

Real Analysis and Probability

R O B E R T B . A S H

University of Illinois

ACADEMIC PRESS

New York San Francisco London
A Subsidiary of Harcourt Brace Jovanovich, Publishers

COPYRIGHT © 1972, BY ACADEMIC PRESS, INC.

ALL RIGHTS RESERVED

**NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM,
BY PHOTOSTAT, MICROFILM, RETRIEVAL SYSTEM, OR ANY
OTHER MEANS, WITHOUT WRITTEN PERMISSION FROM
THE PUBLISHERS.**

**ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003**

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DD

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 75-182642

AMS(MOS) 1970 Subject Classifications: 28-01, 60-01

PRINTED IN THE UNITED STATES OF AMERICA

Preface

This book, the first of a projected two volume series, is designed for a graduate course in modern probability. The first four chapters, along with the Appendix: On General Topology, provide the background in analysis needed for the study of probability. This material is available as a separate book called “Measure, Integration, and Functional Analysis.”

Chapter 5 offers a summary of an undergraduate probability course, from a measure-theoretic point of view. Chapter 6 is concerned with the general concept of conditional probability and expectation. The approach to problems that involve conditioning, given events of probability zero, is the gateway to many areas of probability theory. Chapter 7 deals with strong laws of large numbers, first from the classical viewpoint, and then via martingale theory. Basic properties and applications of martingale sequences are developed systematically. Chapter 8 considers the one-dimensional central limit problem, emphasizing the fundamental role of Prokhorov’s weak compactness theorem. Projected topics for the second volume include L^2 theory, prediction, renewal theory, ergodic theory, and general theory of continuous time stochastic processes.

Chapters 3 and 4 are necessary for the general theory of stochastic processes to be developed in Volume 2, but are not required for Chapters 5–8. Thus it is possible to go directly from Chapter 2 to Chapter 5, although the student of probability or analysis will encounter the material of Chapters 3 and 4 early in his career.

Problems are given at the end of each section. Fairly detailed solutions are given to many problems, and instructors may obtain solutions to those problems not worked out in the text by writing to the publisher.

It is a pleasure to thank Professors Melvin Gardner, Samuel Saslaw, and William Stout, who used the manuscript in their classes and made many helpful suggestions, Mrs. Dee Keel for another beautiful typing job, and the staff at Academic Press for their encouragement and cooperation.

Summary of Notation

We indicate here the notational conventions to be used throughout the book. The numbering system is standard; for example, 2.7.4 means Chapter 2, Section 7, Part 4. In the Appendix on General Topology, the letter A is used; thus A6.3 means Section 6, Part 3 of the appendix.

The symbol ■ will be used to mark the end of a proof.

1 Sets

If A and B are subsets of a set Ω , $A \cup B$ will denote the *union* of A and B , and $A \cap B$ the *intersection* of A and B . The union and intersection of a family of sets A_i are denoted by $\bigcup_i A_i$ and $\bigcap_i A_i$. The *complement* of A (relative to Ω) is denoted by A^c .

The statement “ B is a *subset* of A ” is denoted by $B \subset A$; the inclusion need not be proper, that is, we have $A \subset A$ for any set A . We also write $B \subset A$ as $A \supset B$, to be read “ A is an *overset* (or *superset*) of B .”

The notation $A - B$ will always mean, unless otherwise specified, the set of points that belong to A but not to B . It is referred to as the *difference* between A and B ; a *proper difference* is a set $A - B$, where $B \subset A$.

The *symmetric difference* between A and B is by definition the union of $A - B$ and $B - A$; it is denoted by $A \Delta B$.

If $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = A$, we say that the A_n form an *increasing* sequence of sets (increasing to A) and write $A_n \uparrow A$. Similarly, if $A_1 \supset A_2 \supset \cdots$ and $\bigcap_{n=1}^{\infty} A_n = A$, we say that the A_n form a *decreasing* sequence of sets (decreasing to A) and write $A_n \downarrow A$.

The word “includes” will always imply a subset relation, and the word “contains” a membership relation. Thus if \mathcal{C} and \mathcal{D} are collections of sets, “ \mathcal{C} includes \mathcal{D} ” means that $\mathcal{D} \subset \mathcal{C}$. Equivalently, we may say that \mathcal{C} contains all sets in \mathcal{D} , in other words, each $A \in \mathcal{D}$ is also a member of \mathcal{C} .

A *countable* set is one that is either finite or countably infinite.

2 Real Numbers

The set of real numbers will be denoted by R , and R^n will denote n -dimensional Euclidean space. In R , the interval $(a, b]$ is defined as $\{x \in R : a < x \leq b\}$, and (a, ∞) as $\{x \in R : x > a\}$; other types of intervals are defined similarly. If $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ are points in R^n , $a \leq b$ will mean $a_i \leq b_i$ for all i . The interval $(a, b]$ is defined as $\{x \in R^n : a_i < x_i \leq b_i, i = 1, \dots, n\}$, and other types of intervals are defined similarly.

The set of *extended real numbers* is the two-point compactification $R \cup \{\infty\} \cup \{-\infty\}$, denoted by \bar{R} ; the set of n -tuples (x_1, \dots, x_n) , with each $x_i \in \bar{R}$, is denoted by \bar{R}^n . We adopt the following rules of arithmetic in \bar{R} .

$$a + \infty = \infty + a = \infty, \quad a - \infty = -\infty + a = -\infty, \quad a \in R,$$

$$\infty + \infty = \infty, \quad -\infty - \infty = -\infty \quad (\infty - \infty \text{ is not defined}),$$

$$b \cdot \infty = \infty \cdot b = \begin{cases} \infty & \text{if } b \in \bar{R}, \quad b > 0, \\ -\infty & \text{if } b \in \bar{R}, \quad b < 0, \end{cases}$$

$$\frac{a}{\infty} = \frac{a}{-\infty} = 0, \quad a \in R \quad \left(\frac{\infty}{\infty} \text{ is not defined} \right),$$

$$0 \cdot \infty = \infty \cdot 0 = 0.$$

The rules are convenient when developing the properties of the abstract Lebesgue integral, but it should be emphasized that \bar{R} is not a field under these operations.

Unless otherwise specified (notably in the definition of a positive linear functional in Chapter 4), *positive* means (strictly) greater than zero, and *non-negative* means greater than or equal to zero.

The set of *complex numbers* is denoted by C , and the set of n -tuples of complex numbers by C^n .

3 Functions

If f is a function from Ω to Ω' (written as $f: \Omega \rightarrow \Omega'$) and $B \subset \Omega'$, the *preimage* of B under f is given by $f^{-1}(B) = \{\omega \in \Omega: f(\omega) \in B\}$. It follows from the definition that $f^{-1}(\bigcup_i B_i) = \bigcup_i f^{-1}(B_i)$, $f^{-1}(\bigcap_i B_i) = \bigcap_i f^{-1}(B_i)$, $f^{-1}(A - B) = f^{-1}(A) - f^{-1}(B)$; hence $f^{-1}(A^c) = [f^{-1}(A)]^c$. If \mathcal{C} is a class of sets, $f^{-1}(\mathcal{C})$ means the collection of sets $f^{-1}(B)$, $B \in \mathcal{C}$.

If $f: R \rightarrow R$, f is *increasing* iff $x < y$ implies $f(x) \leq f(y)$; *decreasing* iff $x < y$ implies $f(x) \geq f(y)$. Thus, “increasing” and “decreasing” do not have the strict connotation. If $f_n: \Omega \rightarrow \bar{R}$, $n = 1, 2, \dots$, the f_n are said to form an *increasing sequence* iff $f_n(\omega) \leq f_{n+1}(\omega)$ for all n and ω ; a *decreasing sequence* is defined similarly.

If f and g are functions from Ω to \bar{R} , statements such as $f \leq g$ are always interpreted as holding pointwise, that is, $f(\omega) \leq g(\omega)$ for all $\omega \in \Omega$. Similarly, if $f_i: \Omega \rightarrow \bar{R}$ for each $i \in I$, $\sup_i f_i$ is the function whose value at ω is $\sup\{f_i(\omega): i \in I\}$.

If f_1, f_2, \dots form an increasing sequence of functions with limit f [that is, $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$ for all ω], we write $f_n \uparrow f$. (Similarly, $f_n \downarrow f$ is used for a decreasing sequence.)

Sometimes, a set such as $\{\omega \in \Omega: f(\omega) \leq g(\omega)\}$ is abbreviated as $\{f \leq g\}$; similarly, the preimage $\{\omega \in \Omega: f(\omega) \in B\}$ is written as $\{f \in B\}$.

If $A \subset \Omega$, the *indicator* of A is the function defined by $I_A(\omega) = 1$ if $\omega \in A$ and by $I_A(\omega) = 0$ if $\omega \notin A$. The phrase “characteristic function” is often used in the literature, but we shall not adopt this term here.

If f is a function of two variables x and y , the symbol $f(x, \cdot)$ is used for the mapping $y \rightarrow f(x, y)$ with x fixed.

The *composition* of two functions $X: \Omega \rightarrow \Omega'$ and $f: \Omega' \rightarrow \Omega$ is denoted by $f \circ X$ or $f(X)$.

If $f: \Omega \rightarrow \bar{R}$, the *positive* and *negative parts* of f are defined by $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$, that is,

$$f^+(\omega) = \begin{cases} f(\omega) & \text{if } f(\omega) \geq 0, \\ 0 & \text{if } f(\omega) < 0, \end{cases}$$

$$f^-(\omega) = \begin{cases} -f(\omega) & \text{if } f(\omega) \leq 0, \\ 0 & \text{if } f(\omega) > 0. \end{cases}$$

4 Topology

A *metric space* is a set Ω with a function d (called a *metric*) from $\Omega \times \Omega$ to the nonnegative reals, satisfying $d(x, y) \geq 0$, $d(x, y) = 0$ iff $x = y$, $d(x, y) =$

$d(y, x)$, and $d(x, z) \leq d(x, y) + d(y, z)$. If $d(x, y)$ can be 0 for $x \neq y$, but d satisfies the remaining properties, d is called a *pseudometric* (the term *semimetric* is also used in the literature).

A *ball* (or *open ball*) in a metric or pseudometric space is a set of the form $B(x, r) = \{y \in \Omega: d(x, y) < r\}$ where x , the *center* of the ball, is a point of Ω , and r , the *radius*, is a positive real number. A *closed ball* is a set of the form $\bar{B}(x, r) = \{y \in \Omega: d(x, y) \leq r\}$.

A *topological space* is a set Ω with a collection \mathcal{T} of subsets of Ω , called a *topology*, such that \emptyset and Ω belong to \mathcal{T} and \mathcal{T} is closed under finite intersection and arbitrary union. The members of \mathcal{T} are called *open sets*. A *base* for \mathcal{T} is a collection of sets \mathcal{B} such that each open set is a union of sets in \mathcal{B} .

A *neighborhood* of a point $x \in \Omega$ is an open set containing x ; an *overneighborhood* of x is an overset of a neighborhood of x . A *base for the neighborhood system at x* (or simply a *base at x*) is a collection \mathcal{U} of neighborhoods of x such that for each neighborhood V of x there is a set $U \in \mathcal{U}$ with $U \subset V$. A *base for the overneighborhood system at x* is defined similarly, with neighborhood replaced by overneighborhood.

If Ω is a topological space, $C(\Omega)$ denotes the class of continuous real-valued functions, and $C_b(\Omega)$ the class of bounded, real-valued, continuous functions, on Ω . The phrase “lower semicontinuous” is abbreviated LSC, and “upper semicontinuous” is abbreviated USC. Sequences in Ω are denoted by $\{x_n, n = 1, 2, \dots\}$ and nets by $\{x_n, n \in D\}$.

5 Vector Spaces

The terms “vector space” and “linear space” are synonymous. All vector spaces are over the real or complex field, and the complex field is assumed unless the phrase “real vector space” is used.

A *Hamel basis* for a vector space L is a maximal linearly independent subset B of L . (Linear independence means that if $x_1, \dots, x_n \in B$, $n = 1, 2, \dots$, and c_1, \dots, c_n are scalars, then $\sum_{i=1}^n c_i x_i = 0$ iff all $c_i = 0$.) Alternatively, a Hamel basis is a linearly independent subset B with the property that each $x \in L$ is a finite linear combination of elements in B . [An *orthonormal basis* for a Hilbert space (Chapter 3) is a different concept.]

The terms “subspace” and “linear manifold” are synonymous, each referring to a subset M of a vector space L that is itself a vector space under the operations of addition and scalar multiplication in L . If there is a topology on L and M is a closed set in the topology, then M is called a *closed subspace*.

If B is an arbitrary subset of L , the *linear manifold generated by B* , denoted by $L(B)$, is the smallest linear manifold containing all elements of B , that is, the collection of finite linear combinations of elements of B . Assuming a topology on L , the *space spanned by B* , denoted by $S(B)$, is the smallest closed subspace containing all elements of B . Explicitly, $S(B)$ is the closure of $L(B)$.

6 Zorn's Lemma

A *partial ordering* on a set S is a relation “ \leq ” that is

- (1) *reflexive*: $a \leq a$,
- (2) *antisymmetric*: if $a \leq b$ and $b \leq a$, then $a = b$, and
- (3) *transitive*: if $a \leq b$ and $b \leq c$, then $a \leq c$.

(All elements a, b, c belong to S .)

If $C \subset S$, C is said to be *totally ordered* iff for all $a, b \in C$, either $a \leq b$ or $b \leq a$. A totally ordered subset of S is also called a *chain* in S .

The form of Zorn's lemma that will be used in the text is as follows:

Let S be a set with a partial ordering “ \leq .” Assume that every chain C in S has an upper bound; in other words, there is an element $x \in S$ such that $x \geq a$ for all $a \in C$. Then S has a maximal element, that is, an element m such that for each $a \in S$ it is not possible to have $m \leq a$ and $m \neq a$.

Fundamentals of Measure and Integration Theory

In this chapter we give a self-contained presentation of the basic concepts of the theory of measure and integration. The principles discussed here and in Chapter 2 will serve as background for the study of probability as well as harmonic analysis, linear space theory, and other areas of mathematics.

1.1 Introduction

It will be convenient to start with a little practice in the algebra of sets. This will serve as a refresher and also as a way of collecting a few results that will often be useful.

Let A_1, A_2, \dots be subsets of a set Ω . If $A_1 \subset A_2 \subset \dots$ and $\bigcup_{n=1}^{\infty} A_n = A$, we say that the A_n form an *increasing* sequence of sets with limit A , or that the A_n increase to A ; we write $A_n \uparrow A$. If $A_1 \supset A_2 \supset \dots$ and $\bigcap_{n=1}^{\infty} A_n = A$, we say that the A_n form a *decreasing* sequence of sets with limit A , or that the A_n decrease to A ; we write $A_n \downarrow A$.

The *De Morgan laws*, namely, $(\bigcup_n A_n)^c = \bigcap_n A_n^c$, $(\bigcap_n A_n)^c = \bigcup_n A_n^c$, imply that

- (1) if $A_n \uparrow A$, then $A_n^c \downarrow A^c$; if $A_n \downarrow A$, then $A_n^c \uparrow A^c$.

It is sometimes useful to write a union of sets as a disjoint union. This may be done as follows:

Let A_1, A_2, \dots be subsets of Ω . For each n we have

$$(2) \quad \bigcup_{i=1}^n A_i = A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \\ \cup \cdots \cup (A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n).$$

Furthermore,

$$(3) \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n).$$

In (2) and (3), the sets on the right are disjoint. If the A_n form an increasing sequence, the formulas become

$$(4) \quad \bigcup_{i=1}^n A_i = A_1 \cup (A_2 - A_1) \cup \cdots \cup (A_n - A_{n-1})$$

and

$$(5) \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_n - A_{n-1})$$

(Take A_0 as the empty set).

The results (1)–(5) are proved using only the definitions of union, intersection, and complementation; see Problem 1.

The following set operation will be of particular interest. If A_1, A_2, \dots are subsets of Ω , we define

$$(6) \quad \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Thus $\omega \in \limsup_n A_n$ iff for every n , $\omega \in A_k$ for some $k \geq n$, in other words,

$$(7) \quad \omega \in \limsup_n A_n \text{ iff } \omega \in A_n \text{ for infinitely many } n.$$

Also define

$$(8) \quad \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Thus $\omega \in \liminf_n A_n$ iff for some n , $\omega \in A_k$ for all $k \geq n$, in other words,

$$(9) \quad \omega \in \liminf_n A_n \text{ iff } \omega \in A_n \text{ eventually, that is, for all but finitely many } n.$$

We call $\limsup_n A_n$ the *upper limit* of the sequence of sets A_n , and $\liminf_n A_n$ the *lower limit*. The terminology is, of course, suggested by the analogous concepts for sequences of real numbers

$$\limsup_n x_n = \inf_n \sup_{k \geq n} x_k,$$

$$\liminf_n x_n = \sup_n \inf_{k \geq n} x_k.$$

See Problem 4 for a further development of the analogy.

The following facts may be verified (Problem 5):

- (10) $(\limsup_n A_n)^c = \liminf_n A_n^c$
- (11) $(\liminf_n A_n)^c = \limsup_n A_n^c$
- (12) $\liminf_n A_n \subset \limsup_n A_n$
- (13) If $A_n \uparrow A$ or $A_n \downarrow A$, then $\liminf_n A_n = \limsup_n A_n = A$.

In general, if $\liminf_n A_n = \limsup_n A_n = A$, A is said to be the *limit* of the sequence A_1, A_2, \dots ; we write $A = \lim_n A_n$.

Problems

1. Establish formulas (1)–(5).
2. Define sets of real numbers as follows. Let $A_n = (-1/n, 1]$ if n is odd, and $A_n = (-1, 1/n]$ if n is even. Find $\limsup_n A_n$ and $\liminf_n A_n$.
3. Let $\Omega = R^2$, A_n the interior of the circle with center at $((-1)^n/n, 0)$ and radius 1. Find $\limsup_n A_n$ and $\liminf_n A_n$.
4. Let $\{x_n\}$ be a sequence of real numbers, and let $A_n = (-\infty, x_n)$. What is the connection between $\limsup_{n \rightarrow \infty} x_n$ and $\limsup_n A_n$ (similarly for \liminf)?
5. Establish formulas (10)–(13).

1.2 Fields, σ -Fields, and Measures

Length, area, and volume, as well as probability are instances of the measure concept that we are going to discuss. A measure is a *set function*, that is, an assignment of a number $\mu(A)$ to each set A in a certain class. Some structure must be imposed on the class of sets on which μ is defined, and probability considerations provide a good motivation for the type of structure required. If Ω is a set whose points correspond to the possible outcomes of a random experiment, certain subsets of Ω will be called “events” and assigned a probability. Intuitively, A is an event if the question “Does ω belong to A ? ” has a definite yes or no answer after the experiment is performed (and the outcome corresponds to the point $\omega \in \Omega$). Now if we can answer the question “Is $\omega \in A$? ” we can certainly answer the question “Is $\omega \in A^c$? ”, and if, for each $i = 1, \dots, n$, we can decide whether or not ω belongs to A_i , then we can determine whether or not ω belongs to $\bigcup_{i=1}^n A_i$ (and similarly for $\bigcap_{i=1}^n A_i$). Thus it is natural to require that the class of

events be closed under complementation, finite union, and finite intersection; furthermore, since the answer to the question “Is $\omega \in \Omega?$ ” is always “yes,” the entire space Ω should be an event. Closure under *countable* union and intersection is difficult to justify physically, and perhaps the most convincing reason for requiring it is that a richer mathematical theory is obtained. We shall have more to say about this point after we give the definition of a measure. First, we concentrate on the underlying class of sets.

