

Editorial Board

B. BOLLOBÁS, W. FULTON, A. KATOK, F. KIRWAN,
P. SARNAK, B. SIMON, B. TOTARO

AN OUTLINE OF ERGODIC THEORY

This informal introduction provides a fresh perspective on isomorphism theory, which is the branch of ergodic theory that explores the conditions under which two measure-preserving systems are essentially equivalent. It contains a primer in basic measure theory, proofs of fundamental ergodic theorems, and material on entropy, martingales, Bernoulli processes, and various varieties of mixing.

Original proofs of classic theorems – including the Shannon – McMillan – Breiman theorem, the Krieger finite generator theorem, and the Ornstein isomorphism theorem – are presented by degrees, together with helpful hints that encourage the reader to develop the proofs on their own. Hundreds of exercises and open problems are also included, making this an ideal text for graduate courses. Professionals needing a quick review, or seeking a different perspective on the subject, will also value this book.

STEVEN KALIKOW is a Visiting Professor in the Department of Mathematical Sciences at the University of Memphis.

RANDALL MCCUTCHEON is Assistant Professor in the Department of Mathematical Sciences at the University of Memphis.

CAMBRIDGE STUDIES IN ADVANCED MATHEMATICS

Editorial Board:

B. Bollobás, W. Fulton, A. Katok, F. Kirwan, P. Sarnak, B. Simon, B. Totaro

All the titles listed below can be obtained from good booksellers or from Cambridge University Press. For a complete series listing visit: <http://www.cambridge.org/series/sSeries.asp?code=CSAM>

Already published

- 73 B. Bollobás *Random graphs (2nd Edition)*
- 74 R. M. Dudley *Real analysis and probability (2nd Edition)*
- 75 T. Sheil-Small *Complex polynomials*
- 76 C. Voisin *Hodge theory and complex algebraic geometry, I*
- 77 C. Voisin *Hodge theory and complex algebraic geometry, II*
- 78 V. Paulsen *Completely bounded maps and operator algebras*
- 79 F. Gesztesy & H. Holden *Soliton equations and their algebro-geometric solutions, I*
- 81 S. Mukai *An introduction to invariants and moduli*
- 82 G. Tourlakis *Lectures in logic and set theory, I*
- 83 G. Tourlakis *Lectures in logic and set theory, II*
- 84 R. A. Bailey *Association schemes*
- 85 J. Carlson, S. Müller-Stach & C. Peters *Period mappings and period domains*
- 86 J. J. Duistermaat & J. A. C. Kolk *Multidimensional real analysis, I*
- 87 J. J. Duistermaat & J. A. C. Kolk *Multidimensional real analysis, II*
- 89 M. C. Golumbic & A. N. Trenk *Tolerance graphs*
- 90 L. H. Harper *Global methods for combinatorial isoperimetric problems*
- 91 I. Moerdijk & J. Mrcun *Introduction to foliations and Lie groupoids*
- 92 J. Kollár, K. E. Smith & A. Corti *Rational and nearly rational varieties*
- 93 D. Applebaum *Lévy processes and stochastic calculus (1st Edition)*
- 94 B. Conrad *Modular forms and the Ramanujan conjecture*
- 95 M. Schechter *An introduction to nonlinear analysis*
- 96 R. Carter *Lie algebras of finite and affine type*
- 97 H. L. Montgomery & R. C. Vaughan *Multiplicative number theory, I*
- 98 I. Chavel *Riemannian geometry (2nd Edition)*
- 99 D. Goldfeld *Automorphic forms and L-functions for the group $GL(n, \mathbb{R})$*
- 100 M. B. Marcus & J. Rosen *Markov processes, Gaussian processes, and local times*
- 101 P. Gille & T. Szamuely *Central simple algebras and Galois cohomology*
- 102 J. Bertoin *Random fragmentation and coagulation processes*
- 103 E. Frenkel *Langlands correspondence for loop groups*
- 104 A. Ambrosetti & A. Malchiodi *Nonlinear analysis and semilinear elliptic problems*
- 105 T. Tao & V. H. Vu *Additive combinatorics*
- 106 E. B. Davies *Linear operators and their spectra*
- 107 K. Kodaira *Complex analysis*
- 108 T. Ceccherini-Silberstein, F. Scarabotti & F. Tolli *Harmonic analysis on finite groups*
- 109 H. Geiges *An introduction to contact topology*
- 110 J. Faraut *Analysis on Lie groups: An Introduction*
- 111 E. Park *Complex topological K-theory*
- 112 D. W. Stroock *Partial differential equations for probabilists*
- 113 A. Kirillov, Jr *An introduction to Lie groups and Lie algebras*
- 114 F. Gesztesy *et al.* *Soliton equations and their algebro-geometric solutions, II*
- 115 E. de Faria & W. de Melo *Mathematical tools for one-dimensional dynamics*
- 116 D. Applebaum *Lévy processes and stochastic calculus (2nd Edition)*
- 117 T. Szamuely *Galois groups and fundamental groups*
- 118 G. W. Anderson, A. Guionnet & O. Zeitouni *Introduction to random matrices*
- 119 C. Perez-Garcia & W. H. Schikhof *Locally convex spaces over non-Archimedean valued fields*
- 120 P. K. Friz & N. B. Victoir *Multidimensional Stochastic Processes as Rough Paths*
- 121 T. Ceccherini-Silberstein, F. Scarabotti & F. Tolli *Representation Theory of the Symmetric Groups*

