \lim_{n \rightarrow \infty} x_n = x_0 \\ \lim_{n \rightarrow \infty} y_n = y_0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \lim_{n \rightarrow \infty} x_n y_n = x_0 y_0, \\ \lim_{n \rightarrow \infty} x_n^{-1} = x_0^{-1}. \end{array} \right.$$

We will, for the remainder of the section, assume that X is a metric group which is locally compact in the topology of the metric.

We are interested in measures for which the translation of A by any element x leaves the measure invariant. For example, the space \mathbf{R} of real numbers is clearly a metric group with ordinary addition for the group operation. Given a set $E \subset \mathbf{R}$, and a point $x \in \mathbf{R}$, xE denotes the set of real numbers of the form $x + y$ with $y \in E$. We showed in §4.5 that Lebesgue measure in \mathbf{R} is invariant under translations in the sense that, for measurable E , $|E| = |xE|$. The notation of an invariant measure in a topological group should be thought of as a generalisation of this property of Lebesgue measure in \mathbf{R} . To be precise, a measure μ defined on the class \mathcal{B} of Borel subsets of X is called a *left Haar measure* if

- (i) μ is invariant under left translations; that is for every $E \in \mathcal{B}$, $x \in X$, $\mu(xE) = \mu(E)$;
- (ii) for every compact set C , $\mu(C) < \infty$;
- (iii) for every non-void open set G , $\mu(G) > 0$.

Conditions (ii) and (iii) eliminate such trivial measures as the zero measure, and the measure which is $+\infty$ except on the null set. A *right Haar measure* is one for which left translation invariance is replaced by invariance for right translations. We give the details of construction for a left Haar measure: obvious modifications would give the right Haar measure.

Let \mathcal{C}_0 be the class of non-empty open subsets of X whose closures are compact. The important consequence of local compactness is that every compact K in X can be covered by a finite number of sets of \mathcal{C}_0 . The sets \emptyset , X added to \mathcal{C}_0 form the class \mathcal{C} . The first step is to define a suitable set function λ on \mathcal{C} .

Suppose $H \in \mathcal{C}_0$, and G is any non-empty open set. Then

$$\mathcal{G} = \{xG: x \in \bar{H}G^{-1}\}$$

is a class of open sets covering \bar{H} since, if $y \in \bar{H}$, $g \in G$, $x = yg^{-1}$, $y = xg \in xG$. But \bar{H} is compact so there is a finite subclass of \mathcal{G} , which covers \bar{H} . Let the smallest number of sets of \mathcal{G} which cover \bar{H} be denoted

$$(H:G).$$

This is a measure of the relative sizes of H and G . It is immediate that, for $A, B, C \in \mathcal{C}_0$

$$1 \leq (A:C) \leq (A:B)(B:C).$$

Now compare all sets with some fixed $H_0 \in \mathcal{C}_0$, and put, for each non-empty open set $G, H \in \mathcal{C}_0$,

$$\lambda_G(H) = \frac{(H:G)}{(H_0:G)}.$$

Now, for fixed H , $\lambda_G(H)$ is a bounded function of G since

$$0 < \frac{1}{(H_0:H)} \leq \lambda_G(H) \leq (H:H_0). \quad (9.6.1)$$

If e is the identity element of the group X and

$$S_n = S\left(e, \frac{1}{n}\right) \quad (n = 1, 2, \dots)$$

is the open sphere centre e radius $1/n$, then for each fixed $H \in \mathcal{C}_0$

$$\lambda_{S_n}(H) \quad (n = 1, 2, \dots)$$

is a bounded sequence of real numbers. Put

$$\lambda(H) = \text{Lim } \lambda_{S_n}(H)$$

where Lim is the generalized limit defined for all sequences in \mathbf{m} using the Hahn-Banach theorem to extend the definition from \mathbf{c} to \mathbf{m} while preserving the norm (see exercise 8.4.(7)). Finally, put $\lambda(\emptyset) = 0$, $\lambda(X) = +\infty$ if X is not compact (and so not in \mathcal{C}_0).