1.2.1 Definitions. Let \mathcal{F} be a collection of subsets of a set Ω . Then \mathcal{F} is called a *field* (the term *algebra* is also used) iff $\Omega \in \mathcal{F}$ and \mathcal{F} is closed under complementation and finite union, that is,

- (a) $\Omega \in \mathcal{F}$.
- (b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
- (c) If $A_1, A_2, \dots, A_n \in \mathcal{F}$ then $\bigcup_{i=1}^n A_i \in \mathcal{F}$.

It follows that \mathcal{F} is closed under finite intersection. For if $A_1, \dots, A_n \in \mathcal{F}$, then

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{F}.$$

If (c) is replaced by closure under *countable* union, that is,

- (d) If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

\mathcal{F} is called a σ -field (the term σ -algebra is also used). Just as above, \mathcal{F} is also closed under countable intersection.

If \mathcal{F} is a field, a countable union of sets in \mathcal{F} can be expressed as the limit of an increasing sequence of sets in \mathcal{F} , and conversely. To see this, note that if $A = \bigcup_{n=1}^{\infty} A_n$, then $\bigcup_{i=1}^n A_i \uparrow A$; conversely, if $A_n \uparrow A$, then $A = \bigcup_{n=1}^{\infty} A_n$. This shows that a σ -field is a field that is closed under limits of increasing sequences.

1.2.2 Examples. The largest σ -field of subsets of a fixed set Ω is the collection of all subsets of Ω . The smallest σ -field consists of the two sets \emptyset and Ω .

Let A be a nonempty proper subset of Ω , and let $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$. Then \mathcal{F} is the smallest σ -field containing A . For if \mathcal{G} is a σ -field and $A \in \mathcal{G}$, then by definition of a σ -field, Ω , \emptyset , and A^c belong to \mathcal{G} , hence $\mathcal{F} \subset \mathcal{G}$. But \mathcal{F} is a σ -field, for if we form complements or unions of sets in \mathcal{F} , we invariably obtain sets in \mathcal{F} . Thus \mathcal{F} is a σ -field that is included in any σ -field containing A , and the result follows.

If A_1, \dots, A_n are arbitrary subsets of Ω , the smallest σ -field containing A_1, \dots, A_n may be described explicitly; see Problem 8.

If \mathcal{S} is a class of sets, the smallest σ -field containing the sets of \mathcal{S} will be written as $\sigma(\mathcal{S})$, and sometimes called the *minimal σ -field over \mathcal{S}* .

Let Ω be the set R of real numbers. Let \mathcal{F} consist of all finite disjoint unions of right-semiclosed intervals. (A right-semiclosed interval is a set of the form $(a, b] = \{x: a < x \leq b\}$, $-\infty \leq a < b < \infty$; by convention we also count (a, ∞) as right-semiclosed for $-\infty \leq a < \infty$. The convention is necessary because $(-\infty, a]$ belongs to \mathcal{F} , and if \mathcal{F} is to be a field, the complement (a, ∞) must also belong to \mathcal{F} .) It may be verified that conditions (a)–(c) of 1.2.1 hold; and thus \mathcal{F} is a field. But \mathcal{F} is not a σ -field; for example, $A_n = (0, 1 - (1/n)] \in \mathcal{F}$, $n = 1, 2, \dots$, and $\bigcup_{n=1}^{\infty} A_n = (0, 1) \notin \mathcal{F}$.

If Ω is the set $\bar{R} = [-\infty, \infty]$ of extended real numbers, then just as above, the collection of finite disjoint unions of right-semiclosed intervals forms a field but not a σ -field. Here, the right-semiclosed intervals are sets of the form $(a, b] = \{x: a < x \leq b\}$, $-\infty \leq a < b \leq \infty$, and, by convention, the sets $[-\infty, b] = \{x: -\infty \leq x \leq b\}$, $-\infty \leq b \leq \infty$. (In this case the convention is necessary because $(b, \infty]$ must belong to \mathcal{F} , and therefore the complement $[-\infty, b]$ also belongs to \mathcal{F} .)

There is a type of reasoning that occurs so often in problems involving σ -fields that it deserves to be displayed explicitly, as in the following typical illustration.

If \mathcal{C} is a class of subsets of Ω and $A \subset \Omega$, we denote by $\mathcal{C} \cap A$ the class $\{B \cap A: B \in \mathcal{C}\}$. If the minimal σ -field over \mathcal{C} is $\sigma(\mathcal{C}) = \mathcal{F}$, let us show that

$$\sigma_A(\mathcal{C} \cap A) = \mathcal{F} \cap A,$$

where $\sigma_A(\mathcal{C} \cap A)$ is the minimal σ -field of subsets of A over $\mathcal{C} \cap A$. (In other words, A rather than Ω is regarded as the entire space.)

Now $\mathcal{C} \subset \mathcal{F}$, hence $\mathcal{C} \cap A \subset \mathcal{F} \cap A$, and it is not hard to verify that $\mathcal{F} \cap A$ is a σ -field of subsets of A . Therefore $\sigma_A(\mathcal{C} \cap A) \subset \mathcal{F} \cap A$.

To establish the reverse inclusion we must show that $B \cap A \in \sigma_A(\mathcal{C} \cap A)$ for all $B \in \mathcal{F}$. This is not obvious, so we resort to the following basic reasoning process, which might be called the *good sets principle*. Let \mathcal{S} be the class of good sets, that is, let \mathcal{S} consist of those sets $B \in \mathcal{F}$ such that

$$B \cap A \in \sigma_A(\mathcal{C} \cap A).$$

Since \mathcal{F} and $\sigma_A(\mathcal{C} \cap A)$ are σ -fields, it follows quickly that \mathcal{S} is a σ -field. But $\mathcal{C} \subset \mathcal{S}$, so that $\sigma(\mathcal{C}) \subset \mathcal{S}$, hence $\mathcal{F} = \mathcal{S}$ and the result follows. Briefly, every set in \mathcal{C} is good and the class of good sets forms a σ -field; consequently, every set in $\sigma(\mathcal{C})$ is good.

One other comment: If \mathcal{C} is closed under finite intersection and $A \in \mathcal{C}$, then $\mathcal{C} \cap A = \{C \in \mathcal{C}: C \subset A\}$. (Observe that if $C \subset A$, then $C = C \cap A$.)

1.2.3 Definitions and Comments. A *measure* on a σ -field \mathcal{F} is a nonnegative, extended real-valued function μ on \mathcal{F} such that whenever A_1, A_2, \dots form a finite or countably infinite collection of disjoint sets in \mathcal{F} , we have

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

If $\mu(\Omega) = 1$, μ is called a *probability measure*.

A *measure space* is a triple $(\Omega, \mathcal{F}, \mu)$ where Ω is a set, \mathcal{F} is a σ -field of subsets of Ω , and μ is a measure on \mathcal{F} . If μ is a probability measure, $(\Omega, \mathcal{F}, \mu)$ is called a *probability space*.

It will be convenient to have a slight generalization of the notion of a measure on a σ -field. Let \mathcal{F} be a *field*, μ a set function on \mathcal{F} (a map from \mathcal{F} to $\bar{\mathbb{R}}$). We say that μ is *countably additive* on \mathcal{F} iff whenever A_1, A_2, \dots form a finite or countably infinite collection of disjoint sets in \mathcal{F} whose union also belongs to \mathcal{F} (this will always be the case if \mathcal{F} is a σ -field) we have

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

If this requirement holds only for finite collections of disjoint sets in \mathcal{F} , μ is said to be *finitely additive* on \mathcal{F} . To avoid the appearance of terms of the form $+\infty - \infty$ in the summation, we always assume that $+\infty$ and $-\infty$ cannot both belong to the range of μ .

If μ is countably additive and $\mu(A) \geq 0$ for all $A \in \mathcal{F}$, μ is called a *measure* on \mathcal{F} , a *probability measure* if $\mu(\Omega) = 1$.

Note that countable additivity actually implies finite additivity. For if $\mu(A) = +\infty$ for all $A \in \mathcal{F}$, or if $\mu(A) = -\infty$ for all $A \in \mathcal{F}$, the result is immediate; therefore assume $\mu(A)$ finite for some $A \in \mathcal{F}$. By considering the sequence $A, \emptyset, \emptyset, \dots$, we find that $\mu(\emptyset) = 0$, and finite additivity is now established by considering the sequence $A_1, \dots, A_n, \emptyset, \emptyset, \dots$, where A_1, \dots, A_n are disjoint sets in \mathcal{F} .

Although the set function given by $\mu(A) = +\infty$ for all $A \in \mathcal{F}$ satisfies the definition of a measure, and similarly $\mu(A) = -\infty$ for all $A \in \mathcal{F}$ defines a countably additive set function, we shall from now on exclude these cases. Thus by the above discussion, we always have $\mu(\emptyset) = 0$.

It is possible to develop a theory of measure with the countable additivity requirement replaced by the weaker condition of finite additivity. The disadvantage of doing this is that the resulting mathematical equipment is much less powerful. However, a convincing physical justification of countable additivity has yet to be given. If the probability $P(A)$ of an event A is to represent the long run relative frequency of A in a sequence of performances

of a random experiment, P must be a finitely additive set function; but only finitely many measurements can be made in a finite time interval, so countable additivity is not inevitable on physical grounds. Dubins and Savage (1965) have considered certain problems in stochastic processes using only finitely additive set functions, and they assert that for their purposes, finite additivity avoids some of the complications of countable additivity without sacrificing power or scope. On the other hand, at the present time almost all applications of measure theory in mathematics (and physics and engineering as well) use countable rather than finite additivity, and we shall follow this practice here.

1.2.4 Examples. Let Ω be any set, and let \mathcal{F} consist of all subsets of Ω . Define $\mu(A)$ as the number of points of A . Thus if A has n members, $n = 0, 1, 2, \dots$, then $\mu(A) = n$; if A is an infinite set, $\mu(A) = \infty$. The set function μ is a measure on \mathcal{F} , called *counting measure* on Ω .

A closely related measure is defined as follows. Let $\Omega = \{x_1, x_2, \dots\}$ be a finite or countably infinite set, and let p_1, p_2, \dots be nonnegative numbers. Take \mathcal{F} as all subsets of Ω , and define

$$\mu(A) = \sum_{x_i \in A} p_i.$$

Thus if $A = \{x_{i_1}, x_{i_2}, \dots\}$, then $\mu(A) = p_{i_1} + p_{i_2} + \dots$. The set function μ is a measure on \mathcal{F} and $\mu\{x_i\} = p_i$, $i = 1, 2, \dots$. A probability measure will be obtained iff $\sum_i p_i = 1$; if all $p_i = 1$, then μ is counting measure.

Now if A is a subset of R , we try to arrive at a definition of the *length* of A . If A is an interval (open, closed, or semiclosed) with endpoints a and b , it is reasonable to take the length of A to be $\mu(A) = b - a$. If A is a complicated set, we may not have any intuition about its length, but we shall see in Section 1.4 that the requirements that $\mu(a, b] = b - a$ for all $a, b \in R$, $a < b$, and that μ be a measure, determine μ on a large class of sets.

Specifically, μ is determined on the collection of *Borel sets* of R , denoted by $\mathcal{B}(R)$ and defined as the smallest σ -field of subsets of R containing all intervals $(a, b]$, $a, b \in R$.

Note that $\mathcal{B}(R)$ is guaranteed to exist; it may be described (admittedly in a rather ethereal way) as the intersection of all σ -fields containing the intervals $(a, b]$. Also, if a σ -field contains, say, all open intervals, it must contain all intervals $(a, b]$, and conversely. For

$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n} \right) \quad \text{and} \quad (a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n} \right].$$

Thus $\mathcal{B}(R)$ is the smallest σ -field containing all open intervals. Similarly we may replace the intervals $(a, b]$ by other classes of intervals, for instance,

- all closed intervals,
- all intervals $[a, b)$, $a, b \in R$,
- all intervals (a, ∞) , $a \in R$,
- all intervals $[a, \infty)$, $a \in R$,
- all intervals $(-\infty, b)$, $b \in R$,
- all intervals $(-\infty, b]$, $b \in R$.

Since a σ -field that contains all intervals of a given type contains all intervals of any other type, $\mathcal{B}(R)$ may be described as the smallest σ -field that contains the class of all intervals of R . Similarly, $\mathcal{B}(R)$ is the smallest σ -field containing all open sets of R . (To see this, recall that an open set is a countable union of open intervals.) Since a set is open iff its complement is closed, $\mathcal{B}(R)$ is the smallest σ -field containing all closed sets of R . Finally, if \mathcal{F}_0 is the field of finite disjoint unions of right-semiclosed intervals (see 1.2.2), then $\mathcal{B}(R)$ is the smallest σ -field containing the sets of \mathcal{F}_0 .

Intuitively, we may think of generating the Borel sets by starting with the intervals and forming complements and countable unions and intersections in all possible ways. This idea is made precise in Problem 11.

The class of Borel sets of \bar{R} , denoted by $\mathcal{B}(\bar{R})$, is defined as the smallest σ -field of subsets of \bar{R} containing all intervals $(a, b]$, $a, b \in \bar{R}$. The above discussion concerning the replacement of the right-semiclosed intervals by other classes of sets applies equally well to \bar{R} .

If $E \in \mathcal{B}(R)$, $\mathcal{B}(E)$ will denote $\{B \in \mathcal{B}(R): B \subset E\}$; this coincides with $\{A \cap E: A \in \mathcal{B}(R)\}$ (see 1.2.2).

We now begin to develop some properties of set functions:

1.2.5 Theorem. Let μ be a finitely additive set function on the field \mathcal{F} .

- (a) $\mu(\emptyset) = 0$.
- (b) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ for all $A, B \in \mathcal{F}$.
- (c) If $A, B \in \mathcal{F}$ and $B \subset A$, then $\mu(A) = \mu(B) + \mu(A - B)$

(hence $\mu(A - B) = \mu(A) - \mu(B)$ if $\mu(B)$ is finite, and $\mu(B) \leq \mu(A)$ if

$$\mu(A - B) \geq 0.$$

- (d) If μ is nonnegative,

$$\mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mu(A_i) \quad \text{for all } A_1, \dots, A_n \in \mathcal{F}.$$

If μ is a measure,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

for all $A_1, A_2, \dots \in \mathcal{F}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

PROOF. (a) Pick $A \in \mathcal{F}$ such that $\mu(A)$ is finite; then

$$\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset).$$

(b) By finite additivity,

$$\begin{aligned}\mu(A) &= \mu(A \cap B) + \mu(A - B), \\ \mu(B) &= \mu(A \cap B) + \mu(B - A).\end{aligned}$$

Add the above equations to obtain

$$\begin{aligned}\mu(A) + \mu(B) &= \mu(A \cap B) + [\mu(A - B) + \mu(B - A) + \mu(A \cap B)] \\ &= \mu(A \cap B) + \mu(A \cup B).\end{aligned}$$

- (c) We may write $A = B \cup (A - B)$, hence $\mu(A) = \mu(B) + \mu(A - B)$.
 (d) We have

$$\begin{aligned}\bigcup_{i=1}^n A_i &= A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \cup \dots \\ &\quad \cup (A_1^c \cap \dots \cap A_{n-1}^c \cap A_n)\end{aligned}$$

[see Section 1.1, formula (2)]. The sets on the right are disjoint and

$$\mu(A_1^c \cap \dots \cap A_{n-1}^c \cap A_n) \leq \mu(A_n) \quad \text{by (c).}$$

The case in which μ is a measure is handled using the identity (3) of Section 1.1. ■

1.2.6 Definitions. A set function μ defined on \mathcal{F} is said to be *finite* iff $\mu(A)$ is finite, that is, not $\pm \infty$, for each $A \in \mathcal{F}$. If μ is finitely additive, it is sufficient to require that $\mu(\Omega)$ be finite; for $\Omega = A \cup A^c$, and if $\mu(A)$ is, say, $+\infty$, so is $\mu(\Omega)$.

A nonnegative, finitely additive set function μ on the field \mathcal{F} is said to be σ -*finite* on \mathcal{F} iff Ω can be written as $\bigcup_{n=1}^{\infty} A_n$ where the A_n belong to \mathcal{F} and $\mu(A_n) < \infty$ for all n . [By formula (3) of Section 1.1, the A_n may be assumed disjoint.] We shall see that many properties of finite measures can be extended quickly to σ -finite measures.

It follows from 1.2.5(c) that a nonnegative, finitely additive set function μ on a field \mathcal{F} is finite iff it is bounded; that is, $\sup\{|\mu(A)| : A \in \mathcal{F}\} < \infty$.

This no longer holds if the nonnegativity assumption is dropped (see Problem 4). It is true, however, that a *countably* additive set function on a σ -field is finite iff it is bounded; this will be proved in 2.1.3.

Countably additive set functions have a basic continuity property, which we now describe.

1.2.7 Theorem. Let μ be a countably additive set function on the σ -field \mathcal{F} .

- (a) If $A_1, A_2, \dots \in \mathcal{F}$ and $A_n \uparrow A$, then $\mu(A_n) \rightarrow \mu(A)$ as $n \rightarrow \infty$.
- (b) If $A_1, A_2, \dots \in \mathcal{F}$, $A_n \downarrow A$, and $\mu(A_1)$ is finite [hence $\mu(A_n)$ is finite for all n since $\mu(A_1) = \mu(A_n) + \mu(A_1 - A_n)$], then $\mu(A_n) \rightarrow \mu(A)$ as $n \rightarrow \infty$.

The same results hold if \mathcal{F} is only assumed to be a field, if we add the hypothesis that the limit sets A belong to \mathcal{F} . [If $A \notin \mathcal{F}$ and $\mu \geq 0$, 1.2.5(c) implies that $\mu(A_n)$ increases to a limit in part (a), and decreases to a limit in part (b), but we cannot identify the limit with $\mu(A)$.]

PROOF. (a) If $\mu(A_n) = \infty$ for some n , then $\mu(A) = \mu(A_n) + \mu(A - A_n) = \infty + \mu(A - A_n) = \infty$. Replacing A by A_k we find that $\mu(A_k) = \infty$ for all $k \geq n$, and we are finished. In the same way we eliminate the case in which $\mu(A_n) = -\infty$ for some n . Thus we may assume that all $\mu(A_n)$ are finite.