An Outline of Ergodic Theory

STEVEN KALIKOW
RANDALL MCCUTCHEON

University of Memphis



CAMBRIDGE
UNIVERSITY PRESS

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

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/9780521194402

© S. Kalikow and R. McCutcheon 2010

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 2010

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloguing in Publication data

ISBN 978-0-521-19440-2 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page</i> vii
Introduction	1
1 Measure-theoretic preliminaries	5
1.1 Basic definitions	5
1.2 Carathéodory's theorem, isomorphism, Lebesgue spaces	7
1.3 Properties of Lebesgue spaces, factors	11
1.4 Random variables, integration, (stationary) processes	16
1.5 Conditional expectation	23
2 Measure-preserving systems, stationary processes	26
2.1 Systems and homomorphisms	26
2.2 Constructing measure-preserving transformations	27
2.3 Types of processes; ergodic, independent and (P, T)	30
2.4 Rohlin tower theorem	32
2.5 Countable generator theorem	37
2.6 Birkhoff ergodic theorem and the strong law	39
2.7 Measure from a monkey sequence	43
2.8 Ergodic decomposition	45
2.9 Ergodic theory on L^2	48
2.10 Conditional expectation of a measure	50
2.11 Subsequential limits, extended monkey method	52
3 Martingales and coupling	55
3.1 Martingales	55
3.2 Coupling; the basics	59
3.3 Applications of coupling	62
3.4 The d_{bar} and variation distances	67
3.5 Preparation for the Shannon–Macmillan–Breiman theorem	69
4 Entropy	72
4.1 The 3-shift is not a factor of the 2-shift	72
4.2 The Shannon–McMillan–Breiman theorem	74
4.3 Entropy of a stationary process	78

4.4	An abstraction: partitions of 1	79
4.5	Measurable partitions, entropy of measure-preserving systems	82
4.6	Krieger finite generator theorem	85
4.7	The induced transformation and f_{bar}	92
5	Bernoulli transformations	96
5.1	The cast of characters	96
5.2	FB \subset IC	99
5.3	IC \subset EX	102
5.4	FD \subset IC	112
5.5	EX \subset VWB	115
5.6	EX \subset FD	121
5.7	VWB \subset IC	122
6	Ornstein isomorphism theorem	124
6.1	Copying in distribution	124
6.2	Coding	128
6.3	Capturing entropy: preparation	129
6.4	Tweaking a copy to get a better copy	135
6.5	Sinai's theorem	139
6.6	Ornstein isomorphism theorem	143
7	Varieties of mixing	146
7.1	The varieties of mixing	146
7.2	Ergodicity vs. weak mixing	147
7.3	Weak mixing vs. mild mixing	150
7.4	Mild mixing vs. strong mixing	153
7.5	Strong mixing vs. Kolmogorov	155
7.6	Kolmogorov vs. Bernoulli	162
	<i>Appendix</i>	167
	<i>References</i>	170
	<i>Index</i>	173

Preface

This book treats *isomorphism theory* – that branch of ergodic theory dealing with the question of when two measure-preserving systems are, in a certain sense, essentially equivalent. Although these topics have received fair treatment in several books, we think that the time is right for a fresh perspective. Indeed, with ergodic theory becoming more fashionable in its connections with number theory and additive combinatorics, yet also more abstract and structure-laden, it is interesting to observe the extent to which progress in its original concerns, classification of measure-preserving systems up to isomorphism, was achieved via combinatorial/probabilistic reasoning. Our hope is that the ergodic theory revival currently underway will find its way to isomorphism theory, and revitalize it as well.

We have also attempted to write a book that teaches general mathematical thinking in a unique manner. Most graduate level textbooks in pure mathematics provide detailed proofs of theorems followed by exercises. We¹ have tried to write this book in such a way as to make the proofs of the theorems themselves the exercises. Optional details, which readers may want more or less of, may be relegated to footnotes or to sections labeled “Remark” or “Comment”.

Indeed, proofs of major theorems are generally presented twice; once labeled “Idea of proof”, in which the reader is called on to flesh out the argument from a very basic outline, then again with the label “Sketch of proof”, in which more details are given. We consider it important that the reader attempt to work through the “Idea” section before or instead of the “Sketch”. This reading strategy benefits:

- (a) the serious beginner wishing to work out the proofs on their own, with helpful guidance from the book;
- (b) the student wanting a basic overview of ergodic theory, who doesn’t want to be overburdened with notation while trying to understand the ideas;
- (c) the professional mathematician who is familiar with the material, but wants a quick review or different perspective on it.

¹ The first author has circulated a rough precursor to this book as an MS Word document on the Internet for a number of years; the pedagogic philosophy employed here (and all original proofs of major theorems) are due to him. R.M.