Lemma. *The set function λ defined on \mathcal{C} has the following properties:*

(i) $0 < \lambda(H) < \infty$ for every $H \in \mathcal{C}_0$;

(ii) if $H_1, H_2 \in \mathcal{C}_0$, $d(H_1, H_2) > 0$ then

$$\lambda(H_1 \cup H_2) = \lambda(H_1) + \lambda(H_2);$$

(iii) for any $H_1, H_2 \in \mathcal{C}_0$,

$$\lambda(H_1 \cup H_2) \leq \lambda(H_1) + \lambda(H_2);$$

(iv) if $H_1, H_2 \in \mathcal{C}_0$, $H_1 \subset H_2$ then

$$\lambda(H_1) \leq \lambda(H_2);$$

(v) for any $x \in X$, $H \in \mathcal{C}_0$, $\lambda(xH) = \lambda(H)$.

Proof. By (9.6.1), $\lambda_{S_n}(H)$ is bounded below and above so that

$$0 < \frac{1}{(H_0: H)} \leq \lambda(H) \leq (H: H_0) < \infty.$$

This establishes (i). Further, for each $H \in \mathcal{C}_0$, G open, the covering ratios $(xH: G) = (H: G)$ for all $x \in X$; so that the sequence $\lambda_{S_n}(H)$ is invariant under left translations: therefore λ is also and (v) is proved.

If $H_1, H_2 \in \mathcal{C}_0$ and $d(H_1, H_2) = \eta > 0$, then for $1/n < \eta$ we must have

$$(H_1 \cup H_2: S_n) = (H_1: S_n) + (H_2: S_n),$$

$$\lambda_{S_n}(H_1 \cup H_2) = \lambda_{S_n}(H_1) + \lambda_{S_n}(H_2),$$

and (ii) is now established by taking generalised limits.

Now for any open G , and $H_1, H_2 \in \mathcal{C}_0$

$$(H_1 \cup H_2: G) \leq (H_1: G) + (H_2: G),$$

so

$$\lambda_{S_n}(H_1 \cup H_2) \leq \lambda_{S_n}(H_1) + \lambda_{S_n}(H_2);$$

this implies (iii) and a similar argument gives (iv). \blacksquare

We now define a set function μ^* for all subsets of X by

$$\mu^*(E) = \inf \sum_{i=1}^{\infty} \lambda(H_i), \quad (9.6.2)$$

where the infimum is taken over all coverings $\{H_i\}$ of E by sets in \mathcal{C} .

Theorem 9.10. *In a locally compact metric group, the set function μ^* given by (9.6.2) is a metric outer measure. The restriction μ of μ^* to the class \mathcal{B} of Borel sets is a left Haar measure.*

Proof. In the definition of outer measure given in §3.1, condition (i) is obvious, (ii) follows from (iv) of the lemma, and subadditivity (iii) follows from (9.6.2) as in the proof of theorem 4.2. Thus μ^* is an outer measure. Now suppose $E_1, E_2 \subset X$ with $d(E_1, E_2) > 0$. If $E_1 \cup E_2$ cannot be covered by a sequence from \mathcal{C}_0 , then at least one of the sets E_1, E_2 cannot be covered by such a sequence and

$$\mu^*(E_1 \cup E_2) = \mu^*(E_1) + \mu^*(E_2) \quad (9.6.3)$$

since both sides are $+\infty$. If E_1, E_2 can be covered by sequences from \mathcal{C}_0 , first choose open sets $G_1 \supset E_1, G_2 \supset E_2$ for which $d(G_1, G_2) > 0$ and let $\{H_i\}$ be a sequence of sets from \mathcal{C}_0 covering $E_1 \cup E_2$ with

$$\sum \lambda(H_i) \leq \mu^*(E_1 \cup E_2) + \epsilon.$$

For each i , put $H_i^1 = G_1 \cap H_i$, $H_i^2 = G_2 \cap H_i$.

Then by (ii) and (iv) of the lemma, for each integer i ,

$$\lambda(H_i) \geq \lambda(H_i^1 \cup H_i^2) = \lambda(H_i^1) + \lambda(H_i^2)$$

and so $\mu^*(E_1) + \mu^*(E_2) \leq \Sigma \lambda(H_i) \leq \mu^*(E_1 \cup E_2) + \epsilon$.

Since this is true for each $\epsilon > 0$, and μ^* is subadditive we have established (7.6.3) so that μ^* is a metric outer measure.