Since the A_n form an increasing sequence, we may use the identity (5) of Section 1.1:

$$A = A_1 \cup (A_2 - A_1) \cup \dots \cup (A_n - A_{n-1}) \cup \dots$$

Therefore, by 1.2.5(c),

$$\begin{aligned} \mu(A) &= \mu(A_1) + \mu(A_2) - \mu(A_1) + \dots + \mu(A_n) - \mu(A_{n-1}) + \dots \\ &= \lim_{n \rightarrow \infty} \mu(A_n). \end{aligned}$$

(b) If $A_n \downarrow A$, then $A_1 - A_n \uparrow A_1 - A$, hence $\mu(A_1 - A_n) \rightarrow \mu(A_1 - A)$ by (a). The result now follows from 1.2.5(c). ■

We shall frequently encounter situations in which finite additivity of a particular set function is easily established, but countable additivity is more difficult. It is useful to have the result that finite additivity plus continuity implies countable additivity.

1.2.8 Theorem. Let μ be a finitely additive set function on the field \mathcal{F} .

- (a) Assume that μ is *continuous from below* at each $A \in \mathcal{F}$, that is, if $A_1, A_2, \dots \in \mathcal{F}$, $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, and $A_n \uparrow A$, then $\mu(A_n) \rightarrow \mu(A)$. It follows that μ is countably additive on \mathcal{F} .
- (b) Assume that μ is *continuous from above* at the empty set, that is, if $A_1, A_2, \dots \in \mathcal{F}$ and $A_n \downarrow \emptyset$, then $\mu(A_n) \rightarrow 0$. It follows that μ is countably additive on \mathcal{F} .

PROOF. (a) Let A_1, A_2, \dots be disjoint sets in \mathcal{F} whose union A belongs to \mathcal{F} . If $B_n = \bigcup_{i=1}^n A_i$ then $B_n \uparrow A$, hence $\mu(B_n) \rightarrow \mu(A)$ by hypothesis. But $\mu(B_n) = \sum_{i=1}^n \mu(A_i)$ by finite additivity, hence $\mu(A) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i)$, the desired result.

(b) Let A_1, A_2, \dots be disjoint sets in \mathcal{F} whose union A belongs to \mathcal{F} , and let $B_n = \bigcup_{i=1}^n A_i$. By 1.2.5(c), $\mu(A) = \mu(B_n) + \mu(A - B_n)$; but $A - B_n \downarrow \emptyset$, so by hypothesis, $\mu(A - B_n) \rightarrow 0$. Thus $\mu(B_n) \rightarrow \mu(A)$, and the result follows as in (a). ■

If μ_1 and μ_2 are measures on the σ -field \mathcal{F} , then $\mu = \mu_1 - \mu_2$ is countably additive on \mathcal{F} , assuming either μ_1 or μ_2 is finite-valued. We shall see later (in 2.1.3) that any countably additive set function on a σ -field can be expressed as the difference of two measures.

For examples of finitely additive set functions that are not countably additive, see Problems 1, 3, and 4.

Problems

- Let Ω be a countably infinite set, and let \mathcal{F} consist of all subsets of Ω . Define $\mu(A) = 0$ if A is finite, $\mu(A) = \infty$ if A is infinite.
 - Show that μ is finitely additive but not countably additive.
 - Show that Ω is the limit of an increasing sequence of sets A_n with $\mu(A_n) = 0$ for all n , but $\mu(\Omega) = \infty$.
- Let μ be counting measure on Ω , where Ω is an infinite set. Show that there is a sequence of sets $A_n \downarrow \emptyset$ with $\lim_{n \rightarrow \infty} \mu(A_n) \neq 0$.
- Let Ω be a countably infinite set, and let \mathcal{F} be the field consisting of all finite subsets of Ω and their complements. If A is finite, set $\mu(A) = 0$, and if A^c is finite, set $\mu(A) = 1$.
 - Show that μ is finitely additive but not countably additive on \mathcal{F} .
 - Show that Ω is the limit of an increasing sequence of sets $A_n \in \mathcal{F}$ with $\mu(A_n) = 0$ for all n , but $\mu(\Omega) = 1$.

4. Let \mathcal{F} be the field of finite disjoint unions of right-semiclosed intervals of R , and define the set function μ on \mathcal{F} as follows.

$$\mu(-\infty, a] = a, \quad a \in R,$$

$$\mu(a, b] = b - a, \quad a, b \in R, \quad a < b,$$

$$\mu(b, \infty) = -b, \quad b \in R,$$

$$\mu(R) = 0,$$

$$\mu\left(\bigcup_{i=1}^n I_i\right) = \sum_{i=1}^n \mu(I_i)$$

if I_1, \dots, I_n are disjoint right-semiclosed intervals.

- (a) Show that μ is finitely additive but not countably additive on \mathcal{F} .
 (b) Show that μ is finite but unbounded on \mathcal{F} .

5. Let μ be a nonnegative, finitely additive set function on the field \mathcal{F} . If A_1, A_2, \dots are disjoint sets in \mathcal{F} and $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, show that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu(A_n).$$

6. Let $f: \Omega \rightarrow \Omega'$, and let \mathcal{C} be a class of subsets of Ω' . Show that

$$\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C})),$$

where $f^{-1}(\mathcal{C}) = \{f^{-1}(A): A \in \mathcal{C}\}$. (Use the good sets principle.)

7. If A is a Borel subset of R , show that the smallest σ -field of subsets of A containing the sets open in A (in the relative topology inherited from R) is $\{B \in \mathcal{B}(R): B \subset A\}$.
8. Let A_1, \dots, A_n be arbitrary subsets of a set Ω . Describe (explicitly) the smallest σ -field \mathcal{F} containing A_1, \dots, A_n . How many sets are there in \mathcal{F} ? (Give an upper bound that is attainable under certain conditions.)
9. (a) Let \mathcal{C} be an arbitrary class of subsets of Ω , and let \mathcal{G} be the collection of all finite unions $\bigcup_{i=1}^n A_i$, $n = 1, 2, \dots$, where each A_i is a finite intersection $\bigcap_{j=1}^r B_{ij}$, with B_{ij} a set in \mathcal{C} or its complement.
 Show that \mathcal{G} is the minimal field (not σ -field) over \mathcal{C} .
 (b) Show that the minimal field can also be described as the collection \mathcal{D} of all finite disjoint unions $\bigcup_{i=1}^n A_i$, where the A_i are as above.
 (c) If $\mathcal{F}_1, \dots, \mathcal{F}_n$ are fields of subsets of Ω , show that the smallest field including $\mathcal{F}_1, \dots, \mathcal{F}_n$ consists of all finite (disjoint) unions of sets $A_1 \cap \dots \cap A_n$ with $A_i \in \mathcal{F}_i$, $i = 1, \dots, n$.
10. Let μ be a finite measure on the σ -field \mathcal{F} . If $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, and $A = \lim_n A_n$ (see Section 1.1), show that $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.

11. Let \mathcal{C} be any class of subsets of Ω , with $\emptyset, \Omega \in \mathcal{C}$. Define $\mathcal{C}_0 = \mathcal{C}$, and for any ordinal $\alpha > 0$ write, inductively,

$$\mathcal{C}_\alpha = (\bigcup \{\mathcal{C}_\beta : \beta < \alpha\})',$$

where \mathcal{D}' denotes the class of all countable unions of differences of sets in \mathcal{D} .

Let $\mathcal{S} = \bigcup \{\mathcal{C}_\alpha : \alpha < \beta_1\}$, where β_1 is the first uncountable ordinal, and let \mathcal{F} be the minimal σ -field over \mathcal{C} . Since each $\mathcal{C}_\alpha \subset \mathcal{F}$, we have $\mathcal{S} \subset \mathcal{F}$. Also, the \mathcal{C}_α increase with α , and $\mathcal{C} \subset \mathcal{C}_\alpha$ for all α .

- (a) Show that \mathcal{S} is a σ -field (hence $\mathcal{S} = \mathcal{F}$ by minimality of \mathcal{F}).
- (b) If the cardinality of \mathcal{C} is at most c , the cardinality of the reals, show that $\text{card } \mathcal{F} \leq c$ also.

1.3 Extension of Measures

In Section 1.2.4, we discussed the concept of length of a subset of R . The problem was to extend the set function given on intervals by $\mu(a, b] = b - a$ to a larger class of sets. If \mathcal{F}_0 is the field of finite disjoint unions of right-semiclosed intervals, there is no problem extending μ to \mathcal{F}_0 : if A_1, \dots, A_n are disjoint right-semiclosed intervals, we set $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. The resulting set function on \mathcal{F}_0 is finitely additive, but countable additivity is not clear at this point. Even if we can prove countable additivity on \mathcal{F}_0 , we still have the problem of extending μ to the minimal σ -field over \mathcal{F}_0 , namely, the Borel sets.

We are going to consider a generalization of the above problem. Instead of working only with length, we shall examine set functions given by $\mu(a, b] = F(b) - F(a)$ where F is an increasing right-continuous function from R to R . The extension technique to be developed is not restricted to set functions defined on subsets of R ; we shall prove a general result concerning the extension of a measure from a field \mathcal{F}_0 to the minimal σ -field over \mathcal{F}_0 .

It will be convenient to consider finite measures at first, and nothing is lost if we normalize and work with probability measures:

1.3.1 Lemma. Let \mathcal{F}_0 be a field of subsets of a set Ω , and let P be a probability measure on \mathcal{F}_0 . Suppose that the sets A_1, A_2, \dots belong to \mathcal{F}_0 and increase to a limit A , and that the sets A'_1, A'_2, \dots belong to \mathcal{F}_0 and increase to A' . (A and A' need not belong to \mathcal{F}_0 .) If $A \subset A'$, then

$$\lim_{m \rightarrow \infty} P(A_m) \leq \lim_{n \rightarrow \infty} P(A'_n).$$

PROOF. If m is fixed, $A_m \cap A_n' \uparrow A_m \cap A' = A_m$ as $n \rightarrow \infty$, hence

$$P(A_m \cap A_n') \rightarrow P(A_m)$$

by 1.2.7(a). But $P(A_m \cap A_n') \leq P(A_n')$ by 1.2.5(c), hence

$$P(A_m) = \lim_{n \rightarrow \infty} P(A_m \cap A_n') \leq \lim_{n \rightarrow \infty} P(A_n').$$

Let $m \rightarrow \infty$ to finish the proof. ■

We are now ready for the first extension of P to a larger class of sets:

1.3.2 Lemma. Let P be a probability measure on the field \mathcal{F}_0 . Let \mathcal{G} be the collection of all limits of increasing sequences of sets in \mathcal{F}_0 , that is, $A \in \mathcal{G}$ iff there are sets $A_n \in \mathcal{F}_0$, $n = 1, 2, \dots$, such that $A_n \uparrow A$. (Note that \mathcal{G} can also be described as the collection of all countable unions of sets in \mathcal{F}_0 ; see 1.2.1.)

Define μ on \mathcal{G} as follows. If $A_n \in \mathcal{F}_0$, $n = 1, 2, \dots$, $A_n \uparrow A$ ($\in \mathcal{G}$), set $\mu(A) = \lim_{n \rightarrow \infty} P(A_n)$; μ is well defined by 1.3.1, and $\mu = P$ on \mathcal{F}_0 . Then:

- (a) $\emptyset \in \mathcal{G}$ and $\mu(\emptyset) = 0$; $\Omega \in \mathcal{G}$ and $\mu(\Omega) = 1$; $0 \leq \mu(A) \leq 1$ for all $A \in \mathcal{G}$.
- (b) If $G_1, G_2 \in \mathcal{G}$, then $G_1 \cup G_2, G_1 \cap G_2 \in \mathcal{G}$ and $\mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) = \mu(G_1) + \mu(G_2)$.
- (c) If $G_1, G_2 \in \mathcal{G}$ and $G_1 \subset G_2$, then $\mu(G_1) \leq \mu(G_2)$.
- (d) If $G_n \in \mathcal{G}$, $n = 1, 2, \dots$, and $G_n \uparrow G$, then $G \in \mathcal{G}$ and $\mu(G_n) \rightarrow \mu(G)$.

PROOF. (a) This is clear since $\mu = P$ on \mathcal{F}_0 and P is a probability measure.

(b) Let $A_{n1} \in \mathcal{F}_0, A_{n1} \uparrow G_1; A_{n2} \in \mathcal{F}_0, A_{n2} \uparrow G_2$. We have $P(A_{n1} \cup A_{n2}) + P(A_{n1} \cap A_{n2}) = P(A_{n1}) + P(A_{n2})$ by 1.2.5(b); let $n \rightarrow \infty$ to complete the argument.

(c) This follows from 1.3.1.

(d) Since G is a countable union of sets in \mathcal{F}_0 , $G \in \mathcal{G}$. Now for each n we can find sets $A_{nm} \in \mathcal{F}_0$, $m = 1, 2, \dots$, with $A_{nm} \uparrow G_n$ as $m \rightarrow \infty$. The situation may be represented schematically as follows:

$$\begin{array}{ccccccc} A_{11} & A_{12} & \cdots & A_{1m} & \cdots & \uparrow G_1 \\ A_{21} & A_{22} & \cdots & A_{2m} & \cdots & \uparrow G_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} & \cdots & \uparrow G_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

Let $D_m = A_{1m} \cup A_{2m} \cup \dots \cup A_{mm}$ (the D_m form an increasing sequence). The key step in the proof is the observation that

$$A_{nm} \subset D_m \subset G_m \quad \text{for } n \leq m \quad (1)$$

and, therefore,

$$P(A_{nm}) \leq P(D_m) \leq \mu(G_m) \quad \text{for } n \leq m. \quad (2)$$

Let $m \rightarrow \infty$ in (1) to obtain $G_n \subset \bigcup_{m=1}^{\infty} D_m \subset G$; then let $n \rightarrow \infty$ to conclude that $D_m \uparrow G$, hence $P(D_m) \rightarrow \mu(G)$ by definition of μ . Now let $m \rightarrow \infty$ in (2) to obtain $\mu(G_n) \leq \lim_{m \rightarrow \infty} P(D_m) \leq \lim_{m \rightarrow \infty} \mu(G_m)$; then let $n \rightarrow \infty$ to conclude that $\lim_{n \rightarrow \infty} \mu(G_n) = \lim_{m \rightarrow \infty} P(D_m) = \mu(G)$. ■

We now extend μ to the class of all subsets of Ω ; however, the extension will not be countably additive on all subsets, but only on a smaller σ -field. The construction depends on properties (a)–(d) of 1.3.2, and not on the fact that μ was derived from a probability measure on a field. We express this explicitly as follows:

1.3.3 Lemma. Let \mathcal{G} be a class of subsets of a set Ω , μ a nonnegative real-valued set function on \mathcal{G} such that \mathcal{G} and μ satisfy the four conditions (a)–(d) of 1.3.2. Define, for each $A \subset \Omega$,

$$\mu^*(A) = \inf\{\mu(G): G \in \mathcal{G}, G \supset A\}.$$

Then:

- (a) $\mu^* = \mu$ on \mathcal{G} , $0 \leq \mu^*(A) \leq 1$ for all $A \subset \Omega$.
- (b) $\mu^*(A \cup B) + \mu^*(A \cap B) \leq \mu^*(A) + \mu^*(B)$; in particular, $\mu^*(A) + \mu^*(A^c) \geq \mu^*(\Omega) + \mu^*(\emptyset) = \mu(\Omega) + \mu(\emptyset) = 1$ by 1.3.2(a).
- (c) If $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.
- (d) If $A_n \uparrow A$, then $\mu^*(A_n) \rightarrow \mu^*(A)$.

PROOF. (a) This is clear from the definition of μ^* and from 1.3.2(c).

(b) If $\varepsilon > 0$, choose $G_1, G_2 \in \mathcal{G}$, $G_1 \supset A$, $G_2 \supset B$, such that $\mu(G_1) \leq \mu^*(A) + \varepsilon/2$, $\mu(G_2) \leq \mu^*(B) + \varepsilon/2$. By 1.3.2(b),

$$\begin{aligned} \mu^*(A) + \mu^*(B) + \varepsilon &\geq \mu(G_1) + \mu(G_2) = \mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) \\ &\geq \mu^*(A \cup B) + \mu^*(A \cap B). \end{aligned}$$

Since ε is arbitrary, the result follows.

- (c) This follows from the definition of μ^* .
- (d) By (c), $\mu^*(A) \geq \lim_{n \rightarrow \infty} \mu^*(A_n)$. If $\varepsilon > 0$, for each n we may choose $G_n \in \mathcal{G}$, $G_n \supset A_n$, such that

$$\mu(G_n) \leq \mu^*(A_n) + \varepsilon 2^{-n}.$$

Now $A = \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} G_n \in \mathcal{G}$; hence

$$\begin{aligned}\mu^*(A) &\leq \mu^*\left(\bigcup_{n=1}^{\infty} G_n\right) \quad \text{by (c)} \\ &= \mu\left(\bigcup_{n=1}^{\infty} G_n\right) \quad \text{by (a)} \\ &= \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=1}^n G_k\right) \quad \text{by 1.3.2(d).}\end{aligned}$$

The proof will be accomplished if we prove that

$$\mu\left(\bigcup_{i=1}^n G_i\right) \leq \mu^*(A_n) + \varepsilon \sum_{i=1}^n 2^{-i}, \quad n = 1, 2, \dots.$$

This is true for $n = 1$, by choice of G_1 . If it holds for a given n , we apply 1.3.2(b) to the sets $\bigcup_{i=1}^n G_i$ and G_{n+1} to obtain

$$\mu\left(\bigcup_{i=1}^{n+1} G_i\right) = \mu\left(\bigcup_{i=1}^n G_i\right) + \mu(G_{n+1}) - \mu\left[\left(\bigcup_{i=1}^n G_i\right) \cap G_{n+1}\right].$$

Now $(\bigcup_{i=1}^n G_i) \cap G_{n+1} \supset G_n \cap G_{n+1} \supset A_n \cap A_{n+1} = A_n$, so that the induction hypothesis yields

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{n+1} G_i\right) &\leq \mu^*(A_n) + \varepsilon \sum_{i=1}^n 2^{-i} + \mu^*(A_{n+1}) + \varepsilon 2^{-(n+1)} - \mu^*(A_n) \\ &\leq \mu^*(A_{n+1}) + \varepsilon \sum_{i=1}^{n+1} 2^{-i}. \blacksquare\end{aligned}$$

Our aim in this section is to prove that a σ -finite measure on a field \mathcal{F}_0 has a unique extension to the minimal σ -field over \mathcal{F}_0 . In fact an arbitrary measure μ on \mathcal{F}_0 can be extended to $\sigma(\mathcal{F}_0)$, but the extension is not necessarily unique. In proving this more general result (see Problem 3), the following concept plays a key role.

1.3.4 Definition. An *outer measure* on Ω is a nonnegative, extended real-valued set function λ on the class of all subsets of Ω , satisfying

- (a) $\lambda(\emptyset) = 0$,
- (b) $A \subset B$ implies $\lambda(A) \leq \lambda(B)$ (monotonicity),
- (c) $\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \lambda(A_n)$ (countable subadditivity).

The set function μ^* of 1.3.3 is an outer measure on Ω . Parts 1.3.4(a) and (b) follow from 1.3.3(a), 1.3.2(a), and 1.3.3(c), and 1.3.4(c) is proved as follows:

$$\begin{aligned}\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) &= \lim_{n \rightarrow \infty} \mu^*\left(\bigcup_{i=1}^n A_i\right) \quad \text{by 1.3.3(d).} \\ &\leq \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu^*(A_i) \quad \text{by 1.3.3(b),}\end{aligned}$$

as desired.