This book differs from other books of its type not only in the careful compartmentalization of detail; much in our approach to classical theorems (those of Shannon–McMillan–Breiman and Ornstein, for example) would be considered unorthodox, however presented. Indeed, our aim is not to present “distinctively modern”, “slick” or “book” proofs. Our aim is to present *our* proofs, stumbled upon by active engagement with the subject matter. Our hope is that readers will follow our example, and come to favor their own proofs as well.

Introduction

Note: this introduction is written in an intuitive style, so a scientifically oriented non-mathematician might get something out of it. It is the only part of the book that requires no mathematical expertise.

Question: *What is ergodic theory?*

Let's start with two examples.

Example 1: Imagine a potentially oddly shaped billiard table having no pockets and a frictionless surface. Part of the table is painted white and part of the table is painted black. A billiard ball is placed in a random spot and shot along a trajectory with a random velocity. You meanwhile are blindfolded and don't know the shape of the table. However, as the billiard ball careens around, you receive constant updates on when it's in the black part of the table, and when it's in the white part of the table. From this information you are to deduce as much as you can about the entire setup: for example, whether or not it is possible that the table is in the shape of a rectangle.

Example 2: (This example is extremely vague by intention.) Imagine you are receiving a sequence of signals from outer space. The signal seems to be in some sense random, but there are recurring patterns whose frequencies are stationary (that is, do not alter over time). We are unable to detect a precise signal but we can encode it by interpreting five successive signals as one signal: unfortunately, this code loses information. Furthermore, we make occasional mistakes. We wish to get as much knowledge as possible about the original process.

Measure preserving transformations. The subject matter encompassing the previous two examples is called ergodic theory. Ergodic theory models situations (like the examples) under the abstraction of *measure-preserving transformations*. To understand that concept, we need to understand what *measures* are and what *transformations* are.

A measure is a concept of size that tells you how big a set is, or, in the language of probability, how probable an event is. (A probability space is a space whose total measure is equal to 1.) It has to act like a notion of size should: the measure of the union of two disjoint sets has to equal the sum of the measures of the sets, for example. A transformation is a way of mapping a space to itself

by assigning one point to another. In many modeling applications, the transformation indicates evolution in time: for example, it may map the position and direction of a billiard ball at the current time to the position and direction of the ball one second later.

Ergodic theory is the study of transformations on probability spaces that preserve measure. So, for example, if a set A of points has measure $\frac{1}{3}$, and a transformation T is measure-preserving, then the set of points which are mapped into A by T will also have measure $\frac{1}{3}$. When the measure is interpreted as a probability, the measure-preserving property indicates the stationarity or time-invariance of the expected frequencies of certain events (like the probability that the billiard ball of Example 1 lies in the white part of the table).

Processes. When you apply a transformation over and over, checking after each application whether some event has occurred and recording the result, you get a *process*. The language of processes and the language of transformations are really just two different ways of describing the same thing. You can get a process out of Example 1 if you record at one-second intervals, by writing down either B or W , the location of the moving billiard ball. For example, the output $BBWWB\dots$ represents the ball having been in the black, black, white, white and black areas at times 0, 1, 2, 3 and 4, respectively. You can also get processes out of real world systems, without foreknowledge of any transformation acting. For example, say you record each day at noon whether it's rainy or sunny by writing down R or S . If you did this every day into both the future and the past, you would output a doubly infinite string of R s and S s, thus: $\dots SRRS(S)RRRS\dots$. Here the parentheses identify the current day (it is sunny today, will be rainy tomorrow, and was sunny yesterday, etc.).

To transfer this example to the language of transformations, note that the set of all doubly infinite strings of R s and S s forms a space, and a natural transformation of this space is the *shift*, which moves time forward one day. (Hence the shift takes the above sequence to $\dots SRRSS(R)RRS\dots$) An appropriate measure can be derived from the probabilities of rain and sun respectively on the various days. This measure will be preserved by the shift precisely when the original process was stationary.

In this book, we will usually study measure-preserving transformations using the language of stationary processes. Here is a summary of the important concepts and theorems we will cover.

- (1) *Isomorphism*: Suppose that in Example 1 we were to change which part of the table is painted white and which is painted black. Then you would have a different process. But our new process could end up being equivalent to the original process in the sense that if you know the output of either process infinitely far into both the past and future, it would tell you the

output of the other process. Very roughly, one says that two processes are *isomorphic* when there is a nice way to map outputs of one to outputs of the other so that they determine each other. In general, determining whether there is such a map can be nearly impossible; much of this book is about ways to do it in a few cases.