Now apply theorem 4.1 to μ^* to obtain a measure μ on a class \mathcal{M} of μ^* -measurable sets. Since μ^* is a metric outer measure, this class \mathcal{M} includes the open sets and therefore the Borel sets \mathcal{B} (see exercise 4.3 (4)); so that the restriction of μ^* to \mathcal{B} defines a measure on \mathcal{B} . If we now examine the conditions for μ to be left Haar measure we see that (v) of the lemma implies that μ^* is left translation invariant.

If K is any compact set in X , there is a finite subclass of \mathcal{C}_0 which covers K so that

$$\mu^*(K) \leq \sum_{i=1}^n \lambda(H_i) < \infty$$

so that condition (ii) for a Haar measure is satisfied. Now suppose G is any non-void open set in X . If $x \in G$, pick $\epsilon > 0$ such that $S(x, \epsilon) \subset G$ and put $\bar{E} = S(x, \frac{1}{2}\epsilon)$ so that $\bar{E} \subset G$. Since X is locally compact we may assume ϵ is small enough to make \bar{E} compact so that $\bar{E} \in \mathcal{C}_0$. If $\mu(G) = \infty$ then $\mu(G) > 0$; so we may suppose $\mu(G) < \infty$. For each $\eta > 0$ there is a sequence $\{H_i\}$ from \mathcal{C}_0 such that

$$\bigcup_{i=1}^{\infty} H_i \supset G \supset \bar{E}, \quad \Sigma \lambda(H_i) \leq \mu^*(G) + \eta.$$

But \bar{E} is compact so a finite number of the H_i must cover \bar{E} . Then if

$$\bigcup_{i=1}^m H_i \supset \bar{E},$$

$$\lambda(\bar{E}) \leq \lambda\left(\bigcup_{i=1}^m H_i\right) \leq \sum_{i=1}^m \lambda(H_i) \leq \mu^*(G) + \eta,$$

and since η is arbitrary we have

$$\mu(G) = \mu^*(G) \geq \lambda(\bar{E}) > 0$$

so that μ satisfies condition (iii) for a Haar measure. \blacksquare

Corollary. For any compact metric group X there is a left Haar measure \mathbf{P} defined on a σ -field \mathcal{F} which includes the Borel sets such that $(X, \mathcal{F}, \mathbf{P})$ is a probability space.

Proof. If X is compact, the above construction gives a left Haar measure in which

$$0 < \mu(X) < \infty$$

with μ defined on a σ -field \mathcal{F} which is complete with respect to μ . If we put

$$P(E) = \frac{\mu(E)}{\mu(X)} \quad \text{for } E \in \mathcal{F}$$

it is clear that (X, \mathcal{F}, P) is a probability space. \blacksquare

Exercises 9.6

1. Suppose Ω is the set of positive real numbers with the usual metric and multiplication for the group operation. If $(1, e)$ is the reference set H_0 used in the definition of Haar measure μ in Ω , show that

$$\mu(a, b) = \log b/a \quad \text{for each interval } (a, b) \subset \Omega.$$

(Here e is the base for Napierian logarithms.)

2. With $X = \mathbf{R}$ and addition for the group operation define Haar measure μ with $(0, 1)$ taken as the reference set H_0 . Show that μ is Lebesgue measure in \mathbf{R} .

3. Let X be the set of 2×2 matrices of the form

$$\begin{pmatrix} x & y \\ 0 & x \end{pmatrix}$$

with $x > 0$ and multiplication for the group operation. Define a metric in X by using the Euclidean metric in \mathbf{R}^2 . Show that in the topology of this metric, X is a locally compact metric group.

Define

$$F \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} = -\frac{y}{x}.$$

Map the Lebesgue–Stieltjes measure μ_F in the right half-plane $x > 0$ of \mathbf{R}^2 onto the set X , and show that the result is both a left and a right Haar measure.

4. X is the set of 2×2 matrices of the form

$$\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} \quad (x > 0)$$

metrised by the Euclidean metric in \mathbf{R}^2 . As in question 3, obtain a measure in X by mapping the Lebesgue–Stieltjes measure μ_F of question 3 into X . Show that this is a left Haar measure but not a right Haar measure.

5. If μ is a left Haar measure on X and ν is defined by $\nu(E) = \mu(E^{-1})$, show ν is a right Haar measure.

6. The left Haar measure of theorem 9.10 is regular in the sense that $\mu^*(E) = \inf\{\mu(G): G \supset E, G \text{ open}\}$.