We now identify a σ -field on which μ^* is countably additive:

1.3.5 Theorem. Under the hypothesis of 1.3.3, let

$$\mathcal{H} = \{H \subset \Omega : \mu^*(H) + \mu^*(H^c) = 1\}$$

[$\mathcal{G} \subset \mathcal{H}$ by 1.3.2(a) and (b) and $\mathcal{H} = \{H \subset \Omega : \mu^*(H) + \mu^*(H^c) \leq 1\}$ by 1.3.3(b)]. Then \mathcal{H} is a σ -field and μ^* is a probability measure on \mathcal{H} .

PROOF. Clearly \mathcal{H} is closed under complementation, and $\Omega \in \mathcal{H}$ by 1.3.3(a) and 1.3.2(a). If $H_1, H_2 \subset \Omega$, then by 1.3.3(b),

$$\mu^*(H_1 \cup H_2) + \mu^*(H_1 \cap H_2) \leq \mu^*(H_1) + \mu^*(H_2) \quad (1)$$

and since

$$\begin{aligned}(H_1 \cup H_2)^c &= H_1^c \cap H_2^c, \quad (H_1 \cap H_2)^c = H_1^c \cup H_2^c, \\ \mu^*(H_1 \cup H_2)^c + \mu^*(H_1 \cap H_2)^c &\leq \mu^*(H_1^c) + \mu^*(H_2^c).\end{aligned} \quad (2)$$

If $H_1, H_2 \in \mathcal{H}$, add (1) and (2); the sum of the left sides is at least 2 by 1.3.3(b), and the sum of the right sides is 2. Thus the sum of the left sides is 2 as well. If $a = \mu^*(H_1 \cup H_2) + \mu^*(H_1 \cup H_2)^c$, $b = \mu^*(H_1 \cap H_2) + \mu^*(H_1 \cap H_2)^c$, then $a + b = 2$, hence $a \leq 1$ or $b \leq 1$. If $a \leq 1$, then $a = 1$, so $b = 1$ also. Consequently $H_1 \cup H_2 \in \mathcal{H}$ and $H_1 \cap H_2 \in \mathcal{H}$. We have therefore shown that \mathcal{H} is a field. Now equality holds in (1), for if not, the sum of the left sides of (1) and (2) would be less than the sum of the right sides, a contradiction. Thus μ^* is finitely additive on \mathcal{H} .

To show that \mathcal{H} is a σ -field, let $H_n \in \mathcal{H}$, $n = 1, 2, \dots$, $H_n \uparrow H$; $\mu^*(H) + \mu^*(H^c) \geq 1$ by 1.3.3(b). But $\mu^*(H) = \lim_{n \rightarrow \infty} \mu^*(H_n)$ by 1.3.3(d), hence for any $\varepsilon > 0$, $\mu^*(H) \leq \mu^*(H_n) + \varepsilon$ for large n . Since $\mu^*(H^c) \leq \mu^*(H_n^c)$ for all n by 1.3.3(c), and $H_n \in \mathcal{H}$, we have $\mu^*(H) + \mu^*(H^c) \leq 1 + \varepsilon$. Since ε is arbitrary, $H \in \mathcal{H}$, making \mathcal{H} a σ -field.

Since $\mu^*(H_n) \rightarrow \mu^*(H)$, μ^* is countably additive by 1.2.8(a). ■

We now have our first extension theorem:

1.3.6 Theorem. A finite measure on a field \mathcal{F}_0 can be extended to a measure on $\sigma(\mathcal{F}_0)$.

PROOF. Nothing is lost by considering a probability measure. The result then follows from 1.3.1–1.3.5 if we observe that $\mathcal{F}_0 \subset \mathcal{G} \subset \mathcal{H}$, hence $\sigma(\mathcal{F}_0) \subset \mathcal{H}$. Thus μ^* restricted to $\sigma(\mathcal{F}_0)$ is the desired extension. ■

In fact there is very little difference between $\sigma(\mathcal{F}_0)$ and \mathcal{H} ; if $B \in \mathcal{H}$, then B can be expressed as $A \cup N$, where $A \in \sigma(\mathcal{F}_0)$ and N is a subset of a set $M \in \sigma(\mathcal{F}_0)$ with $\mu^*(M) = 0$. To establish this, we introduce the idea of completion of a measure space.

1.3.7 Definitions. A measure μ on a σ -field \mathcal{F} is said to be *complete* iff whenever $A \in \mathcal{F}$ and $\mu(A) = 0$ we have $B \in \mathcal{F}$ for all $B \subset A$.

In 1.3.5, μ^* on \mathcal{H} is complete, for if $B \subset A \in \mathcal{H}$, $\mu^*(A) = 0$, then $\mu^*(B) + \mu^*(B^c) \leq \mu^*(A) + \mu^*(B^c) = \mu^*(B^c) \leq 1$; thus $B \in \mathcal{H}$.

The *completion* of a measure space $(\Omega, \mathcal{F}, \mu)$ is defined as follows. Let \mathcal{F}_μ be the class of sets $A \cup N$, where A ranges over \mathcal{F} and N over all subsets of sets of measure 0 in \mathcal{F} .

Now \mathcal{F}_μ is a σ -field including \mathcal{F} , for it is clearly closed under countable union, and if $A \cup N \in \mathcal{F}_\mu$, $N \subset M \in \mathcal{F}$, $\mu(M) = 0$, then $(A \cup N)^c = A^c \cap N^c = (A^c \cap M^c) \cup (A^c \cap (N^c - M^c))$ and $A^c \cap (N^c - M^c) = A^c \cap (M - N) \subset M$, so $(A \cup N)^c \in \mathcal{F}_\mu$.

We extend μ to \mathcal{F}_μ by setting $\mu(A \cup N) = \mu(A)$. This is a valid definition, for if $A_1 \cup N_1 = A_2 \cup N_2 \in \mathcal{F}_\mu$, we have

$$\mu(A_1) = \mu(A_1 \cap A_2) + \mu(A_1 - A_2) = \mu(A_1 \cap A_2)$$

since $A_1 - A_2 \subset N_2$. Thus $\mu(A_1) \leq \mu(A_2)$, and by symmetry, $\mu(A_1) = \mu(A_2)$. The measure space $(\Omega, \mathcal{F}_\mu, \mu)$ is called the *completion* of $(\Omega, \mathcal{F}, \mu)$, and \mathcal{F}_μ the completion of \mathcal{F} relative to μ .

Note that the completion is in fact complete, for if $M \subset A \cup N \in \mathcal{F}_\mu$ where $A \in \mathcal{F}$, $\mu(A) = 0$, $N \subset B \in \mathcal{F}$, $\mu(B) = 0$, then $M \subset A \cup B \in \mathcal{F}$, $\mu(A \cup B) = 0$; hence $M \in \mathcal{F}_\mu$.

1.3.8 Theorem. In 1.3.6, $(\Omega, \mathcal{H}, \mu^*)$ is the completion of $(\Omega, \sigma(\mathcal{F}_0), \mu^*)$.

PROOF. We must show that $\mathcal{H} = \mathcal{F}_{\mu^*}$, where $\mathcal{F} = \sigma(\mathcal{F}_0)$. If $A \in \mathcal{H}$, by definition of $\mu^*(A)$ and $\mu^*(A^c)$ we can find sets $G_n, G_n' \in \sigma(\mathcal{F}_0)$, $n = 1, 2, \dots$, with $G_n \subset A \subset G_n'$ and $\mu^*(G_n) \rightarrow \mu^*(A)$, $\mu^*(G_n') \rightarrow \mu^*(A)$. Let $G = \bigcup_{n=1}^{\infty} G_n$, $G' = \bigcap_{n=1}^{\infty} G_n'$. Then $A = G \cup (A - G)$, $G \in \sigma(\mathcal{F}_0)$, $A - G \subset G' - G \in \sigma(\mathcal{F}_0)$, $\mu^*(G' - G) \leq \mu^*(G_n' - G_n) \rightarrow 0$, so that $\mu^*(G' - G) = 0$. Thus $A \in \mathcal{F}_{\mu^*}$.

Conversely if $B \in \mathcal{F}_{\mu^*}$, then $B = A \cup N$, $A \in \mathcal{F}$, $N \subset M \in \mathcal{F}$, $\mu^*(M) = 0$. Since $\mathcal{F} \subset \mathcal{H}$ we have $A \in \mathcal{H}$, and since $(\Omega, \mathcal{H}, \mu^*)$ is complete we have $N \in \mathcal{H}$. Thus $B \in \mathcal{H}$. ■

To prove the uniqueness of the extension from \mathcal{F}_0 to \mathcal{F} , we need the following basic result:

1.3.9 Monotone Class Theorem. Let \mathcal{F}_0 be a field of subsets of Ω , and \mathcal{C} a class of subsets of Ω that is *monotone* (if $A_n \in \mathcal{C}$ and $A_n \uparrow A$ or $A_n \downarrow A$, then $A \in \mathcal{C}$). If $\mathcal{C} \supset \mathcal{F}_0$, then $\mathcal{C} \supset \sigma(\mathcal{F}_0)$, the minimal σ -field over \mathcal{F}_0 .

PROOF. The technique of the proof might be called “boot strapping.” Let $\mathcal{F} = \sigma(\mathcal{F}_0)$ and let \mathcal{M} be the smallest monotone class containing all sets of \mathcal{F}_0 . We show that $\mathcal{M} = \mathcal{F}$, in other words, *the smallest monotone class and the smallest σ -field over a field coincide*. The proof is completed by observing that $\mathcal{M} \subset \mathcal{C}$.

Fix $A \in \mathcal{M}$ and let $\mathcal{M}_A = \{B \in \mathcal{M} : A \cap B, A \cap B^c \text{ and } A^c \cap B \in \mathcal{M}\}$; then \mathcal{M}_A is a monotone class. In fact $\mathcal{M}_A = \mathcal{M}$; for if $A \in \mathcal{F}_0$, then $\mathcal{F}_0 \subset \mathcal{M}_A$ since \mathcal{F}_0 is a field, hence $\mathcal{M} \subset \mathcal{M}_A$ by minimality of \mathcal{M} ; consequently $\mathcal{M}_A = \mathcal{M}$. But this shows that for any $B \in \mathcal{M}$ we have $A \cap B, A \cap B^c, A^c \cap B \in \mathcal{M}$ for any $A \in \mathcal{F}_0$, so that $\mathcal{M}_B \supset \mathcal{F}_0$. Again by minimality of \mathcal{M} , $\mathcal{M}_B = \mathcal{M}$.

Now \mathcal{M} is a field (for if $A, B \in \mathcal{M} = \mathcal{M}_A$, then $A \cap B, A \cap B^c, A^c \cap B \in \mathcal{M}$) and a monotone class that is also a field is a σ -field (see 1.2.1), hence \mathcal{M} is a σ -field. Thus $\mathcal{F} \subset \mathcal{M}$ by minimality of \mathcal{F} , and in fact $\mathcal{F} = \mathcal{M}$ because \mathcal{F} is a monotone class including \mathcal{F}_0 . ■

We now prove the fundamental extension theorem.

1.3.10 Carathéodory Extension Theorem. Let μ be a measure on the field \mathcal{F}_0 of subsets of Ω , and assume that μ is σ -finite on \mathcal{F}_0 , so that Ω can be decomposed as $\bigcup_{n=1}^{\infty} A_n$, where $A_n \in \mathcal{F}_0$ and $\mu(A_n) < \infty$ for all n . Then μ has a unique extension to a measure on the minimal σ -field \mathcal{F} over \mathcal{F}_0 .

PROOF. Since \mathcal{F}_0 is a field, the A_n may be taken as disjoint (replace A_n by $A_1^c \cap \dots \cap A_{n-1}^c \cap A_n$, as in formula (3) of 1.1). Let $\mu_n(A) = \mu(A \cap A_n)$, $A \in \mathcal{F}_0$; then μ_n is a finite measure on \mathcal{F}_0 , hence by 1.3.6 it has an extension μ_n^* to \mathcal{F} . Since $\mu = \sum_n \mu_n$, the set function $\mu^* = \sum_n \mu_n^*$ is an extension of μ , and it is a measure on \mathcal{F} since the order of summation of any double series of nonnegative terms can be reversed.

Now suppose that λ is a measure on \mathcal{F} and $\lambda = \mu$ on \mathcal{F}_0 . Define $\lambda_n(A) = \lambda(A \cap A_n)$, $A \in \mathcal{F}$. Then λ_n is a finite measure on \mathcal{F} and $\lambda_n = \mu_n = \mu_n^*$ on \mathcal{F}_0 , and it follows that $\lambda_n = \mu_n^*$ on \mathcal{F} . For $\mathcal{C} = \{A \in \mathcal{F} : \lambda_n(A) = \mu_n^*(A)\}$ is a monotone class (by 1.2.7) that contains all sets of \mathcal{F}_0 , hence $\mathcal{C} = \mathcal{F}$ by 1.3.9. But then $\lambda = \sum_n \lambda_n = \sum_n \mu_n^* = \mu^*$, proving uniqueness. ■

The intuitive idea of constructing a minimal σ -field by forming complements and countable unions and intersections in all possible ways suggests that if \mathcal{F}_0 is a field and $\mathcal{F} = \sigma(\mathcal{F}_0)$, sets in \mathcal{F} can be approximated in some sense by sets in \mathcal{F}_0 . The following result formalizes this notion:

1.3.11 Approximation Theorem. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let \mathcal{F}_0 be a field of subsets of Ω such that $\sigma(\mathcal{F}_0) = \mathcal{F}$. Assume that μ is σ -finite on \mathcal{F}_0 , and let $\varepsilon > 0$ be given. If $A \in \mathcal{F}$ and $\mu(A) < \infty$, there is a set $B \in \mathcal{F}_0$ such that $\mu(A \Delta B) < \varepsilon$.

PROOF. Let \mathcal{G} be the class of all countable unions of sets of \mathcal{F}_0 . The conclusion of 1.3.11 holds for any $A \in \mathcal{G}$, by 1.2.7(a). By 1.3.3, if μ is finite and $A \in \mathcal{F}$, A can be approximated arbitrarily closely (in the sense of 1.3.11) by a set in \mathcal{G} , and therefore 1.3.11 is proved for finite μ . In general, let Ω be the disjoint union of sets $A_n \in \mathcal{F}_0$ with $\mu(A_n) < \infty$, and let $\mu_n(C) = \mu(C \cap A_n)$, $C \in \mathcal{F}$.

Then μ_n is a finite measure on \mathcal{F} , hence if $A \in \mathcal{F}$, there is a set $B_n \in \mathcal{F}_0$ such that $\mu_n(A \Delta B_n) < \varepsilon 2^{-n}$. Since

$$\begin{aligned}\mu_n(A \Delta B_n) &= \mu((A \Delta B_n) \cap A_n) \\ &= \mu[(A \Delta (B_n \cap A_n)) \cap A_n] = \mu_n(A \Delta (B_n \cap A_n)),\end{aligned}$$

and $B_n \cap A_n \in \mathcal{F}_0$, we may assume that $B_n \subset A_n$. (The observation that $B_n \cap A_n \in \mathcal{F}_0$ is the point where we use the hypothesis that μ is σ -finite on \mathcal{F}_0 , not merely on \mathcal{F} .) If $C = \bigcup_{n=1}^{\infty} B_n$, then $C \cap A_n = B_n$, so that

$$\mu_n(A \Delta C) = \mu((A \Delta C) \cap A_n) = \mu((A \Delta B_n) \cap A_n) = \mu_n(A \Delta B_n),$$

hence $\mu(A \Delta C) = \sum_{n=1}^{\infty} \mu_n(A \Delta C) < \varepsilon$. But $\bigcap_{k=1}^N B_k = A \uparrow C - A$ as $N \rightarrow \infty$,

and $A - \bigcup_{k=1}^N B_k \downarrow A = C$. If $A \in \mathcal{F}$ and $\mu(A) < \infty$, it follows from 1.2.7 that $\mu(A \Delta \bigcup_{k=1}^N B_k) \rightarrow \mu(A \Delta C)$ as $N \rightarrow \infty$, hence is less than ε for large enough N . Set $B = \bigcup_{k=1}^{\infty} B_k \in \mathcal{F}_0$. ■

1.3.12 Example. Let Ω be the rationals, \mathcal{F}_0 the field of finite disjoint unions of right-semiclosed intervals $(a, b] = \{\omega \in \Omega : a < \omega \leq b\}$, a, b rational [counting (a, ∞) and Ω itself as right-semiclosed; see 1.2.2]. Let $\mathcal{F} = \sigma(\mathcal{F}_0)$. Then:

- (a) \mathcal{F} consists of all subsets of Ω .
- (b) If $\mu(A)$ is the number of points in A (μ is counting measure), then μ is σ -finite on \mathcal{F} but not on \mathcal{F}_0 .
- (c) There are sets $A \in \mathcal{F}$ of finite measure that cannot be approximated by sets in \mathcal{F}_0 , that is, there is no sequence $A_n \in \mathcal{F}_0$ with $\mu(A \Delta A_n) \rightarrow 0$.
- (d) If $\lambda = 2\mu$, then $\lambda = \mu$ on \mathcal{F}_0 but not on \mathcal{F} .

Thus both the approximation theorem and the Carathéodory extension theorem fail in this case.

PROOF. (a) We have $\{x\} = \bigcap_{n=1}^{\infty} (x - (1/n), x]$, and therefore all singletons are in \mathcal{F} . But then all sets are in \mathcal{F} since Ω is countable.

(b) Since Ω is a countable union of singletons, μ is σ -finite on \mathcal{F} . But every nonempty set in \mathcal{F}_0 has infinite measure, so μ is not σ -finite on \mathcal{F}_0 .

(c) If A is any finite nonempty subset of Ω , then $\mu(A \Delta B) = \infty$ for all nonempty $B \in \mathcal{F}_0$, since any nonempty set in \mathcal{F}_0 must contain infinitely many points not in A .

(d) Since $\lambda\{x\} = 2$ and $\mu\{x\} = 1$, $\lambda \neq \mu$ on \mathcal{F} . But $\lambda(A) = \mu(A) = \infty$, $A \in \mathcal{F}_0$ (except for $A = \emptyset$). ■

Problems

- Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let \mathcal{F}_{μ} be the completion of \mathcal{F} relative to μ . If $A \subset \Omega$, define

$$\mu_0(A) = \sup\{\mu(B) : B \in \mathcal{F}, B \subset A\}, \quad \mu^0(A) = \inf\{\mu(B) : B \in \mathcal{F}, B \supset A\}.$$

If $A \in \mathcal{F}_{\mu}$, show that $\mu_0(A) = \mu^0(A) = \mu(A)$. Conversely, if $\mu_0(A) = \mu^0(A) < \infty$, show that $A \in \mathcal{F}_{\mu}$.

- Show that the monotone class theorem (1.3.9) fails if \mathcal{F}_0 is not assumed to be a field.