- (2) *Ergodicity*: An *ergodic* process or transformation is one that cannot be expressed as a combination of two simpler processes (or transformations). For example, consider the process that picks a random person and then spits out an enormous sequence of *Ls* and *Rs* according to which hand that person uses to twist open all the doorknobs they encounter their whole life. That process is certainly not going to be ergodic because the character of the output will be divided in a predictable way according to whether the person chosen is left-handed or right-handed. Assuming the proportion of left-handed people in the general population to be 0.09, the whole process would then be expressible as 0.09(left-hand process) + 0.91 (right-hand process).
- (3) *Birkhoff ergodic theorem*: When an ergodic transformation is repeatedly applied to form an ergodic process, then with probability 1, the frequency of time an output of that process spends in a given set is the measure of that set, e.g. if the measure of a set is $\frac{1}{3}$, then it will spend (in an asymptotic sense) $\frac{1}{3}$ of the time in that set.
- (4) *Rohlin tower theorem*: Fix an arbitrary positive integer, say 678. For any measure-preserving transformation T that does not simply rotate finite sets of points around, you can break almost the whole space into 678 equally sized disjoint sets A_1, \dots, A_{678} such that if you arrange the sets as the rungs of a ladder, the transformation consists in simply walking up the ladder; that is, $TA_i = A_{i+1}$.
- (5) *Shannon–McMillan–Breiman theorem*: Consider an ergodic process that spits out doubly infinite strings of *as* and *bs*. If you pick a random doubly infinite string, then with probability 1, when you look at its sequence of finite initial strings (e.g. *a*, *ab*, *abb*, *abba*, ...), that sequence will have probabilities that asymptotically approach a fixed rate of exponential decay. Moreover, that rate of decay will not depend on the sequence you choose.
- (6) *Entropy*: The exponential rate of decay just mentioned is called the *entropy* of the process. Recall that essentially all of the doubly infinite strings have the same exponential decay rate. Call the ones that do *reasonable names*. Then the number of reasonable names is approximately equal to the reciprocal of this exponentially decaying probability, that is, it is a quantity that increases at a fixed exponential rate. Thus entropy can also be thought of as

being the asymptotic exponential growth rate of the number of reasonable names.

- (7) *Kolmogorov entropy invariance*: Any two isomorphic processes must have the same entropy. This provides a quick way to identify two processes as *not* being isomorphic, namely, having different entropies.
- (8) *Independent process*: A stationary process on an alphabet in which the next letter to come in the output string is always independent of the ones that came previously is called an *independent process*. For example, repeatedly rolling a die (even a loaded die) gives an independent process.
- (9) *Ornstein isomorphism theorem*: Says that two stationary independent processes are isomorphic if and only if they have the same entropy. Indeed, the standard proofs of the theorem say even more, as they give a condition which is natural to check in many cases such that any two processes that are of equal entropy and satisfy the condition must be isomorphic. This has led to the surprising realization that a great many classes of measure-preserving systems that don't seem at all similar to die rolling or coin tossing are in fact isomorphic to independent processes.

The above list spans the core topics of isomorphism theory. In the final chapter of the book, we touch briefly on additional topics in both isomorphism theory and ergodic theory, more broadly construed. In an appendix, we list some of our favorite open problems.

1

Measure-theoretic preliminaries

1. Discussion. In this opening chapter, we offer a review of the basic facts we need from measure theory for the rest of the book (it doubles as an introduction to our pedagogic method). For readers seeking a true introduction to the subject, we recommend first perusing, e.g. Folland (1984); experts meanwhile may safely jump to Chapter 2.

2. Comment. When an exercise is given in the middle of a proof, the end of the exercise will be signaled by a dot: •

The conclusion of a proof is signaled by the box sign at the right margin, thus: □

1.1. Basic definitions

In this subchapter, we discuss algebras, σ -algebras, generation of a σ -algebra by a family of subsets, completion with respect to a measure and relevant definitions.

3. Definition. Let Ω be a set. An *algebra* of subsets of Ω is a non-empty collection \mathcal{A} of subsets of Ω that is closed under finite unions and complementation. A σ -*algebra* is a collection \mathcal{A} of subsets of Ω that is closed under countable unions and complementation.

4. Comment. Every algebra of subsets of Ω contains the *trivial algebra* $\{\emptyset, \Omega\}$.

5. Exercise. Let Ω be a set and let \mathcal{C} be a family of subsets of Ω . Show that the intersection of all σ -algebras of subsets of Ω containing \mathcal{C} is itself a σ -algebra.

6. Definition. Let Ω be a set and let \mathcal{C} be a family of subsets of Ω . The σ -*algebra generated by* \mathcal{C} is the intersection of all σ -algebras of subsets of Ω containing \mathcal{C} .

7. Definition. Let \mathcal{A} be an algebra of subsets of Ω . A *premeasure* on \mathcal{A} is a finitely additive set function p taking \mathcal{A} to the non-negative reals that also

never violates countable additivity except for “undefined” cases caused by \mathcal{A} not being a σ -algebra.²

If \mathcal{A} is a σ -algebra, then p is called a *measure*.