7. Haar measure is obviously not unique since for any Haar measure $\mu, c > 0$ the measure $c\mu$ is also a Haar measure. However, on a compact metric group, with the condition $\mu(X) = 1$ it can be proved that the Haar measure is essentially unique.

8. If A, B are two compact sets with $\mu(A) = \mu(B) = 0$, does it follow that $\mu(AB) = 0$?

9. If μ is a Haar measure in X , then X has a discrete topology if and only if $\mu\{x\} \neq 0$ for at least one $x \in X$.

10. If a Haar measure μ on X is finite prove that X is compact.

11. In a locally compact metric group X show that a Haar measure μ on X is σ -finite if and only if X is σ -compact.

INDEX OF NOTATION

$A - B$, 9	l_p, l_∞ , 209
$\mathcal{B}, \mathcal{B}^n$, 43	\mathcal{L}^k , 79
\mathcal{B}^* , 103	$\mathcal{L}_F, \mathcal{L}_F^k$, 96
$C_0(X)$, 250	\mathcal{L}_∞ , 168
$C(X)$, 251	$\mathcal{L}_2(\Omega, \mu)$, 194
\mathbf{C} , 2	\bar{M}, \mathcal{M} , 166
\mathbf{C} , 46	$\bar{M}(\phi_a)$, 199
$C(a, b)$, 209	\mathbf{M}, \mathbf{m} , 46
$C^*(X)$, 48	$\mathcal{P}, \mathcal{P}^n$, 15
c , 6, 47†	\mathbf{Q} , 2
$\mathcal{C} \times \mathcal{D}$, 134	\mathbf{R}, \mathbf{R}^n , 2
$\mathcal{C} * \mathcal{D}$, 135	\mathbf{R}^* , 34, 51
$\mathcal{C}(\mathcal{F}_i; i \in I)$, 136	\mathbf{R}^+ , 51
D, D_+, D_-, D^+, D^- , 224	\mathbf{R}^I , 158
$d(x, E), d(E, F)$, 27	$S(x, r)$, 24
\bar{E} , 26	$\bar{S}(x, r)$, 25
E_x, E^y , 135	s , 46
$E \times F$, 2	\mathcal{P} , 82
\mathcal{E} , 15	$\mathcal{S}(\mathcal{R})$, 18
\mathcal{E}^n , 18	X^I , 157
f^{-1} , 4	$\{x_i\}$, 5
$f \circ g$, 4	$\{x; P\}$, 1
$f: A \rightarrow B$, 3	\mathbf{Z} , 2
$f_x(y)$, 136	a.e., 109
(f, g) , 198	diam, 27
\mathcal{F}_σ , 44	ess inf, ess sup, 167
$\mathcal{F}(A)$, 19	lim inf, lim sup, 12
$\mathcal{F}(\mathcal{C})$, 19	μ_+, μ_- , 62
\mathcal{G}_δ , 44	μ_F , 96
\mathcal{I} , 100	μf^{-1} , 154
K^* , 211	ρ_∞ , 168
\mathcal{K} , 43	ρ_1 , 174
L_1, \mathcal{L}_1 , 174	ρ_p , 175
L_p, \mathcal{L}_p , 175	$\sigma - \mathcal{R}$, 77
l_2 , 48	$\tau(x, A)$, 235

† Note that the symbol c has two distinct uses, which should not be confused.

χ_E , 12 \aleph_0 , 6 $\in, \notin, \subset, \Rightarrow$, 1 \emptyset , 2 Π , 3 \exists , 4 \sim , 6 \cup, \cap, Δ , 9 \cup, \cap , 10 $\| \cdot \|$, 45 \int , see chapter 5 \ll , 148 \vee, \wedge , 241

The symbol **□** is used to signal
the end of a proof.

GENERAL INDEX

- absolute continuity of functions, 231; of measures, 120, 148, 236
 - integration, 128
- additive set function, 51, 65, 214
- algebra of subsets, 15
- algebraic numbers, 13
- almost everywhere, 108
 - uniform convergence, 168
- approximation in measure, 84
 - to measurable functions, 131
- area, 69, 79
 - under a curve, 146
- atom, 64, 237
- axiom of choice, 7, 19, 93

- Baire sets, 132, 250
 - measure, 250
- Baire's category theorem, 42
- Banach space, 194, 209
- Bessel's inequality, 204
- Birkhoff's theorem, 190
- Borel field, 16
 - measurable function, 107, 154
 - sets, 43
- Borelian sets, 43
- bounded, 27
 - convergence theorem, 126
 - linear functional, 210, 251
 - variation, 228
- bounds, upper and lower, 21
- Brownian motion, 161