3. This problem deals with the extension of an arbitrary (not necessarily σ -finite) measure on a field.
- (a) Let λ be an outer measure on the set Ω (see 1.3.4). We say that the set E is λ -measurable iff

$$\lambda(A) = \lambda(A \cap E) + \lambda(A \cap E^c) \quad \text{for all } A \subset \Omega.$$

(The equals sign may be replaced by “ \geq ” by subadditivity of λ .) If \mathcal{M} is the class of all λ -measurable sets, show that \mathcal{M} is a σ -field, and that if E_1, E_2, \dots are disjoint sets in \mathcal{M} whose union is E , and $A \subset \Omega$, we have

$$\lambda(A \cap E) = \sum_n \lambda(A \cap E_n). \quad (1)$$

In particular, λ is a measure on \mathcal{M} . [Use the definition of λ -measurability to show that \mathcal{M} is a field and that (1) holds for finite sequences. If E_1, E_2, \dots are disjoint sets in \mathcal{M} and $F_n = \bigcup_{i=1}^n E_i \uparrow E$, show that

$$\lambda(A) \geq \lambda(A \cap F_n) + \lambda(A \cap E^c) = \sum_{i=1}^n \lambda(A \cap E_i) + \lambda(A \cap E^c),$$

and then let $n \rightarrow \infty$.]

- (b) Let μ be a measure on a field \mathcal{F}_0 of subsets of Ω . If $A \subset \Omega$, define

$$\mu^*(A) = \inf \left\{ \sum_n \mu(E_n) : A \subset \bigcup_n E_n, E_n \in \mathcal{F}_0 \right\}.$$

Show that μ^* is an outer measure on Ω and that $\mu^* = \mu$ on \mathcal{F}_0 .

- (c) In (b), if \mathcal{M} is the class of μ^* -measurable sets, show that $\mathcal{F}_0 \subset \mathcal{M}$. Thus by (a) and (b), μ may be extended to the minimal σ -field over \mathcal{F}_0 .
- (d) In (b), if μ is σ -finite on \mathcal{F}_0 , show that $(\Omega, \mathcal{M}, \mu^*)$ is the completion of $(\Omega, \sigma(\mathcal{F}_0), \mu)$.

1.4 Lebesgue–Stieltjes Measures and Distribution Functions

We are now in a position to construct a large class of measures on the Borel sets of R . If F is an increasing, right-continuous function from R to R , we set $\mu(a, b] = F(b) - F(a)$; we then extend μ to a finitely additive set function on the field $\mathcal{F}_0(R)$ of finite disjoint unions of right-semiclosed intervals. If we can show that μ is countably additive on $\mathcal{F}_0(R)$, the Carathéodory extension theorem extends μ to $\mathcal{B}(R)$.

1.4.1 Definitions. A *Lebesgue-Stieltjes measure* on R is a measure μ on $\mathcal{B}(R)$ such that $\mu(I) < \infty$ for each bounded interval I . A *distribution function* on R is a map $F: R \rightarrow R$ that is *increasing* [$a < b$ implies $F(a) \leq F(b)$] and *right-continuous* [$\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$]. We are going to show that the formula $\mu(a, b] = F(b) - F(a)$ sets up a one-to-one correspondence between Lebesgue-Stieltjes measures and distribution functions, where two distribution functions that differ by a constant are identified.

1.4.2 Theorem. Let μ be a Lebesgue-Stieltjes measure on R . Let $F: R \rightarrow R$ be defined, up to an additive constant, by $F(b) - F(a) = \mu(a, b]$. [For example, fix $F(0)$ arbitrarily and set $F(x) - F(0) = \mu(0, x]$, $x > 0$; $F(0) - F(x) = \mu(x, 0]$, $x < 0$.] Then F is a distribution function.

PROOF. If $a < b$, then $F(b) - F(a) = \mu(a, b] \geq 0$. If $\{x_n\}$ is a sequence of points such that $x_1 > x_2 > \dots > x$, then $F(x_n) - F(x) = \mu(x, x_n] \rightarrow 0$ by 1.2.7(b). ■

Now let F be a distribution function on R . It will be convenient to work in the compact space \bar{R} , so we extend F to a map of \bar{R} into \bar{R} by defining $F(\infty) = \lim_{x \rightarrow \infty} F(x)$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$; the limits exist by monotonicity. Define $\mu(a, b] = F(b) - F(a)$, $a, b \in \bar{R}$, $a < b$, and let $\mu[-\infty, b] = F(b) - F(-\infty) = \mu(-\infty, b]$; then μ is defined on all right-semiclosed intervals of \bar{R} (counting $[-\infty, b]$ as right-semiclosed; see 1.2.2).

If I_1, \dots, I_k are disjoint right-semiclosed intervals of \bar{R} , we define $\mu(\bigcup_{j=1}^k I_j) = \sum_{j=1}^k \mu(I_j)$. Thus μ is extended to the field $\mathcal{F}_0(\bar{R})$ of finite disjoint unions of right-semiclosed intervals of \bar{R} , and μ is finitely additive on $\mathcal{F}_0(\bar{R})$. To show that μ is in fact countably additive on $\mathcal{F}_0(\bar{R})$, we make use of 1.2.8(b), as follows:

1.4.3 Lemma. The set function μ is countably additive on $\mathcal{F}_0(\bar{R})$.

PROOF. First assume that $F(\infty) - F(-\infty) < \infty$, so that μ is finite. Let A_1, A_2, \dots be a sequence of sets in $\mathcal{F}_0(\bar{R})$ decreasing to \emptyset . If $(a, b]$ is one of the intervals of A_n , then by right continuity of F , $\mu(a', b] = F(b) - F(a') \rightarrow F(b) - F(a) = \mu(a, b]$ as $a' \rightarrow a$ from above.

Thus we can find sets $B_n \in \mathcal{F}_0(\bar{R})$ whose closures \bar{B}_n (in \bar{R}) are included in A_n , with $\mu(B_n)$ approximating $\mu(A_n)$. If $\varepsilon > 0$ is given, the finiteness of μ allows us to choose the B_n so that $\mu(A_n) - \mu(B_n) < \varepsilon 2^{-n}$. Now $\bigcap_{n=1}^{\infty} \bar{B}_n = \emptyset$, and it follows that $\bigcap_{k=1}^n \bar{B}_k = \emptyset$ for sufficiently large n . (Perhaps the easiest way to see this is to note that the sets $\bar{R} - \bar{B}_n$ form an open covering of the

compact set \bar{R} , hence there is a finite subcovering, so that $\bigcup_{k=1}^n (\bar{R} - \bar{B}_k) = \bar{R}$ for some n . Therefore $\bigcap_{k=1}^n \bar{B}_k = \emptyset$.) Now

$$\begin{aligned}\mu(A_n) &= \mu\left(A_n - \bigcap_{k=1}^n B_k\right) + \mu\left(\bigcap_{k=1}^n B_k\right) \\ &= \mu\left(A_n - \bigcap_{k=1}^n B_k\right) \\ &\leq \mu\left(\bigcup_{k=1}^n (A_k - B_k)\right) \quad \text{since } A_n \subset A_{n-1} \subset \dots \subset A_1 \\ &\leq \sum_{k=1}^n \mu(A_k - B_k) \quad \text{by 1.2.5(d)} \\ &< \varepsilon.\end{aligned}$$

Thus $\mu(A_n) \rightarrow 0$.

Now if $F(\infty) - F(-\infty) = \infty$, define $F_n(x) = F(x)$, $|x| \leq n$; $F_n(x) = F(n)$, $x \geq n$; $F_n(x) = F(-n)$, $x \leq -n$. If μ_n is the set function corresponding to F_n , then $\mu_n \leq \mu$ and $\mu_n \rightarrow \mu$ on $\mathcal{F}_0(\bar{R})$. Let A_1, A_2, \dots be disjoint sets in $\mathcal{F}_0(\bar{R})$ such that $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_0(\bar{R})$. Then $\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n)$ (Problem 5, Section 1.2) so if $\sum_{n=1}^{\infty} \mu(A_n) = \infty$, we are finished. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then

$$\begin{aligned}\mu(A) &= \lim_{n \rightarrow \infty} \mu_n(A) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} \mu_n(A_k)\end{aligned}$$

since the μ_n are finite. Now since $\sum_{k=1}^{\infty} \mu(A_k) < \infty$, we may write

$$\begin{aligned}0 &\leq \mu(A) - \sum_{k=1}^{\infty} \mu(A_k) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} [\mu_n(A_k) - \mu(A_k)] \\ &\leq 0 \quad \text{since } \mu_n \leq \mu.\blacksquare\end{aligned}$$

We now complete the construction of Lebesgue–Stieltjes measures:

1.4.4 Theorem. Let F be a distribution function on R , and let $\mu(a, b] = F(b) - F(a)$, $a < b$. There is a unique extension of μ to a Lebesgue–Stieltjes measure on R .

PROOF. Extend μ to a countably additive set function on $\mathcal{F}_0(\bar{R})$ as above. Let $\mathcal{F}_0(R)$ be the field of all finite disjoint unions of right-semiclosed intervals of R [counting (a, ∞) as right-semiclosed; see 1.2.2], and extend μ to $\mathcal{F}_0(R)$

as in the discussion after 1.4.2. [Take $\mu(a, \infty) = F(\infty) - F(a)$; $\mu(-\infty, b] = F(b) - F(-\infty)$, $a, b \in R$; $\mu(R) = F(\infty) - F(-\infty)$; note that there is no other possible choice for μ on these sets, by 1.2.7(a).] Now the map

$$\begin{aligned} (a, b] &\rightarrow (a, b] & \text{if } a, b \in R & \text{or if } b \in R, a = -\infty, \\ (a, \infty] &\rightarrow (a, \infty] & \text{if } a \in R & \text{or if } a = -\infty. \end{aligned}$$

sets up a one-to-one, μ -preserving correspondence between a subset of $\mathcal{F}_0(\bar{R})$ (everything in $\mathcal{F}_0(\bar{R})$ except sets including intervals of the form $[-\infty, b]$) and $\mathcal{F}_0(R)$. It follows that μ is countably additive on $\mathcal{F}_0(R)$. Furthermore, μ is σ -finite on $\mathcal{F}_0(R)$ since $\mu(-n, n] < \infty$; note that μ need not be σ -finite on $\mathcal{F}_0(\bar{R})$ since the sets $(-n, n]$ do not cover \bar{R} . By the Carathéodory extension theorem, μ has a unique extension to $\mathcal{B}(R)$. The extension is a Lebesgue-Stieltjes measure because $\mu(a, b] = F(b) - F(a) < \infty$ for $a, b \in R, a < b$. ■

1.4.5 Comments and Examples. If F is a distribution function and μ the corresponding Lebesgue-Stieltjes measure, we have seen that $\mu(a, b] = F(b) - F(a)$, $a < b$. The measure of any interval, right-semiclosed or not, may be expressed in terms of F . For if $F(x^-)$ denotes $\lim_{y \rightarrow x^-} F(y)$, then

$$\begin{array}{lll} (1) & \mu(a, b] = F(b) - F(a), & (3) \quad \mu[a, b] = F(b) - F(a^-), \\ (2) & \mu(a, b) = F(b^-) - F(a), & (4) \quad \mu[a, b) = F(b^-) - F(a^-). \end{array}$$

(Thus if F is continuous at a and b , all four expressions are equal.) For example, to prove (2), observe that

$$\mu(a, b) = \lim_{n \rightarrow \infty} \mu\left(a, b - \frac{1}{n}\right] = \lim_{n \rightarrow \infty} \left[F\left(b - \frac{1}{n}\right) - F(a)\right] = F(b^-) - F(a).$$

Statement (3) follows because

$$\mu[a, b] = \lim_{n \rightarrow \infty} \mu\left(a - \frac{1}{n}, b\right] = \lim_{n \rightarrow \infty} \left[F(b) - F\left(a - \frac{1}{n}\right)\right] = F(b) - F(a^-);$$

(4) is proved similarly. The proof of (3) works even if $a = b$, so that $\mu\{x\} = F(x) - F(x^-)$. Thus

(5) F is continuous at x iff $\mu\{x\} = 0$; the magnitude of a discontinuity of F at x coincides with the measure of $\{x\}$.

The following formulas are obtained from (1)–(3) by allowing a to approach $-\infty$ or b to approach $+\infty$.

$$\begin{array}{lll} (6) & \mu(-\infty, x] = F(x) - F(-\infty), & (9) \quad \mu[x, \infty) = F(\infty) - F(x^-), \\ (7) & \mu(-\infty, x) = F(x^-) - F(-\infty), & (10) \quad \mu(R) = F(\infty) - F(-\infty). \\ (8) & \mu(x, \infty) = F(\infty) - F(x), & \end{array}$$

(The formulas (6), (8), and (10) have already been observed in the proof of 1.4.4.)

If μ is finite, then F is bounded; since F may always be adjusted by an additive constant, nothing is lost in this case if we set $F(-\infty) = 0$.

We may now generate a large number of measures on $\mathcal{B}(R)$. For example, iff: $R \rightarrow R$, $f \geq 0$, and f is integrable (Riemann for now) on any finite interval, then if we fix $F(0)$ arbitrarily and define

$$\begin{aligned} F(x) - F(0) &= \int_0^x f(t) dt, \quad x > 0; \\ F(0) - F(x) &= \int_x^0 f(t) dt, \quad x < 0, \end{aligned}$$

then F is a (continuous) distribution function and thus gives rise to a Lebesgue–Stieltjes measure; specifically,

$$\mu(a, b] = \int_a^b f(x) dx.$$

In particular, we may take $f(x) = 1$ for all x , and $F(x) = x$; then $\mu(a, b] = b - a$. The set function μ is called the *Lebesgue measure* on $\mathcal{B}(R)$. The completion of $\mathcal{B}(R)$ relative to Lebesgue measure is called the class of *Lebesgue measurable sets*, written $\bar{\mathcal{B}}(R)$. Thus a Lebesgue measurable set is the union of a Borel set and a subset of a Borel set of Lebesgue measure 0. The extension of Lebesgue measure to $\bar{\mathcal{B}}(R)$ is called “Lebesgue measure” also.

Now let μ be a Lebesgue–Stieltjes measure that is concentrated on a countable set $S = \{x_1, x_2, \dots\}$, that is, $\mu(R - S) = 0$. [In general if $(\Omega, \mathcal{F}, \mu)$ is a measure space and $B \in \mathcal{F}$, we say that μ is concentrated on B iff $\mu(\Omega - B) = 0$.] In the present case, such a measure is easily constructed: If a_1, a_2, \dots are nonnegative numbers and $A \subset R$, set $\mu(A) = \sum \{a_i : x_i \in A\}$; μ is a measure on all subsets of R , not merely on the Borel sets (see 1.2.4). If $\mu(I) < \infty$ for each bounded interval I , μ will be a Lebesgue–Stieltjes measure on $\mathcal{B}(R)$; if $\sum_i a_i < \infty$, μ will be a finite measure. The distribution function F corresponding to μ is continuous on $R - S$; if $\mu\{x_n\} = a_n > 0$, F has a jump at x_n of magnitude a_n . If $x, y \in S$ and no point of S lies between x and y , then F is constant on $[x, y]$. For if $x \leq b < y$, then $F(b) - F(x) = \mu(x, b] = 0$.

Now if we take S to be the rational numbers, the above discussion yields a monotone function F from R to R that is continuous at each irrational point and discontinuous at each rational point.

If F is an increasing, right-continuous, real-valued function defined on a closed bounded interval $[a, b]$, there is a corresponding finite measure μ on the Borel subsets of $[a, b]$; explicitly, μ is determined by the requirement that $\mu(a', b'] = F(b') - F(a')$, $a \leq a' < b' \leq b$. The easiest way to establish the

correspondence is to extend F by defining $F(x) = F(b)$, $x \geq b$; $F(x) = F(a)$, $x \leq a$; then take μ as the Lebesgue-Stieltjes measure corresponding to F , restricted to $\mathcal{B}[a, b]$.

We are going to consider Lebesgue-Stieltjes measures and distribution functions in Euclidean n -space. First, some terminology:

1.4.6 Definitions and Comments. If $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n) \in R^n$, the interval $(a, b]$ is defined as $\{x = (x_1, \dots, x_n) \in R^n : a_i < x_i \leq b_i \text{ for all } i = 1, \dots, n\}$; (a, ∞) is defined as $\{x \in R^n : x_i > a_i \text{ for all } i = 1, \dots, n\}$, $(-\infty, b]$ as $\{x \in R^n : x_i \leq b_i \text{ for all } i = 1, \dots, n\}$; other types of intervals are defined similarly. The smallest σ -field containing all intervals $(a, b]$, $a, b \in R^n$, is called the class of *Borel sets* of R^n , written $\mathcal{B}(R^n)$. The Borel sets form the minimal σ -field over many other classes of sets, for example, the open sets, the intervals $[a, b]$, and so on, exactly as in the discussion of the one-dimensional case in 1.2.4. The class of Borel sets of \bar{R}^n , written $\mathcal{B}(\bar{R}^n)$, is defined similarly.

A *Lebesgue-Stieltjes measure* on R^n is a measure μ on $\mathcal{B}(R^n)$ such that $\mu(I) < \infty$ for each bounded interval I .

The notion of a distribution function on R^n , $n \geq 2$, is more complicated than in the one-dimensional case. To see why, assume for simplicity that $n = 3$, and let μ be a finite measure on $\mathcal{B}(R^3)$. Define

$$F(x_1, x_2, x_3) = \mu\{\omega \in R^3 : \omega_1 \leq x_1, \omega_2 \leq x_2, \omega_3 \leq x_3\}, \quad (x_1, x_2, x_3) \in R^3.$$

By analogy with the one-dimensional case, we expect that F is a distribution function corresponding to μ [see formula (6) of 1.4.5]. This will turn out to be correct, but the correspondence is no longer by means of the formula $\mu(a, b] = F(b) - F(a)$. To see this, we compute $\mu(a, b]$ in terms of F .

Introduce the *difference operator* \triangle as follows: If $G: R^n \rightarrow R$, $\triangle_{b_1 a_1} G(x_1, \dots, x_n)$ is defined as

$$G(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_n) - G(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_n).$$

1.4.7. Lemma. If $a \leq b$, that is, $a_i \leq b_i$, $i = 1, 2, 3$, then

- (a) $\mu(a, b] = \triangle_{b_1 a_1} \triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3)$, where
- (b)
$$\begin{aligned} \triangle_{b_1 a_1} \triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3) \\ = F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) \\ - F(a_1, a_2, a_3). \end{aligned}$$

Thus $\mu(a, b]$ is not simply $F(b) - F(a)$.