8. Definition. Let Ω be a set, and let \mathcal{A} be a σ -algebra of subsets of Ω . The pair (Ω, \mathcal{A}) is called a *measurable space*, and the members of \mathcal{A} are called *measurable sets*. Next let μ be a probability measure defined on \mathcal{A} ; that is, a measure satisfying $\mu(\Omega) = 1$. The triple $(\Omega, \mathcal{A}, \mu)$ is called a *probability space*.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. An *event* is a measurable set, that is, a member of \mathcal{A} . Two events A and B are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A *null set* is a set $A \in \mathcal{A}$ with $\mu(A) = 0$. \mathcal{A} is said to be *complete with respect to μ* , or $(\Omega, \mathcal{A}, \mu)$ is simply said to be *complete*, if all subsets of null sets are measurable (and hence null sets).

9. Exercise. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and suppose that \mathcal{B} is not complete with respect to μ . Let $\mathcal{A} = \{B \cup C : B \in \mathcal{B}, \text{ there exists a null set } D \text{ with } C \subset D\}$. Extend μ to \mathcal{A} by the rule $\mu(B \cup C) = \mu(B)$ for the relevant cases. Show that this extension is well defined and that $(\Omega, \mathcal{A}, \mu)$ is a complete probability space.

10. Definition. The *completion* of a probability space $(\Omega, \mathcal{B}, \mu)$ is the probability space $(\Omega, \mathcal{A}, \mu)$ constructed in the previous exercise.

11. Definition. If A and B are sets, the *symmetric difference* of A and B is the set of points in A or B but not both. We denote the symmetric difference by $A \Delta B$.

12. Definition. Suppose that $(\Omega, \mathcal{A}, \mu)$ is a complete measure space and suppose that $\mathcal{C} \subset \mathcal{A}$ is a family of measurable sets. We say that \mathcal{C} *generates \mathcal{A} mod zero* if for every $A \in \mathcal{A}$ there exists B in the σ -algebra \mathcal{B} generated by \mathcal{C} such that $\mu(A \Delta B) = 0$.

13. Comment. Notice that this does not imply that $\mathcal{B} = \mathcal{A}$.

14. Definition. Let Ω be a set and denote its power set by $\mathcal{P}(\Omega)$. An *outer measure* on Ω is a non-increasing, countably sub-additive set function μ^* from $\mathcal{P}(\Omega)$ to the non-negative reals taking the empty set to zero.³

² That is to say, $p : \mathcal{A} \rightarrow [0, \infty]$ and if $(A_i)_{i=1}^{\infty} \subset \mathcal{A}$ is pairwise disjoint with $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ then $p(A) = \sum_{i=1}^{\infty} p(A_i)$. Notice that $p(\emptyset) = 0$.

³ That is, $\mu^* : \mathcal{P}(\Omega) \rightarrow [0, \infty]$ with $\mu^*(\emptyset) = 0$, $\mu^*(A) \leq \mu^*(B)$ whenever $A \subset B$, and $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$ for any sequence $(A_i)_{i=1}^{\infty} \subset \mathcal{P}(\Omega)$.

1.2. Carathéodory's theorem, isomorphism, Lebesgue spaces

In this subchapter we develop the machinery for constructing probability spaces and determining when they are essentially the same. We use this machinery to construct Lebesgue measure on the unit interval and define Lebesgue spaces to be those spaces isomorphic to an interval space.

15. Convention. If $(x_n)_{n=1}^\infty$ is a sequence, we use the notation $(x_n) \subset X$ to relate the fact that $x_n \in X$ for all n .

16. Theorem. (Carathéodory; see e.g. Folland 1984, Theorem 1.11.) Let Ω be a set, \mathcal{A} an algebra of subsets of Ω and p a premeasure on \mathcal{A} for which $p(\Omega) = 1$. For every $B \subset \Omega$ let

$$\mu^*(B) = \inf \left\{ \sum_{i=1}^{\infty} p(A_i) : (A_i)_{i=1}^{\infty} \subset \mathcal{A}, B \subset \bigcup_{i=1}^{\infty} A_i \right\}.$$

Let $\mathcal{B} = \{B \subset \Omega : \mu^*(B) + \mu^*(B^c) = 1\}$. Then μ^* is an outer measure on Ω which agrees with p on \mathcal{A} , \mathcal{B} is a σ -algebra containing \mathcal{A} , and the restriction μ of μ^* to \mathcal{B} is a measure.

(Proof omitted.)

17. Exercise. Show that the measure space arrived at in an application of Carathéodory's theorem is complete.

We now give a couple of applications of Carathéodory's theorem.

18. Definition. Let Λ be a countable set and let $\Omega = \Lambda^{\mathbf{Z}}$. A *cylinder set* is a subset of Ω you get by specifying values for finitely many (possibly zero) coordinates.⁴ The *support* of a cylinder set is the set of coordinates whose values are specified.⁵

19. Example. The set of $(x_i)_{i=-\infty}^\infty \in \Omega$ such that $x_0 = a$, $x_{17} = b$ and $x_{-2} = c$ is a cylinder set. The support of this cylinder set is $\{-2, 0, 17\}$.

20. Theorem. (Tychonoff; see e.g. Folland 1984, Theorem 4.43.) Let X_i be compact topological spaces, $i \in \mathcal{I}$, and let $X = \prod_{i \in \mathcal{I}} X_i$. Then X is compact in the product topology.