- Cantor set, function, 49, 110, 152, 229, 236
- Cantor's lemma, 41
- Carathéodory, 127
- cardinal numbers, 5, 6
- Cartesian product, 2, 38, 134
- category, first and second, 41
- Cauchy integral, 127
 - Riemann integral, 129
 - sequence, 29, 167, 171
- chain, 20
 - maximal, 21
- change of variable, 155
- class, 2
- closed linear span, 199
 - set, 25
 - sphere, 26
- closure, 26
- coarser topology, 40
- collection, 2
- compactification, 33, 34
- compactness, 29, 30, 39
 - local, 31
- complement, 9
- complete measure, 81, 109, 166
 - metric space, 29, 175, 194
 - set in a linear space, 199
- completion, 82
- composition, 4, 154
- conjugate indices, 183
 - space, 213, 215
- consistency conditions, 158
- continuous function, 35
 - set function, 56
- continuum hypothesis, 7, 58
- convergence, 180-1
 - in mean, 174
 - in measure, 171
 - in norm, 204
 - in p th mean, 174
 - of sets, 12
- countable, 7
 - basis for measure, 195, 207
- counting measure, 185
- covering, 30, 225
- cylinder set, 136, 140, 158

- Daniell extension, 242
 - functional, 241
 - integral, 101, 223, 241, 246
 - Kolmogorov theorem, 159
- decomposition, Hahn-Jordan, 61, 64
 - Lebesgue, 149, 239
- de Morgan's laws, 10
- dense, 42, 195
- density, 234
- derivate, 224
- derivative, 224, 240
 - Radon-Nikodym, 152, 223
- derived set, 26
- diameter, 27
- differentiable, 224
- discrete distribution
 - measure, 53, 80, 237
 - probability, 98
 - topology, 28
- disjoint, 11
- dissection, 102
- distance, 23, 27
- distribution function, 96-7
- domain, 3
- dominated convergence theorem, 121, 125-6, 180, 249
- dual space, 21, 215

- Egoroff's theorem, 169

- elementary event
 - figures, 17–18; length, area and volume of, 69
 - functions, 109
- empty set, 2
- enumerable, 7
- equicontinuity, 177
- equivalence relation, 6, 166
- ergodic theorems, maximal, 188; mean, 221; pointwise, 190
 - transformation, 192, 222
- essentially bounded, 167
- extended real numbers, 33
- extension of functions, 4
 - of set functions, 58
 - theorems, 65–6, 77, 244
- Fatou's lemma, 120, 249
- field, 15
- finite-dimensional distributions, inter-section property, 32
- Fourier coefficients, 203
 - series, 203
- Fubini's theorem, 143–4
- function, 3
 - space, 157
- generated ring, 17, 65
 - σ -ring, 18, 77
 - z -class, 17
- Gram–Schmidt orthogonalisation, 201
- groups, 254
- Haar measure, 255, 257, 259
- Hahn–Banach theorem, 212–13
 - decomposition, 61, 64
- Hausdorff space, 250
- Heine–Borel theorem, 30
- Hilbert cube, 48
 - space, 194, 202
- Hölder's inequality, 183, 186
- indefinite integral, 127, 149, 230, 234
- indicator function, 12
- indiscrete topology, 28
- inequalities, 183
- inner measure, 75
 - product, 198
- integrable function, 113–14, 127, 129, 246
 - set, 246
- integral, 100
 - Cauchy, 127
 - Cauchy–Riemann, 129
 - Daniell, 101, 223, 241, 246
 - Lebesgue, 124
 - Lebesgue–Stieltjes, 125
 - Riemann, 100, 128, 129, 176
- integration by parts, 157
- interior point, 28
- intersection, 9, 10
- invariant function, 190
 - measure, 90, 255
- inverse function, 4
 - image, 4
- invertible, 187
- Jacobian, 156
- Jordan decomposition, 61, 64
- jump function, 237
- Kolmogorov, 159
- Kuratowski's lemma, 21
- least upper bound axiom, 21, 30
- Lebesgue convergence theorem, 121
 - covering lemma, 38
 - decomposition theorem, 149, 239
 - density theorem, 235
 - integral, 124
 - measurable, 108
 - measure, 69, 79, 195
 - Stieltjes integral, 125; measure, 95, 198, 236
- Legendre polynomials, 202
- length, 69, 79
- limit of a sequence, 27
 - point, 26
- linear dependence, 199
 - functional, 201, 215, 241, 250
 - space, 45, 194
 - span, 199
 - subspace, 212
- Lindelöf space, 22
- Liouville's theorem, 188
- local compactness, 31, 250, 254
- majorised, *see* dominated
- mapping, 3, 153
- marginal distribution, 163
- maximal ergodic theorem, 188
- mean ergodic theorem, 221
- measurable function, 103, 107, 166, 246
 - set, 74, 79, 96, 246
 - space, 246
 - transformation, 154
- measure, 55
 - Haar, 255, 257, 259
 - Lebesgue, 69, 79, 195
 - Lebesgue–Stieltjes, 95, 198, 236
 - Radon, 254
- measure-preserving transformations, 187
- metric, 23, 185
 - group, 255
 - outer measure, 86, 88, 257
 - space, 23