PROOF.

$$\begin{aligned}
 (a) \quad & \triangle_{b_3 a_3} F(x_1, x_2, x_3) \\
 &= F(x_1, x_2, b_3) - F(x_1, x_2, a_3) \\
 &= \mu\{\omega: \omega_1 \leq x_1, \omega_2 \leq x_2, \omega_3 \leq b_3\} \\
 &\quad - \mu\{\omega: \omega_1 \leq x_1, \omega_2 \leq x_2, \omega_3 \leq a_3\} \\
 &= \mu\{\omega: \omega_1 \leq x_1, \omega_2 \leq x_2, a_3 < \omega_3 \leq b_3\} \\
 &\text{since } a_3 \leq b_3.
 \end{aligned}$$

Similarly,

$$\triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3) = \mu\{\omega: \omega_1 \leq x_1, a_2 < \omega_2 \leq b_2, \\
 a_3 < \omega_3 \leq b_3\}$$

and

$$\triangle_{b_1 a_1} \triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3) = \mu\{\omega: a_1 < \omega_1 \leq b_1, a_2 < \omega_2 \leq b_2, \\
 a_3 < \omega_3 \leq b_3\}.$$

$$\begin{aligned}
 (b) \quad & \triangle_{b_3 a_3} F(x_1, x_2, x_3) = F(x_1, x_2, b_3) - F(x_1, x_2, a_3), \\
 & \triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3) = F(x_1, b_2, b_3) - F(x_1, a_2, b_3) \\
 &\quad - F(x_1, b_2, a_3) + F(x_1, a_2, a_3).
 \end{aligned}$$

Thus $\triangle_{b_1 a_1} \triangle_{b_2 a_2} \triangle_{b_3 a_3} F(x_1, x_2, x_3)$ is the desired expression. ■

The extension of 1.4.7 to n dimensions is clear.

1.4.8 Theorem. Let μ be a finite measure on $\mathcal{B}(R^n)$ and define

$$F(x) = \mu(-\infty, x] = \mu\{\omega: \omega_i \leq x_i, i = 1, \dots, n\}.$$

If $a \leq b$, then

- $\mu(a, b] = \triangle_{b_1 a_1} \cdots \triangle_{b_n a_n} F(x_1, \dots, x_n)$, where
- $\triangle_{b_1 a_1} \cdots \triangle_{b_n a_n} F(x_1, \dots, x_n) = F_0 - F_1 + F_2 - F_3 + \cdots + (-1)^n F_n$;

F_i is the sum of all $\binom{n}{i}$ terms of the form $F(c_1, \dots, c_n)$ with $c_k = a_k$ for exactly i integers in $\{1, 2, \dots, n\}$, and $c_k = b_k$ for the remaining $n - i$ integers.

PROOF. Apply the computations of 1.4.7. ■

We know that a distribution function on R determines a corresponding Lebesgue–Stieltjes measure. This is true in n dimensions if we change the definition of increasing.

Let $F: R^n \rightarrow R$, and, for $a \leq b$, let $F(a, b]$ denote

$$\triangle_{b_1, a_1} \cdots \triangle_{b_n, a_n} F(x_1, \dots, x_n).$$

The function F is said to be *increasing* iff $F(a, b] \geq 0$ whenever $a \leq b$; F is *right-continuous* iff it is right-continuous in all variables together, in other words, for any sequence $x^1 \geq x^2 \geq \cdots \geq x^k \geq \cdots \rightarrow x$ we have $F(x^k) \rightarrow F(x)$.

An increasing right-continuous $F: R^n \rightarrow R$ is said to be a *distribution function* on R^n . (Note that if F arises from a measure μ as in 1.4.8, F is a distribution function.)

If F is a distribution function on R^n , we set $\mu(a, b] = F(a, b]$ (this reduces to $F(b) - F(a)$ if $n = 1$). We are going to show that μ has a unique extension to a Lebesgue–Stieltjes measure on R^n . The technique of the proof is the same in any dimension, but to avoid cumbersome notation and to capture the essential ideas, we sometimes specialize to the case $n = 2$. We break the argument into several steps:

(1) If $a \leq a' \leq b' \leq b$, $I = (a, b]$ is the union of the nine disjoint intervals I_1, \dots, I_9 formed by first constraining the first coordinate in one of the following three ways:

$$a_1 < x \leq a'_1, \quad a'_1 < x \leq b'_1, \quad b'_1 < x \leq b_1,$$

and then constraining the second coordinate in one of the following three ways:

$$a_2 < y \leq a'_2, \quad a'_2 < y \leq b'_2, \quad b'_2 < y \leq b_2.$$

For example, a typical set in the union is

$$\{(x, y): b'_1 < x \leq b_1, \quad a_2 < y \leq a'_2\};$$

in n dimensions we would obtain 3^n such sets.

The result (1) may be verified by looking at Fig. 1.1.

(2) In (1), $F(I) = \sum_{j=1}^9 F(I_j)$, hence $a \leq a' \leq b' \leq b$ implies $F(a', b'] \leq F(a, b]$.

This is verified by brute force, using 1.4.8.

Now a *right-semiclosed interval* $(a, b]$ in \bar{R}^n is, by convention, a set of the form $\{(x_1, \dots, x_n): a_i < x_i \leq b_i, i = 1, \dots, n\}$, $a, b \in \bar{R}^n$, with the proviso that $a_i < x_i \leq b_i$ can be replaced by $a_i \leq x_i \leq b_i$ if $a_i = -\infty$. With this assumption, the set $\mathcal{F}_0(\bar{R}^n)$ of finite disjoint unions of right-semiclosed intervals is a field. (The corresponding convention in R^n is that $a_i < x_i \leq b_i$ can

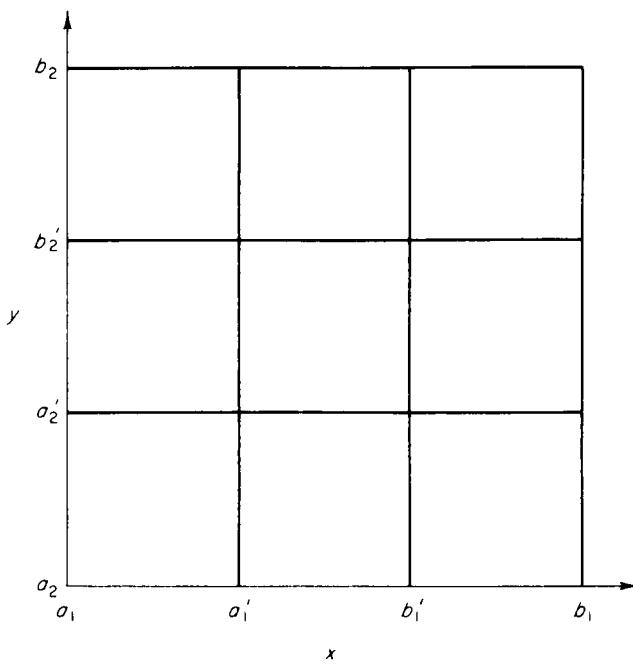


Figure 1.1.

be replaced by $a_i < x_i < b_i$ if $b_i = +\infty$. Both conventions are dictated by considerations similar to those of the one-dimensional case; see 1.2.2.)

(3) If a and b belong to \bar{R}^n but not to R^n , we define $F(a, b]$ as the limit of $F(a', b']$ where $a', b' \in R^n$, a' decreases to a , and b' increases to b . [The definition is sensible because of the monotonicity property in (2).] Similarly if $a \in R^n$, $b \in \bar{R}^n - R^n$, we take $F(a, b] = \lim_{b' \uparrow b} F(a, b']$; if $a \in \bar{R}^n - R^n$, $b \in R^n$, $F(a, b] = \lim_{a' \downarrow a} F(a', b]$.

Thus we define μ on right-semiclosed intervals of \bar{R}^n ; μ extends to a finitely additive set function on $\mathcal{F}_0(\bar{R}^n)$, as in the discussion after 1.4.2. [There is a slight problem here; a given interval I may be expressible as a finite disjoint union of intervals I_1, \dots, I_r , so that for the extension to be well defined we must have $F(I) = \sum_{j=1}^r F(I_j)$; but this follows just as in (2).]

(4) The set function μ is countably additive on $\mathcal{F}_0(\bar{R}^n)$.

First assume that $\mu(\bar{R}^n)$ is finite. If $a \in R^n$, $F(a', b] \rightarrow F(a, b]$ as a' decreases to a by the right-continuity of F and 1.4.8(b); if $a \in \bar{R}^n - R^n$, the same result holds by (3). The argument then proceeds word for word as in 1.4.3.

Now assume $\mu(\bar{R}^n) = \infty$. Then F , restricted to $C_k = \{x: -k < x_i \leq k, i = 1, \dots, n\}$, induces a finite-valued set function μ_k on $\mathcal{F}_0(\bar{R}^n)$ that is concentrated on C_k , so that $\mu_k(B) = \mu_k(B \cap C_k), B \in \mathcal{F}_0(\bar{R}^n)$. Since $\mu_k \leq \mu$ and $\mu_k \rightarrow \mu$ on $\mathcal{F}_0(\bar{R}^n)$, the proof of 1.4.3 applies verbatim.

1.4.9 Theorem. Let F be a distribution function on R^n , and let $\mu(a, b] = F(a, b], a, b \in R^n, a \leq b$. There is a unique extension of μ to a Lebesgue-Stieltjes measure on R^n .

PROOF. Repeat the proof of 1.4.4, with appropriate notational changes. For example, in extending μ to $\mathcal{F}_0(R^n)$, the field of finite disjoint unions of right-semiclosed intervals of R^n , we take (say for $n = 3$)

$$\mu\{(x, y, z): a_1 < x \leq b_1, a_2 < y < \infty, a_3 < z < \infty\} = \lim_{b_2, b_3 \rightarrow \infty} F(a, b].$$

The one-to-one μ -preserving correspondence is given by

$$(a, b] \rightarrow (a, b] \quad \text{if } a, b \in R^n \\ \text{or if } b \in R^n \text{ and at least one component of } a \text{ is } -\infty;$$

also, if the interval $\{(x_1, \dots, x_n): a_i < x_i \leq b_i: i = 1, \dots, n\}$ has some $b_i = \infty$, the corresponding interval in R^n has $a_i < x_i < \infty$. The remainder of the proof is as before. ■

1.4.10 Examples. (a) Let F_1, F_2, \dots, F_n be distribution functions on R , and define $F(x_1, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n)$. Then F is a distribution function on R^n since

$$F(a, b] = \prod_{i=1}^n [F_i(b_i) - F_i(a_i)].$$

In particular, if $F_i(x_i) = x_i, i = 1, \dots, n$, then each F_i corresponds to Lebesgue measure on $\mathcal{B}(R)$. In this case we have $F(x_1, \dots, x_n) = x_1 x_2 \cdots x_n$ and $\mu(a, b] = F(a, b] = \prod_{i=1}^n (b_i - a_i)$. Thus the measure of any rectangular box is its volume; μ is called *Lebesgue measure* on $\mathcal{B}(R^n)$. Just as in one dimension, the completion of $\mathcal{B}(R^n)$ relative to Lebesgue measure is called the class of *Lebesgue measurable sets* in R^n , written $\bar{\mathcal{B}}(R^n)$.

(b) Let f be a nonnegative function from R^n to R such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n < \infty.$$

(For now, we assume the integration is in the Riemann sense.) Define

$$F(x) = \int_{(-\infty, x]} f(t) dt,$$

that is,

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Then

$$\triangle_{b_n a_n} F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{n-1}} \int_{a_n}^{b_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n,$$

and we find by repeating this computation that

$$F(a, b] = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Thus F is a distribution function. If μ is the Lebesgue–Stieltjes measure determined by F , we have

$$\mu(a, b] = \int_{(a, b]} f(x) dx.$$

We have seen that if F is a distribution function on R^n , there is a unique Lebesgue–Stieltjes measure determined by $\mu(a, b] = F(a, b]$, $a \leq b$. Also, if μ is a finite measure on $\mathcal{B}(R^n)$ and $F(x) = \mu(-\infty, x]$, $x \in R^n$, then F is a distribution function on R^n and $\mu(a, b] = F(a, b]$, $a \leq b$. It is possible to associate a distribution function with an arbitrary Lebesgue–Stieltjes measure on R^n , and thus establish a one-to-one correspondence between Lebesgue–Stieltjes measures and distribution functions, provided distribution functions with the same increments $F(a, b]$, $a, b \in R^n$, $a \leq b$, are identified. The result will not be needed, and the details are quite tedious and will be omitted.

Problems

- Let F be the distribution function on R given by $F(x) = 0$, $x < -1$; $F(x) = 1 + x$, $-1 \leq x < 0$; $F(x) = 2 + x^2$, $0 \leq x < 2$; $F(x) = 9$, $x \geq 2$. If μ is the Lebesgue–Stieltjes measure corresponding to F , compute the measure of each of the following sets:
 - $\{2\}$,
 - $[-\frac{1}{2}, 3)$,
 - $(-1, 0] \cup (1, 2)$,
 - $[0, \frac{1}{2}) \cup (1, 2]$,
 - $\{x: |x| + 2x^2 > 1\}$.

2. Let μ be a Lebesgue-Stieltjes measure on R corresponding to a continuous distribution function.
 - (a) If A is a countable subset of R , show that $\mu(A) = 0$.
 - (b) If $\mu(A) > 0$, must A include an open interval?
 - (c) If $\mu(A) > 0$ and $\mu(R - A) = 0$, must A be dense in R ?
 - (d) Do the answers to (b) or (c) change if μ is restricted to be Lebesgue measure?
3. If B is a Borel set in R^n and $a \in R^n$, show that $a + B = \{a + x: x \in B\}$ is a Borel set, and $-B = \{-x: x \in B\}$ is a Borel set. (Use the good sets principle.)
4. Show that if $B \in \mathcal{B}(R^n)$, $a \in R^n$, then $a + B \in \mathcal{B}(R^n)$ and $\mu(a + B) = \mu(B)$, where μ is Lebesgue measure. Thus Lebesgue measure is translation-invariant. (The good sets principle works here also, in conjunction with the monotone class theorem.)
5. Let μ be a Lebesgue-Stieltjes measure on $\mathcal{B}(R^n)$ such that $\mu(a + I) = \mu(I)$ for all $a \in R^n$ and all (right-semiclosed) intervals I in R^n . In other words, μ is translation-invariant on intervals. Show that μ is a constant times Lebesgue measure.
6. (*A set that is not Lebesgue measurable*) Call two real numbers x and y equivalent iff $x - y$ is rational. Choose a member of each distinct equivalence class $B_x = \{y: y - x \text{ rational}\}$ to form a set A ; assume that the representatives are chosen so that $A \subset [0, 1]$. Establish the following:
 - (a) If r and s are distinct rational numbers, $(r + A) \cap (s + A) = \emptyset$; also $R = \bigcup \{r + A: r \text{ rational}\}$.
 - (b) If A is Lebesgue measurable (so that $r + A$ is Lebesgue measurable by Problem 4), then $\mu(r + A) = 0$ for all rational r (μ is Lebesgue measure). Conclude that A cannot be Lebesgue measurable.

The only properties of Lebesgue measure needed in this problem are translation-invariance and finiteness on bounded intervals. Therefore, the result implies that there is no translation-invariant measure λ (except $\lambda \equiv 0$) on the class of all subsets of R such that $\lambda(I) < \infty$ for each bounded interval I .
7. (*The Cantor ternary set*) Let E_1 be the middle third of the interval $[0, 1]$, that is, $E_1 = (\frac{1}{3}, \frac{2}{3})$; thus $x \in [0, 1] - E_1$ iff x can be written in ternary form using 0 or 2 in the first digit. Let E_2 be the union of the middle thirds of the two intervals that remain after E_1 is removed, that is, $E_2 = (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$; thus $x \in [0, 1] - (E_1 \cup E_2)$ iff x can be written in ternary form using 0 or 2 in the first two digits. Continue the construction; let E_n be the union of the middle thirds of the intervals that remain after E_1, \dots, E_{n-1} are removed. The Cantor ternary set C is defined as $[0, 1] - \bigcup_{n=1}^{\infty} E_n$; thus $x \in C$ iff x can be expressed in

ternary form using only digits 0 and 2. Various topological properties of C follow from the definition: C is closed, perfect (every point of C is a limit point of C), and nowhere dense.

Show that C is uncountable and has Lebesgue measure 0.

Comment. In the above construction, we have $m(E_n) = (\frac{1}{3})(\frac{2}{3})^{n-1}$, $n = 1, 2, \dots$, where m is Lebesgue measure. If $0 < \alpha < 1$, the procedure may be altered slightly so that $m(E_n) = \alpha(\frac{1}{3})^n$. We then obtain a set $C(\alpha)$, homeomorphic to C , of measure $1 - \alpha$; such sets are called *Cantor sets of positive measure*.

8. Give an example of a function $F: R^2 \rightarrow R$ such that F is right-continuous and is increasing in each coordinate separately, but F is not a distribution function on R^2 .
9. A distribution function on R is monotone and thus has only countably many points of discontinuity. Is this also true for a distribution function on R^n , $n > 1$?
10. (a) Let F and G be distribution functions on R^n . If $F(a, b] = G(a, b]$ for all $a, b \in R^n$, $a \leq b$, does it follow that F and G differ by a constant?
 (b) Must a distribution function on R^n be increasing in each coordinate separately?
11. If c is the cardinality of the reals, show that there are only c Borel subsets of R^n , but 2^c Lebesgue measurable sets.
12. The following result shows that under appropriate conditions, a Borel set can be approximated from below by a compact set, and from above by an open set. Problems of this type will be examined systematically in Chapter 4.
 If μ is a σ -finite measure on $\mathcal{B}(R^n)$, show that for each $B \in \mathcal{B}(R^n)$,
 - (a) $\mu(B) = \sup\{\mu(K): K \subset B, K \text{ compact}\}$.
 - If μ is in fact a Lebesgue-Stieltjes measure, show that
 - (b) $\mu(B) = \inf\{\mu(V): V \supset B, B \text{ open}\}$.
 - (c) Give an example of a σ -finite measure on $\mathcal{B}(R^n)$ that is not a Lebesgue-Stieltjes measure and for which (b) fails.

1.5 Measurable Functions and Integration

If f is a real-valued function defined on a bounded interval $[a, b]$ of reals, we can talk about the Riemann integral of f , at least if f is piecewise continuous. We are going to develop a much more general integration process, one that applies to functions from an arbitrary set to the extended reals, provided that certain “measurability” conditions are satisfied.

Probability considerations may again be used to motivate the concept of measurability. Suppose that (Ω, \mathcal{F}, P) is a probability space, and that h is a function from Ω to \bar{R} . Thus if the outcome of the experiment corresponds to the point $\omega \in \Omega$, we may compute the number $h(\omega)$. Suppose that we are interested in the probability that $a \leq h(\omega) \leq b$, in other words, we wish to compute $P\{\omega: h(\omega) \in B\}$ where $B = [a, b]$. For this to be possible, the set $\{\omega: h(\omega) \in B\} = h^{-1}(B)$ must belong to the σ -field \mathcal{F} . If $h^{-1}(B) \in \mathcal{F}$ for each interval B (and hence, as we shall see below, for each Borel set B), then h is a “measurable function,” in other words, probabilities of events involving h can be computed. In the language of probability theory, h is a “random variable.”

1.5.1 Definitions and Comments. If $h: \Omega_1 \rightarrow \Omega_2$, h is measurable relative to the σ -fields \mathcal{F}_j of subsets of Ω_j , $j = 1, 2$, iff $h^{-1}(A) \in \mathcal{F}_1$ for each $A \in \mathcal{F}_2$.

It is sufficient that $h^{-1}(A) \in \mathcal{F}_1$ for each $A \in \mathcal{C}$, where \mathcal{C} is a class of subsets of Ω_2 such that the minimal σ -field over \mathcal{C} is \mathcal{F}_2 . For $\{A \in \mathcal{F}_2 : h^{-1}(A) \in \mathcal{F}_1\}$ is a σ -field that contains all sets of \mathcal{C} , hence coincides with \mathcal{F}_2 . This is another application of the good sets principle.