21. Exercise. Prove Tychonoff's theorem. □

⁴ To be more precise: for $r \geq 0$ an integer, $f_1, f_2, \dots, f_r \in \mathbf{Z}$ and $\lambda_1, \dots, \lambda_r \subset \Lambda$, set $C = C(f_1, \dots, f_r, \lambda_1, \dots, \lambda_r) = \{(x_i)_{i=-\infty}^\infty \in \Omega : x_{f_j} = \lambda_j, 1 \leq j \leq r\}$. C is a cylinder set.

⁵ So the support of the cylinder set defined in the previous footnote is $\{f_1, f_2, \dots, f_r\}$.

22. Exercise.

- (a) Show that the family \mathcal{A} of unions of finite (possibly empty) collections of cylinder sets in $\Omega = \Lambda^{\mathbb{Z}}$ forms an algebra. Then show that if Λ is finite and $(A_i)_{i=1}^{\infty}$ are pairwise disjoint members of \mathcal{A} whose union is a member of \mathcal{A} then only finitely many of the A_i are non-empty. *Hint: Apply Tychonoff's theorem to $\Lambda^{\mathbb{Z}}$.*
- (b) (See, e.g. McCutchen 1999, Theorem 3.2.4.) If Λ is finite, any finitely additive set function p from cylinder sets to the non-negative reals⁶ taking Ω to 1 is extendable to a premeasure on \mathcal{A} which may thereby be extended to a measure by Carathéodory's theorem.⁷

23. Exercise. Let $\Omega = [0, 1)$ and denote by \mathcal{A} the set of unions of finite (possibly empty), pairwise disjoint families of half-open intervals $[a, b) \subset [0, 1)$. Show that \mathcal{A} is an algebra of sets. Put $p([a, b)) = b - a$. Show that p has a unique extension to a premeasure on \mathcal{A} .

24. Definition. The outer measure μ^* you get by applying Carathéodory's theorem to the premeasure p of the foregoing exercise is called *Lebesgue outer measure*, which we denote by m^* . We denote by \mathcal{L} the σ -algebra \mathcal{B} coming from Carathéodory's theorem; members of \mathcal{L} are called *Lebesgue measurable sets*. The restriction of m^* to \mathcal{B} is called *Lebesgue measure*, which we denote by m .

25. Remark. Although we've only defined Lebesgue measure, Lebesgue measurable sets, etc. here on the unit interval, one can of course extend this to the whole line in the obvious way; readers should convince themselves of this.

26. Definition. For $A \subset \mathbf{R}$, define the *Lebesgue inner measure* of A to be the quantity $m_*(A) = \sup\{m(B) : B \in \mathcal{L}, B \subset A\}$.⁸

27. Exercise. Prove that for any set A and any interval I , $m_*(A) = |I| - m^*(I \setminus A)$.

⁶ In other words, if C_1, \dots, C_r are pairwise disjoint cylinder sets whose union is a cylinder set C , then $p(C) = \sum_{i=1}^r p(C_i)$.

⁷ It is instructive, and the reader is encouraged, to explore just what such a finitely additive function looks like. For a quick example, suppose that $\Lambda = \{a, d\}$. The cylinder set C_1 that sees the occurrence of "add" at the zero place (that is, $C(0, 1, 2, a, d, d)$) and the cylinder set C_2 that sees "ada" at the zero place are disjoint and their union is the cylinder set C_3 that sees "ad" at the zero place; hence any premeasure p must satisfy $p(C_1) + p(C_2) = p(C_3)$, but these are the only sorts of conditions.

⁸ The reader should check that this supremum is a maximum; i.e. it is attained.

28. Exercise. Show that for any Lebesgue measurable set $A \subset [0, 1]$, $m(A) = \sup\{m(K) : K \subset A, K \text{ is closed}\}$ and $m(A) = \inf\{m(U) : A \subset U, U \text{ is open}\}$. (We may sometimes say, accordingly, that Lebesgue measure is *inner regular* with respect to closed sets and *outer regular* with respect to open sets.)

29. Definition. Let (Ω, \mathcal{A}) and (Ω', \mathcal{A}') be measurable spaces. A function $T : \Omega \rightarrow \Omega'$ satisfying $T^{-1}A' \in \mathcal{A}$ for every $A' \in \mathcal{A}'$ is said to be $(\mathcal{A}, \mathcal{A}')$ -*measurable*, or simply *measurable* when \mathcal{A} and \mathcal{A}' are understood. Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ be probability spaces. A measurable function $T : \Omega \rightarrow \Omega'$ satisfying $\mu(T^{-1}A') = \mu'(A')$ for every $A' \in \mathcal{A}'$ is said to be *measure-preserving*.

30. Theorem. (*Urysohn's lemma*; see e.g. Dudley 2002, Lemma 2.6.3.) Let X be a normal topological space, and let A and B be disjoint closed subsets of X . There exists a continuous function $f : X \rightarrow [0, 1]$ such that $f(x) = 0$ for all $x \in A$ and $f(x) = 1$ for all $x \in B$.

(Proof omitted.)