- metrisable, 25
- Minkowski's inequality, 184, 186
- monotone class, 16, 79
 - class theorem, 18
 - convergence theorem, 119
 - sequence, 12
 - set function, 60
 - monotonic function, 8, 224, 226
- mutually singular measures, 149
- neighbourhood, 26
- non-atomic, 64, 238
- non-measurable set, 93
- norm, 45, 194, 211
- normal numbers, 193
- normed linear space, 44–5
- nowhere dense, 41, 49
- null set, 2
- open covering, 30
 - set, 24
 - sphere, 24
- ordered pairs, 2
- ordering, 20
- ordinate set, 145
- orthogonal system, 199, 200
- orthogonalisation, 201
- orthonormal system, 200
- outer measure, 59, 74
 - metric, 86, 88, 257
- parallelogram law, 209
- Parseval's identity, 204
- partial ordering, 20
- perfect set, 44, 49
- phase space, 187
- point of density, 235
 - of dispersion, 235
- pointwise convergence, 166
- positive linear functional, 241, 250
- probability measure, 96, 98
 - space, 96
- product field, 134
 - measure, 139, 141, 164
 - ring, 134
 - σ -field, 134
 - space, 3, 5, 134
 - z -class, 134
- projection, 40, 135
- proper subset, 1
- Rademacher functions, 202
- Radon measure, 254
 - Nikodym derivative, 152, 223; theorem, 149
- range of a function, 3
- rectangle, 134
- reflexive, 20, 218
- regular measure, 86
 - outer measure, 75
- representation of linear functionals, 250
- restriction, 4
 - of a set function, 58, 75
- Riemann integral, 100, 128, 129, 176
- Riesz–Fisher theorem, 205
- Riesz's lemma, 188
 - representation theorem, 250
- ring, 15
 - (in algebraic sense), 18
- scalar product, 198
- Schröder–Bernstein theorem, 6
- Schwarz's inequality, 184
- section, 135–6
- semi-ring, 15
- separable, 48, 187, 194
- separating functional, 219
- sequence, 4
 - monotone, 12, 18
- set, 1
 - function, 51
- shift, 193
- sigma additive (σ -additive), 54
 - algebra (σ -algebra), 16
 - compact (σ -compact), 43
 - field (σ -field), 16
 - finite (σ -finite), 59
 - ring (σ -ring), 16
- simple function, 102
- singular, 149, 238
- statistical mechanics, 188
- Stieltjes measure, 95–6, 99, 163
 - see also* Lebesgue–Stieltjes
- Stone's theorem, 247
- strong derivative, 240
- subadditive, 59, 213
- support, 250
- supremum, 21, 29
- thick, 164
- topological group, 254
 - space, 25
- transcendental numbers, 14
- transformation, 3, 89, 154
 - measure-preserving, 187
- transitive, 20
- triangle inequality, 23
- trigonometric functions, 202
- Tychonoff's theorem, 39
- uniform absolute continuity, 178
 - continuity, 37
 - convergence, 167
- union, 9, 10
- uniqueness of extension, 77

upper bound, 21

vector lattice, 241

Venn diagram, 9

Vitali covering, 225

volume, 69, 79

Von Neumann, 15

Weierstrass property, 31

well ordered, 2, 22

Wiener measure, 162

z -class, 16, 134

Zorn's lemma, 21