The notation $h: (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ will mean that $h: \Omega_1 \rightarrow \Omega_2$, measurable relative to \mathcal{F}_1 and \mathcal{F}_2 .

If \mathcal{F} is a σ -field of subsets of Ω , (Ω, \mathcal{F}) is sometimes called a *measurable space*, and the sets in \mathcal{F} are sometimes called *measurable sets*.

Notice that measurability of h does not imply that $h(A) \in \mathcal{F}_2$ for each $A \in \mathcal{F}_1$. For example, if $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$, then any $h: \Omega_1 \rightarrow \Omega_2$ is measurable, regardless of \mathcal{F}_1 , but if $A \in \mathcal{F}_1$ and $h(A)$ is a nonempty proper subset of Ω_2 , then $h(A) \notin \mathcal{F}_2$. Actually, in measure theory, the inverse image is a much more desirable object than the direct image since the basic set operations are preserved by inverse images but not in general by direct images. Specifically, we have $h^{-1}(\bigcup_i B_i) = \bigcup_i h^{-1}(B_i)$, $h^{-1}(\bigcap_i B_i) = \bigcap_i h^{-1}(B_i)$, and $h^{-1}(B^c) = [h^{-1}(B)]^c$. We also have $h(\bigcup_i A_i) = \bigcup_i h(A_i)$, but $h(\bigcap_i A_i) \subset \bigcap_i h(A_i)$, and the inclusion may be proper. Furthermore, $h(A^c)$ need not equal $[h(A)]^c$, in fact when h is a constant function the two sets are disjoint.

If (Ω, \mathcal{F}) is a measurable space and $h: \Omega \rightarrow R^n$ (or \bar{R}^n), h is said to be *Borel measurable* [on (Ω, \mathcal{F})] iff h is measurable relative to the σ -fields \mathcal{F} and \mathcal{B} , the class of Borel sets. If Ω is a Borel subset of R^k (or \bar{R}^k) and we use the term “Borel measurable,” we always assume that $\mathcal{F} = \mathcal{B}$.

A continuous map h from R^k to R^n is Borel measurable; for if \mathcal{C} is the class of open subsets of R^n , then $h^{-1}(A)$ is open, hence belongs to $\mathcal{B}(R^k)$, for each $A \in \mathcal{C}$.

If A is a subset of R that is not a Borel set (Section 1.4, Problems 6 and 11) and I_A is the *indicator* of A , that is, $I_A(\omega) = 1$ for $\omega \in A$ and 0 for $\omega \notin A$, then I_A is not Borel measurable; for $\{\omega: I_A(\omega) = 1\} = A \notin \mathcal{B}(R)$.

To show that a function $h: \Omega \rightarrow R$ (or \bar{R}) is Borel measurable, it is sufficient to show that $\{\omega: h(\omega) > c\} \in \mathcal{F}$ for each real c . For if \mathcal{C} is the class of sets $\{x: x > c\}$, $c \in R$, then $\sigma(\mathcal{C}) = \mathcal{B}(R)$. Similarly, $\{\omega: h(\omega) > c\}$ can be replaced by $\{\omega: h(\omega) \geq c\}$, $\{\omega: h(\omega) < c\}$ or $\{\omega: h(\omega) \leq c\}$, or equally well by $\{\omega: a \leq h(\omega) \leq b\}$ for all real a and b , and so on.

If $(\Omega, \mathcal{F}, \mu)$ is a measure space the terminology “ h is Borel measurable on $(\Omega, \mathcal{F}, \mu)$ ” will mean that h is Borel measurable on (Ω, \mathcal{F}) and μ is a measure on \mathcal{F} .

1.5.2 Definition. Let (Ω, \mathcal{F}) be a measurable space, fixed throughout the discussion. If $h: \Omega \rightarrow \bar{R}$, h is said to be *simple* iff h is Borel measurable and takes on only finitely many distinct values. Equivalently, h is simple iff it can be written as a finite sum $\sum_{i=1}^r x_i I_{A_i}$ where the A_i are disjoint sets in \mathcal{F} and I_{A_i} is the indicator of A_i ; the x_i need not be distinct.

We assume the standard arithmetic of \bar{R} ; if $a \in R$, $a + \infty = \infty$, $a - \infty = -\infty$, $a/\infty = a/-\infty = 0$, $a \cdot \infty = \infty$ if $a > 0$, $a \cdot \infty = -\infty$ if $a < 0$, $0 \cdot \infty = 0 \cdot (-\infty) = 0$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$, with commutativity of addition and multiplication. It is then easy to check that sums, differences, products, and quotients of simple functions are simple, as long as the operations are well-defined, in other words we do not try to add $+\infty$ and $-\infty$, divide by 0, or divide ∞ by ∞ .

Let μ be a measure on \mathcal{F} , again fixed throughout the discussion. If $h: \Omega \rightarrow \bar{R}$ is Borel measurable we are going to define the *abstract Lebesgue integral* of h with respect to μ , written as $\int_{\Omega} h d\mu$, $\int_{\Omega} h(\omega) \mu(d\omega)$, or $\int_{\Omega} h(\omega) d\mu(\omega)$.

1.5.3 Definition of the Integral. First let h be simple, say $h = \sum_{i=1}^r x_i I_{A_i}$ where the A_i are disjoint sets in \mathcal{F} . We define

$$\int_{\Omega} h d\mu = \sum_{i=1}^r x_i \mu(A_i)$$

as long as $+\infty$ and $-\infty$ do not both appear in the sum; if they do, we say that the integral does not exist. Strictly speaking, it must be verified that if h has a different representation, say $\sum_{j=1}^s y_j I_{B_j}$, then

$$\sum_{i=1}^r x_i \mu(A_i) = \sum_{j=1}^s y_j \mu(B_j).$$

(For example, if $A = B \cup C$, where $B \cap C = \emptyset$, then $xI_A = xI_B + xI_C$.) The proof is based on the observation that

$$h = \sum_{i=1}^r \sum_{j=1}^s z_{ij} I_{A_i \cap B_j},$$

where $z_{ij} = x_i = y_j$. Thus

$$\begin{aligned}\sum_{i,j} z_{ij} \mu(A_i \cap B_j) &= \sum_i x_i \sum_j \mu(A_i \cap B_j) \\ &= \sum_i x_i \mu(A_i) \\ &= \sum_j y_j \mu(B_j) \quad \text{by a symmetrical argument.}\end{aligned}$$

If h is nonnegative Borel measurable, define

$$\int_{\Omega} h \, d\mu = \sup \left\{ \int_{\Omega} s \, d\mu : s \text{ simple, } 0 \leq s \leq h \right\}.$$

This agrees with the previous definition if h is simple. Furthermore, we may if we like restrict s to be finite-valued.

Notice that according to the definition, the integral of a nonnegative Borel measurable function always exists; it may be $+\infty$.

Finally, if h is an arbitrary Borel measurable function, let $h^+ = \max(h, 0)$, $h^- = \max(-h, 0)$, that is,

$$\begin{aligned}h^+(\omega) &= h(\omega) \quad \text{if } h(\omega) \geq 0; \quad h^+(\omega) = 0 \quad \text{if } h(\omega) < 0; \\ h^-(\omega) &= -h(\omega) \quad \text{if } h(\omega) \leq 0; \quad h^-(\omega) = 0 \quad \text{if } h(\omega) > 0.\end{aligned}$$

The function h^+ is called the *positive part* of h , h^- the *negative part*. We have $|h| = h^+ + h^-$, $h = h^+ - h^-$, and h^+ and h^- are Borel measurable. For example, $\{\omega : h^+(\omega) \in B\} = \{\omega : h(\omega) \geq 0, h(\omega) \in B\} \cup \{\omega : h(\omega) < 0, 0 \in B\}$. The first set is $h^{-1}[0, \infty] \cap h^{-1}(B) \in \mathcal{F}$; the second is $h^{-1}[-\infty, 0]$ if $0 \in B$, and \emptyset if $0 \notin B$. Thus $(h^+)^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{B}(\bar{R})$, and similarly for h^- . Alternatively, if h_1 and h_2 are Borel measurable, then $\max(h_1, h_2)$ and $\min(h_1, h_2)$ are Borel measurable; to see this, note that

$$\{\omega : \max(h_1(\omega), h_2(\omega)) \leq c\} = \{\omega : h_1(\omega) \leq c\} \cap \{\omega : h_2(\omega) \leq c\}$$

and $\{\omega : \min(h_1(\omega), h_2(\omega)) \leq c\} = \{\omega : h_1(\omega) \leq c\} \cup \{\omega : h_2(\omega) \leq c\}$. It follows that if h is Borel measurable, so are h^+ and h^- .

We define

$$\int_{\Omega} h \, d\mu = \int_{\Omega} h^+ \, d\mu - \int_{\Omega} h^- \, d\mu \quad \text{if this is not of the form } +\infty - \infty;$$

if it is, we say that the integral does not exist. The function h is said to be μ -integrable (or simply integrable if μ is understood) iff $\int_{\Omega} h \, d\mu$ is finite, that is, iff $\int_{\Omega} h^+ \, d\mu$ and $\int_{\Omega} h^- \, d\mu$ are both finite.

If $A \in \mathcal{F}$, we define

$$\int_A h \, d\mu = \int_{\Omega} h I_A \, d\mu.$$

(The proof that $h I_A$ is Borel measurable is similar to the first proof above that h^+ is Borel measurable.)

If h is a step function from R to R and μ is Lebesgue measure, $\int_R h d\mu$ agrees with the Riemann integral. However, the integral of h with respect to Lebesgue measure exists for many functions that are not Riemann integrable, as we shall see in Section 1.7.

Before examining the properties of the integral, we need to know more about Borel measurable functions. One of the basic reasons why such functions are useful in analysis is that a pointwise limit of Borel measurable functions is still Borel measurable.

1.5.4 Theorem. If h_1, h_2, \dots are Borel measurable functions from Ω to \bar{R} and $h_n(\omega) \rightarrow h(\omega)$ for all $\omega \in \Omega$, then h is Borel measurable.

PROOF. It is sufficient to show that $\{\omega : h(\omega) > c\} \in \mathcal{F}$ for each real c . We have

$$\begin{aligned}\{\omega : h(\omega) > c\} &= \left\{ \omega : \lim_{n \rightarrow \infty} h_n(\omega) > c \right\} \\ &= \left\{ \omega : h_n(\omega) \text{ is eventually } > c + \frac{1}{r} \text{ for some } r = 1, 2, \dots \right\} \\ &= \bigcup_{r=1}^{\infty} \left\{ \omega : h_n(\omega) > c + \frac{1}{r} \text{ for all but finitely many } n \right\} \\ &= \bigcup_{r=1}^{\infty} \liminf_n \left\{ \omega : h_n(\omega) > c + \frac{1}{r} \right\} \\ &= \bigcup_{r=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{ \omega : h_k(\omega) > c + \frac{1}{r} \right\} \in \mathcal{F}. \quad \blacksquare\end{aligned}$$

To show that the class of Borel measurable functions is closed under algebraic operations, we need the following basic approximation theorem.

1.5.5 Theorem. (a) A nonnegative Borel measurable function h is the limit of an increasing sequence of nonnegative, finite-valued, simple functions h_n .

(b) An arbitrary Borel measurable function f is the limit of a sequence of finite-valued simple functions f_n , with $|f_n| \leq |f|$ for all n .

PROOF. (a) Define

$$h_n(\omega) = \frac{k-1}{2^n} \quad \text{if} \quad \frac{k-1}{2^n} \leq h(\omega) < \frac{k}{2^n}, \quad k = 1, 2, \dots, n2^n,$$

and let $h_n(\omega) = n$ if $h(\omega) \geq n$. [Or equally well, $h_n(\omega) = (k-1)/2^n$ if $(k-1)/2^n < h(\omega) \leq k/2^n$, $k = 1, 2, \dots, n2^n$; $h_n(\omega) = n$ if $h(\omega) > n$; $h_n(\omega) = 0$ if $h(\omega) = 0$.] The h_n have the desired properties (Problem 1).

(b) Let g_n and h_n be nonnegative, finite-valued, simple functions with $g_n \uparrow f^+$ and $h_n \uparrow f^-$; take $f_n = g_n - h_n$. ■

1.5.6 Theorem. If h_1 and h_2 are Borel measurable functions from Ω to \bar{R} , so are $h_1 + h_2$, $h_1 - h_2$, $h_1 h_2$, and h_1/h_2 [assuming these are well-defined, in other words, $h_1(\omega) + h_2(\omega)$ is never of the form $+\infty - \infty$ and $h_1(\omega)/h_2(\omega)$ is never of the form ∞/∞ or $a/0$].

PROOF. As in 1.5.5, let s_{1n}, s_{2n} be finite-valued simple functions with $s_{1n} \rightarrow h_1$, $s_{2n} \rightarrow h_2$. Then $s_{1n} + s_{2n} \rightarrow h_1 + h_2$,

$$s_{1n} s_{2n} I_{\{h_1 \neq 0\}} I_{\{h_2 \neq 0\}} \rightarrow h_1 h_2,$$

and

$$\frac{s_{1n}}{s_{2n} + (1/n)I_{\{s_{2n}=0\}}} \rightarrow \frac{h_1}{h_2}.$$

Since

$$s_{1n} \pm s_{2n}, \quad s_{1n} s_{2n} I_{\{h_1 \neq 0\}} I_{\{h_2 \neq 0\}}, \quad s_{1n} \left(s_{2n} + \frac{1}{n} I_{\{s_{2n}=0\}} \right)^{-1}$$

are simple, the result follows from 1.5.4. ■

We are going to extend 1.5.4 and part of 1.5.6 to Borel measurable functions from Ω to \bar{R}^n ; to do this, we need the following useful result.

1.5.7 Lemma. A composition of measurable functions is measurable; specifically, if $g: (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ and $h: (\Omega_2, \mathcal{F}_2) \rightarrow (\Omega_3, \mathcal{F}_3)$, then $h \circ g: (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_3, \mathcal{F}_3)$.

PROOF. If $B \in \mathcal{F}_3$, then $(h \circ g)^{-1}(B) = g^{-1}(h^{-1}(B)) \in \mathcal{F}_1$. ■

Since some books contain the statement “A composition of measurable functions need not be measurable,” some explanation is called for. If $h: R \rightarrow R$, some authors call h “measurable” iff the preimage of a Borel set is a Lebesgue measurable set. We shall call such a function *Lebesgue measurable*. Note that every Borel measurable function is Lebesgue measurable, but not conversely. (Consider the indicator of a Lebesgue measurable set that is not a

Borel set; see Section 1.4, Problem 11.) If g and h are Lebesgue measurable, the composition $h \circ g$ need not be Lebesgue measurable. For let \mathcal{B} be the Borel sets, and $\bar{\mathcal{B}}$ the Lebesgue measurable sets. If $B \in \mathcal{B}$ then $h^{-1}(B) \in \bar{\mathcal{B}}$; but $g^{-1}(h^{-1}(B))$ is known to belong to $\bar{\mathcal{B}}$ only when $h^{-1}(B) \in \mathcal{B}$, so we cannot conclude that $(h \circ g)^{-1}(B) \in \mathcal{B}$. For an explicit example, see Royden (1968, p. 70). If $g^{-1}(A) \in \bar{\mathcal{B}}$ for all $A \in \mathcal{B}$, not just for all $A \in \mathcal{B}$, then we are in the situation described in Lemma 1.5.7, and $h \circ g$ is Lebesgue measurable; similarly, if h is Borel measurable (and g is Lebesgue measurable), then $h \circ g$ is Lebesgue measurable.

It is rarely necessary to replace Borel measurability of functions from R to R (or R^k to R^n) by the slightly more general concept of Lebesgue measurability; in this book, the only instance is in Section 1.7. The integration theory that we are developing works for extended real-valued functions on an arbitrary measure space $(\Omega, \mathcal{F}, \mu)$. Thus there is no problem in integrating Lebesgue measurable functions; set $\Omega = R$, $\mathcal{F} = \bar{\mathcal{B}}$.

We may now assert that if h_1, h_2, \dots are Borel measurable functions from Ω to \bar{R}^n and h_n converges pointwise to h , then h is Borel measurable; furthermore, if h_1 and h_2 are Borel measurable functions from Ω to \bar{R}^n , so are $h_1 + h_2$ and $h_1 - h_2$, assuming these are well-defined. The reason is that if $h(\omega) = (h_1(\omega), \dots, h_n(\omega))$ describes a map from Ω to \bar{R}^n , Borel measurability of h is equivalent to Borel measurability of all the component functions h_i .

1.5.8 Theorem. Let $h: \Omega \rightarrow \bar{R}^n$; if p_i is the projection map of \bar{R}^n onto \bar{R} , taking (x_1, \dots, x_n) to x_i , set $h_i = p_i \circ h$, $i = 1, \dots, n$. Then h is Borel measurable iff h_i is Borel measurable for all $i = 1, \dots, n$.

PROOF. Assume h Borel measurable. Since

$$p_i^{-1}\{x_i : a_i \leq x_i \leq b_i\} = \{x \in \bar{R}^n : a_i \leq x_i \leq b_i, -\infty \leq x_j \leq \infty, j \neq i\},$$

which is an interval of \bar{R}^n , p_i is Borel measurable. Thus

$$h: (\Omega, \mathcal{F}) \rightarrow (\bar{R}^n, \mathcal{B}(\bar{R}^n)), \quad p_i: (\bar{R}^n, \mathcal{B}(\bar{R}^n)) \rightarrow (\bar{R}, \mathcal{B}(\bar{R})),$$

and therefore by 1.5.7, $h_i: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$.

Conversely, assume each h_i to be Borel measurable. Then

$$h^{-1}\{x \in \bar{R}^n : a_i \leq x_i \leq b_i, i = 1, \dots, n\} = \bigcap_{i=1}^n \{\omega \in \Omega : a_i \leq h_i(\omega) \leq b_i\} \in \mathcal{F},$$

and the result follows. ■

We now proceed to some properties of the integral. In the following result, all functions are assumed Borel measurable from Ω to \bar{R} .

1.5.9 Theorem. (a) If $\int_{\Omega} h d\mu$ exists and $c \in R$, then $\int_{\Omega} ch d\mu$ exists and equals $c \int_{\Omega} h d\mu$.

(b) If $g(\omega) \geq h(\omega)$ for all ω , then $\int_{\Omega} g d\mu \geq \int_{\Omega} h d\mu$ in the sense that if $\int_{\Omega} h d\mu$ exists and is greater than $-\infty$, then $\int_{\Omega} g d\mu$ exists and $\int_{\Omega} g d\mu \geq \int_{\Omega} h d\mu$; if $\int_{\Omega} g d\mu$ exists and is less than $+\infty$, then $\int_{\Omega} h d\mu$ exists and $\int_{\Omega} h d\mu \leq \int_{\Omega} g d\mu$. Thus if both integrals exist, $\int_{\Omega} g d\mu \geq \int_{\Omega} h d\mu$, whether or not the integrals are finite.