31. Exercise. Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ be measure spaces and suppose that \mathcal{A}' is generated by a family of sets \mathcal{B} . Let $T : \Omega \rightarrow \Omega'$. Show that:

- (a) if $T^{-1}B \in \mathcal{A}$ for every $B \in \mathcal{B}$ then T is measurable;
- (b) if $\mu(T^{-1}B) = \mu'(B)$ for every $B \in \mathcal{B}$ then T is measure-preserving.

32. Definition. Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ be probability spaces and suppose $\pi : \Omega \rightarrow \Omega'$ is a measure-preserving transformation. We say that π is a *homomorphism*, or a *factor map*, and that $(\Omega', \mathcal{A}', \mu')$ is a *factor* of $(\Omega, \mathcal{A}, \mu)$.

33. Definition. Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ be probability spaces and suppose $T : \Omega \rightarrow \Omega'$ is a homomorphism. If there exist full measure sets⁹ $X \subset \Omega$ and $X' \subset \Omega'$ such that the restriction of T to X is a bijection to X' , and $T^{-1} : X' \rightarrow X$ is measurable, then we will say that T is an *isomorphism*, and that the spaces $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ are *isomorphic*.

34. Comment. When two spaces $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ are isomorphic, then, once appropriate null sets are disregarded, they are “essentially the same space”. In other words, they are in fact the same, up to relabeling.

35. Exercise. Show that “is isomorphic to” is an equivalence relation and that completeness is an isomorphism invariant.

⁹ A set $X \subset \Omega$ is said to be of full measure if $\mu(\Omega \setminus X) = 0$.

36. Convention. Our attitude toward null sets is that they “don’t count”. Accordingly we will assume that all probability spaces are complete.

Without this convention, $([0, 1], \mathcal{L}, m)$ would not be isomorphic to $([0, 1], \mathcal{B}, m)$, where \mathcal{B} is the σ -algebra of Borel sets (that is, the σ -algebra generated by the open sets) for the rather uninteresting (though non-trivial) reason that \mathcal{L} has more null sets than \mathcal{B} does.

37. Definition. An *interval space* consists of an interval $[0, t]$ equipped with Lebesgue measure, where $0 \leq t \leq 1$, to which are appended countably many points having a combined positive measure $1 - t$; *with atoms* if $t < 1$, *without atoms* if $t = 1$.¹⁰

38. Exercise. Show that an interval space is a complete probability space.

39. Definition. A *Lebesgue space* is a probability space that is isomorphic to some interval space. A Lebesgue space is *non-atomic* if it is isomorphic to $([0, 1], \mathcal{L}, m)$.

40. Remark. A classic reference in the theory of Lebesgue spaces, axiomatically defined, is Rohlin (1952), though this and most of the literature on axiomatic treatments is fraught with vagueness and ambiguity (if not confusion), due in part to a cavalier attitude toward sets of measure zero. (For an interesting and well-motivated modern axiomatic treatment, see Rudolph (1990).) An arguably more sensible theory of spaces measurably isomorphic to the unit interval is that of regular Borel probability measures on Polish spaces; however, we are choosing to skirt most of the issues entirely by simply defining Lebesgue spaces to be those that are isomorphic to an interval space. (Not all of the issues: see the axiomatic criterion in Theorem 53 below.)

There are tremendous technical advantages to doing analysis on Lebesgue spaces. Following standard ergodic theory practice, we shall deal almost exclusively with non-atomic Lebesgue spaces in this book. Since, in essence, the only non-atomic Lebesgue space is the unit interval, one can be deceived into thinking that this is unduly restrictive. However, the concept is actually quite general. Indeed, just about every probability space you are likely to encounter is Lebesgue or at least has a Lebesgue completion; in particular, spaces derived from completing regular Borel measures on compact metrizable spaces are Lebesgue; non-Lebesgue spaces are pathological examples, generally deriving

¹⁰ So, let $\theta = [0, t] \cup C$, where C is a countable set whose intersection with $[0, t]$ is empty, let $f : C \rightarrow [0, 1 - t]$ be a function satisfying $\sum_{c \in C} f(c) = 1 - t$, let \mathcal{A} consist of all $L \cup D$, where L is a Lebesgue measurable subset of $[0, t]$ and D is any subset of C , and for $A = L \cup D \in \mathcal{A}$, put $\mu(A) = m(L) + \sum_{c \in D} f(c)$. $(\theta, \mathcal{A}, \mu)$ is an interval space.

from non-separable topologies or via constructions requiring the axiom of choice.

1.3. Properties of Lebesgue spaces, factors

In this subchapter we collect facts about probability spaces for which the hypothesis that you are dealing with a Lebesgue space is needed. Among these is the fact that any factor of a Lebesgue space is a Lebesgue space.

41. Exercise. Let $(\Omega, \mathcal{A}, \mu)$ be a non-atomic Lebesgue space. Let $\epsilon > 0$ and suppose $A \in \mathcal{A}$ with $\mu(A) > \epsilon$. Then there exists $B \in \mathcal{A}$ with $B \subset A$ satisfying $\mu(B) = \epsilon$.

We need the following classical theorem.

42. Theorem. (*Lusin; see e.g. Folland 1984, Theorem 7.10.*) Let $f : [0, 1] \rightarrow \mathbf{R}$ be Lebesgue measurable. For every $\epsilon > 0$ there exists a continuous function $g : [0, 1] \rightarrow \mathbf{R}$ such that $m(\{x : f(x) \neq g(x)\}) < \epsilon$.

Sketch of proof.

43. Exercise. Show that we may assume without loss of generality that $0 < f(x) < 1$ for all $x \in [0, 1]$. *Hint: Consider the function $h(x) = \frac{1}{2} \left(1 + \frac{f(x)}{1 + \sup(|f|)} \right)$.* •

Suppose, then, that $0 < f(x) < 1$. For every $x \in [0, 1]$, write $f(x) = \sum_{n=1}^{\infty} a_n(x) 2^{-n}$ for the binary decimal expansion of $f(x)$ ¹¹ and let $H_n = \{x \in [0, 1] : a_n(x) = 1\}$. The sets H_n are clearly Lebesgue measurable. Accordingly, by Exercise 28 one may approximate H_n by an open set U_n and by a closed set K_n such that $K_n \subset H_n \subset U_n$, with $m(U_n \setminus K_n) < \delta_n$, where $\sum_{n=1}^{\infty} \delta_n < \epsilon$. Pick, by Urysohn's lemma, continuous functions h_n satisfying $h_n(x) = 1$ for $x \in K_n$ and $h_n(x) = 0$ for $x \in U_n^c$. Let $g(x) = \sum_{n=1}^{\infty} h_n(x) 2^{-n}$.

44. Exercise. Show that g is continuous and that $m(\{x : f(x) \neq g(x)\}) < \epsilon$. □

Below, we will see that a measure-preserving bijection T between probability spaces needn't be an isomorphism (due to non-measurability of T^{-1}) even when the range space is Lebesgue. The content of the next theorem is that this can't happen when the domain and range are each Lebesgue.

¹¹ There is an issue about uniqueness here for countably many cases but one can agree to choose in these cases in some canonical fashion; for example, one can agree to avoid expansions having only finitely many occurrences of 0 where possible.

45. Theorem. Let $(\Omega, \mathcal{A}, \mu)$ and (X, \mathcal{B}, ν) be Lebesgue spaces and suppose $T : \Omega \rightarrow X$ is injective and measure preserving. Then TA is measurable for every measurable set A .

Sketch of proof.

46. Exercise. Prove that it is sufficient for the general case to establish the conclusion when both $(\Omega, \mathcal{A}, \mu)$ and (X, \mathcal{B}, ν) are $([0, 1], \mathcal{L}, m)$. •

So, let $T : [0, 1] \rightarrow [0, 1]$ be injective and measure-preserving. Let $A \in \mathcal{L}$. We must show that $TA \in \mathcal{L}$.

47. Exercise. It is sufficient to show that for arbitrary $\epsilon > 0$, $m^*(TA) < m_*(TA) + 2\epsilon$. •

Let $\epsilon > 0$ be arbitrary. By Lusin's theorem, there is a continuous function $g : [0, 1] \rightarrow [0, 1]$ such that the set $B = \{x : Tx \neq g(x)\}$ satisfies $m(B) < \epsilon$. ϵ -approximate $A \setminus B$ and $A^c \setminus B$ from the inside by closed sets K_1 and K_2 , respectively.

48. Exercise. Show that $g(K_1)$ and $g(K_2)$ are measurable. *Hint: recall that the continuous image of a compact set is compact.* Show that $m(K_1) + m(K_2) > 1 - 2\epsilon$. Finally show that $g(K_1) \subset TA \subset g(K_2)^c$ and complete the proof. □

49. Exercise. Obtain the following more general version of Theorem 45 by adapting its proof. Let $(\Omega, \mathcal{A}, \mu)$ and (X, \mathcal{B}, ν) be Lebesgue spaces and suppose $T : \Omega \rightarrow X$ is measure-preserving. Prove that if for some $B \subset X$, $T^{-1}(B)$ is measurable, then B is measurable.

We need a way of verifying that spaces we construct are Lebesgue spaces.

50. Definition. Let Ω be a set. A sequence $(C_i)_{i=1}^\infty$ of subsets of Ω is said to *separate points* on Ω if for every $x, y \in \Omega$ with $x \neq y$ there is some $i \in \mathbf{N}$ such that either $x \in C_i$ and $y \in C_i^c$ or $y \in C_i$ and $x \in C_i^c$. A sequence $(P_i)_{i=1}^\infty$ of countable partitions of Ω separates points if a sequence consisting of all the cells of all the P_i separates points.

51. Exercise. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Show that $\rho(A, B) = \mu(A \Delta B)$ defines a metric on \mathcal{A} , modulo identification of sets differing by a null set.

52. Definition. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. A property P predicated of points $\omega \in \Omega$ is said to hold *almost everywhere*, or *a.e.*, if the set of x for which $P(x)$ fails has zero measure.