- (c) If $\int_{\Omega} h d\mu$ exists, then $|\int_{\Omega} h d\mu| \leq \int_{\Omega} |h| d\mu$.
- (d) If $h \geq 0$ and $B \in \mathcal{F}$, then $\int_B h d\mu = \sup\{\int_B s d\mu : 0 \leq s \leq h, s \text{ simple}\}$.
- (e) If $\int_{\Omega} h d\mu$ exists, so does $\int_A h d\mu$ for each $A \in \mathcal{F}$; if $\int_{\Omega} h d\mu$ is finite, then $\int_A h d\mu$ is also finite for each $A \in \mathcal{F}$.

PROOF. (a) It is immediate that this holds when h is simple. If h is non-negative and $c > 0$, then

$$\begin{aligned} \int_{\Omega} ch d\mu &= \sup\left\{\int_{\Omega} s d\mu; 0 \leq s \leq ch, s \text{ simple}\right\} \\ &= c \sup\left\{\int_{\Omega} \frac{s}{c} d\mu; 0 \leq \frac{s}{c} \leq h, \frac{s}{c} \text{ simple}\right\} = c \int_{\Omega} h d\mu. \end{aligned}$$

In general, if $h = h^+ - h^-$ and $c > 0$, then $(ch)^+ = ch^+$, $(ch)^- = ch^-$; hence $\int_{\Omega} ch d\mu = c \int_{\Omega} h^+ d\mu - c \int_{\Omega} h^- d\mu$ by what we have just proved, so that $\int_{\Omega} ch d\mu = c \int_{\Omega} h d\mu$. If $c < 0$, then

$$(ch)^+ = -ch^-, \quad (ch)^- = -ch^+,$$

so

$$\int_{\Omega} ch d\mu = -c \int_{\Omega} h^- d\mu + c \int_{\Omega} h^+ d\mu = c \int_{\Omega} h d\mu.$$

(b) If g and h are nonnegative and $0 \leq s \leq h$, s simple, then $0 \leq s \leq g$; hence $\int_{\Omega} h d\mu \leq \int_{\Omega} g d\mu$. In general, $h \leq g$ implies $h^+ \leq g^+$, $h^- \geq g^-$. If $\int_{\Omega} h d\mu > -\infty$, we have $\int_{\Omega} g^- d\mu \leq \int_{\Omega} h^- d\mu < \infty$; hence $\int_{\Omega} g d\mu$ exists and equals

$$\int_{\Omega} g^+ d\mu - \int_{\Omega} g^- d\mu \geq \int_{\Omega} h^+ d\mu - \int_{\Omega} h^- d\mu = \int_{\Omega} h d\mu.$$

The case in which $\int_{\Omega} g d\mu < \infty$ is handled similarly.

(c) We have $-|h| \leq h \leq |h|$ so by (a) and (b), $-\int_{\Omega} |h| d\mu \leq \int_{\Omega} h d\mu \leq \int_{\Omega} |h| d\mu$ and the result follows. (Note that $|h|$ is Borel measurable by 1.5.6 since $h = h^+ + h^-$.)

(d) If $0 \leq s \leq h$, then $\int_B s d\mu \leq \int_B h d\mu$ by (b); hence

$$\int_B h d\mu \geq \sup\left\{\int_B s d\mu : 0 \leq s \leq h\right\}.$$

If $0 \leq t \leq hI_B$, t simple, then $t = tI_B \leq h$ so $\int_{\Omega} t d\mu \leq \sup\{\int_{\Omega} sI_B d\mu : 0 \leq s \leq h, s \text{ simple}\}$. Take the sup over t to obtain $\int_B h d\mu \leq \sup\{\int_B s d\mu : 0 \leq s \leq h, s \text{ simple}\}$.

(e) This follows from (b) and the fact that $(hI_A)^+ = h^+I_A \leq h^+$, $(hI_A)^- = h^-I_A \leq h^-$. ■

Problems

1. Show that the functions proposed in the proof of 1.5.5(a) have the desired properties. Show also that if h is bounded, the approximating sequence converges to h uniformly on Ω .
2. Let f and g be extended real-valued Borel measurable functions on (Ω, \mathcal{F}) , and define

$$\begin{aligned} h(\omega) &= f(\omega) && \text{if } \omega \in A, \\ &= g(\omega) && \text{if } \omega \in A^c, \end{aligned}$$

where A is a set in \mathcal{F} . Show that h is Borel measurable.

3. If f_1, f_2, \dots are extended real-valued Borel measurable functions on (Ω, \mathcal{F}) , $n = 1, 2, \dots$, show that $\sup_n f_n$ and $\inf_n f_n$ are Borel measurable (hence $\limsup_{n \rightarrow \infty} f_n$ and $\liminf_{n \rightarrow \infty} f_n$ are Borel measurable).
4. Let $(\Omega, \mathcal{F}, \mu)$ be a complete measure space. If $f: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ and $g: \Omega \rightarrow \Omega'$, $g = f$ except on a subset of a set $A \in \mathcal{F}$ with $\mu(A) = 0$, show that g is measurable (relative to \mathcal{F} and \mathcal{F}').
5. (a) Let f be a function from R^k to R^m , not necessarily Borel measurable. Show that $\{x: f \text{ is discontinuous at } x\}$ is an F_σ (a countable union of closed subsets of R^k), and hence is a Borel set. Does this result hold in spaces more general than the Euclidean space R^n ?
(b) Show that there is no function from R to R whose discontinuity set is the irrationals. (In 1.4.5 we constructed a distribution function whose discontinuity set was the rationals.)
6. How many Borel measurable functions are there from R^n to R^k ?
7. We have seen that a pointwise limit of measurable functions is measurable. We may also show that under certain conditions, a pointwise limit of measures is a measure. The following result, known as Steinhaus' lemma, will be needed in the problem: If $\{a_{nk}\}$ is a double sequence of real numbers satisfying

- (i) $\sum_{k=1}^{\infty} a_{nk} = 1$ for all n ,
- (ii) $\sum_{k=1}^{\infty} |a_{nk}| \leq c < \infty$ for all n , and
- (iii) $a_{nk} \rightarrow 0$ as $n \rightarrow \infty$ for all k ,

there is a sequence $\{x_n\}$, with $x_n = 0$ or 1 for all n , such that $t_n = \sum_{k=1}^{\infty} a_{nk} x_k$ fails to converge to a finite or infinite limit.

To prove this, choose positive integers n_1 and k_1 arbitrarily; having chosen $n_1, \dots, n_r, k_1, \dots, k_r$, choose $n_{r+1} > n_r$ such that $\sum_{k \leq k_r} |a_{n_{r+1}k}| < \frac{1}{8}$; this is possible by (iii). Then choose $k_{r+1} > k_r$ such that $\sum_{k > k_{r+1}} |a_{n_{r+1}k}| < \frac{1}{8}$; this is possible by (ii). Set $x_k = 0$, $k_{2s-1} < k \leq k_{2s}$, $x_k = 1$, $k_{2s} < k \leq k_{2s+1}$, $s = 1, 2, \dots$. We may write $t_{n_{r+1}}$ as $h_1 + h_2 + h_3$, where h_1 is the sum of $a_{n_{r+1}k}x_k$ for $k \leq k_r$, h_2 corresponds to $k_r < k \leq k_{r+1}$, and h_3 to $k > k_{r+1}$. If r is odd, then $x_k = 0$, $k_r < k \leq k_{r+1}$; hence $|t_{n_{r+1}}| < \frac{1}{4}$. If r is even, then $h_2 = \sum_{k_r < k \leq k_{r+1}} a_{n_{r+1}k}$; hence by (i),

$$h_2 = 1 - \sum_{k \leq k_r} a_{n_{r+1}k} - \sum_{k > k_{r+1}} a_{n_{r+1}k} > \frac{3}{4}.$$

Thus $t_{n_{r+1}} > \frac{3}{4} - |h_1| - |h_3| > \frac{1}{2}$, so $\{t_n\}$ cannot converge.

- (a) *Vitali–Hahn–Saks Theorem.* Let (Ω, \mathcal{F}) be a measurable space, and let P_n , $n = 1, 2, \dots$, be probability measures on \mathcal{F} . If $P_n(A) \rightarrow P(A)$ for all $A \in \mathcal{F}$, then P is a probability measure on \mathcal{F} ; furthermore, if $\{B_k\}$ is a sequence of sets in \mathcal{F} decreasing to \emptyset , then $\sup_n P_n(B_k) \downarrow 0$ as $k \rightarrow \infty$. [Let A be the disjoint union of sets $A_k \in \mathcal{F}$; without loss of generality, assume $A = \Omega$ (otherwise add A^c to both sides). It is immediate that P is finitely additive, so by Problem 5, Section 1.2, $\alpha = \sum_k P(A_k) \leq P(\Omega) = 1$. If $\alpha < 1$, set $a_{nk} = (1 - \alpha)^{-1}[P_n(A_k) - P(A_k)]$ and apply Steinhaus' lemma.]
- (b) Extend the Vitali–Hahn–Saks theorem to the case where the P_n are not necessarily probability measures, but $P_n(\Omega) \leq c < \infty$ for all n . [For further extensions, see Dunford and Schwartz (1958).]

1.6 Basic Integration Theorems

We are now ready to present the main properties of the integral. The results in this section will be used many times in the text. As above, $(\Omega, \mathcal{F}, \mu)$ is a fixed measure space, and all functions to be considered map Ω to $\bar{\mathbb{R}}$.

1.6.1 Theorem. Let h be a Borel measurable function such that $\int_{\Omega} h d\mu$ exists. Define $\lambda(B) = \int_B h d\mu$, $B \in \mathcal{F}$. Then λ is countably additive on \mathcal{F} ; thus if $h \geq 0$, λ is a measure.

PROOF. Let h be a nonnegative simple function $\sum_{i=1}^n x_i I_{A_i}$. Then $\lambda(B) = \int_B h d\mu = \sum_{i=1}^n x_i \mu(B \cap A_i)$; since μ is countably additive, so is λ .

Now let h be nonnegative Borel measurable, and let $B = \bigcup_{n=1}^{\infty} B_n$, the B_n disjoint sets in \mathcal{F} . If s is simple and $0 \leq s \leq h$, then

$$\int_B s \, d\mu = \sum_{n=1}^{\infty} \int_{B_n} s \, d\mu$$

by what we have proved for nonnegative simple functions

$$\leq \sum_{n=1}^{\infty} \int_{B_n} h \, d\mu$$

by 1.5.9(b) (or the definition of the integral).

Take the sup over s to obtain, by 1.5.9(d), $\lambda(B) \leq \sum_{n=1}^{\infty} \lambda(B_n)$.

Now $B_n \subset B$, hence $I_{B_n} \leq I_B$, so by 1.5.9(b), $\lambda(B_n) \leq \lambda(B)$. If $\lambda(B_n) = \infty$ for some n , we are finished, so assume all $\lambda(B_n)$ finite. Fix n and let $\varepsilon > 0$. It follows from 1.5.9(d), 1.5.9(b), and the fact that the maximum of a finite number of simple functions is simple that we can find a simple function s , $0 \leq s \leq h$, such that

$$\int_{B_i} s \, d\mu \geq \int_{B_i} h \, d\mu - \frac{\varepsilon}{n}, \quad i = 1, 2, \dots, n.$$

Now

$$\lambda(B_1 \cup \dots \cup B_n) = \int_{\bigcup_{i=1}^n B_i} h \, d\mu \geq \int_{\bigcup_{i=1}^n B_i} s \, d\mu = \sum_{i=1}^n \int_{B_i} s \, d\mu$$

by what we have proved for nonnegative simple functions, hence

$$\lambda(B_1 \cup \dots \cup B_n) \geq \sum_{i=1}^n \int_{B_i} h \, d\mu - \varepsilon = \sum_{i=1}^n \lambda(B_i) - \varepsilon.$$

Since $\lambda(B) \geq \lambda(\bigcup_{i=1}^n B_i)$ and ε is arbitrary, we have

$$\lambda(B) \geq \sum_{i=1}^{\infty} \lambda(B_i).$$

Finally let $h = h^+ - h^-$ be an arbitrary Borel measurable function. Then $\lambda(B) = \int_B h^+ \, d\mu - \int_B h^- \, d\mu$. Since $\int_{\Omega} h^+ \, d\mu < \infty$ or $\int_{\Omega} h^- \, d\mu < \infty$, the result follows. ■

The proof of 1.6.1 shows that λ is the difference of two measures λ^+ and λ^- , where $\lambda^+(B) = \int_B h^+ \, d\mu$, $\lambda^-(B) = \int_B h^- \, d\mu$; at least one of the measures λ^+ and λ^- must be finite.

1.6.2 Monotone Convergence Theorem. Let h_1, h_2, \dots form an increasing sequence of nonnegative Borel measurable functions, and let $h(\omega) = \lim_{n \rightarrow \infty} h_n(\omega)$, $\omega \in \Omega$. Then $\int_{\Omega} h_n \, d\mu \rightarrow \int_{\Omega} h \, d\mu$. (Note that $\int_{\Omega} h_n \, d\mu$ increases with n by 1.5.9(b); for short, $0 \leq h_n \uparrow h$ implies $\int_{\Omega} h_n \, d\mu \uparrow \int_{\Omega} h \, d\mu$.)

PROOF. By 1.5.9(b), $\int_{\Omega} h_n d\mu \leq \int_{\Omega} h d\mu$ for all n , hence $k = \lim_{n \rightarrow \infty} \int_{\Omega} h_n d\mu \leq \int_{\Omega} h d\mu$. Let $0 < b < 1$, and let s be a nonnegative, finite-valued, simple function with $s \leq h$. Let $B_n = \{\omega : h_n(\omega) \geq bs(\omega)\}$. Then $B_n \uparrow \Omega$ since $h_n \uparrow h$ and s is finite-valued. Now $k \geq \int_{\Omega} h_n d\mu \geq \int_{B_n} h_n d\mu$ by 1.5.9(b), and $\int_{B_n} h_n d\mu \geq b \int_{B_n} s d\mu$ by 1.5.9(a) and (b). By 1.6.1 and 1.2.7, $\int_{B_n} s d\mu \rightarrow \int_{\Omega} s d\mu$, hence (let $b \rightarrow 1$) $k \geq \int_{\Omega} s d\mu$. Take the sup over s to obtain $k \geq \int_{\Omega} h d\mu$. ■

1.6.3 Additivity Theorem. Let f and g be Borel measurable, and assume that $f + g$ is well-defined. If $\int_{\Omega} f d\mu$ and $\int_{\Omega} g d\mu$ exist and $\int_{\Omega} f d\mu + \int_{\Omega} g d\mu$ is well-defined (not of the form $+\infty - \infty$ or $-\infty + \infty$), then

$$\int_{\Omega} (f + g) d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu.$$

In particular, if f and g are integrable, so is $f + g$.

PROOF. If f and g are nonnegative simple functions, this is immediate from the definition of the integral. Assume f and g are nonnegative Borel measurable, and let t_n, u_n be nonnegative simple functions increasing to f and g , respectively. Then $0 \leq s_n = t_n + u_n \uparrow f + g$. Now $\int_{\Omega} s_n d\mu = \int_{\Omega} t_n d\mu + \int_{\Omega} u_n d\mu$ by what we have proved for nonnegative simple functions; hence by 1.6.2, $\int_{\Omega} (f + g) d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu$.

Now if $f \geq 0, g \leq 0, h = f + g \geq 0$ (so g must be finite), we have $f = h + (-g)$; hence $\int_{\Omega} f d\mu = \int_{\Omega} h d\mu - \int_{\Omega} g d\mu$. If $\int_{\Omega} g d\mu$ is finite, then $\int_{\Omega} h d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu$, and if $\int_{\Omega} g d\mu = -\infty$, then since $h \geq 0$,

$$\int_{\Omega} f d\mu \geq - \int_{\Omega} g d\mu = \infty,$$

contradicting the hypothesis that $\int_{\Omega} f d\mu + \int_{\Omega} g d\mu$ is well-defined. Similarly if $f \geq 0, g \leq 0, h \leq 0$, we obtain $\int_{\Omega} h d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu$ by replacing all functions by their negatives. (Explicitly, $-g \geq 0, -f \leq 0, -h = -f - g \geq 0$, and the above argument applies.)

Let

$$\begin{aligned} E_1 &= \{\omega : f(\omega) \geq 0, & g(\omega) \geq 0\}, \\ E_2 &= \{\omega : f(\omega) \geq 0, & g(\omega) < 0, \quad h(\omega) \geq 0\}, \\ E_3 &= \{\omega : f(\omega) \geq 0, & g(\omega) < 0, \quad h(\omega) < 0\}, \\ E_4 &= \{\omega : f(\omega) < 0, & g(\omega) \geq 0, \quad h(\omega) \geq 0\}, \\ E_5 &= \{\omega : f(\omega) < 0, & g(\omega) \geq 0, \quad h(\omega) < 0\}, \\ E_6 &= \{\omega : f(\omega) < 0, & g(\omega) < 0\}. \end{aligned}$$

The above argument shows that $\int_{E_i} h \, d\mu = \int_{E_i} f \, d\mu + \int_{E_i} g \, d\mu$. Now $\int_{\Omega} f \, d\mu = \sum_{i=1}^6 \int_{E_i} f \, d\mu$, $\int_{\Omega} g \, d\mu = \sum_{i=1}^6 \int_{E_i} g \, d\mu$ by 1.6.1, so that $\int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu = \sum_{i=1}^6 \int_{E_i} h \, d\mu$, and this equals $\int_{\Omega} h \, d\mu$ by 1.6.1, if we can show that $\int_{\Omega} h \, d\mu$ exists; that is, $\int_{\Omega} h^+ \, d\mu$ and $\int_{\Omega} h^- \, d\mu$ are not both infinite.

If this is the case, $\int_{E_i} h^+ \, d\mu = \int_{E_j} h^- \, d\mu = \infty$ for some i, j (1.6.1 again), so that $\int_{E_i} h \, d\mu = \infty$, $\int_{E_j} h \, d\mu = -\infty$. But then $\int_{E_i} f \, d\mu$ or $\int_{E_i} g \, d\mu = \infty$; hence $\int_{\Omega} f \, d\mu$ or $\int_{\Omega} g \, d\mu = \infty$. (Note that $\int_{\Omega} f^+ \, d\mu \geq \int_{E_i} f^+ \, d\mu$.) Similarly $\int_{\Omega} f \, d\mu$ or $\int_{\Omega} g \, d\mu = -\infty$, and this is a contradiction. ■

1.6.4 Corollaries. (a) If h_1, h_2, \dots are nonnegative Borel measurable,

$$\int_{\Omega} \left(\sum_{n=1}^{\infty} h_n \right) \, d\mu = \sum_{n=1}^{\infty} \int_{\Omega} h_n \, d\mu.$$

Thus any series of nonnegative Borel measurable functions may be integrated term by term.

(b) If h is Borel measurable, h is integrable iff $|h|$ is integrable.

(c) If g and h are Borel measurable with $|g| \leq h$, h integrable, then g is integrable.

PROOF. (a) $\sum_{k=1}^n h_k \uparrow \sum_{k=1}^{\infty} h_k$, and the result follows from 1.6.2 and 1.6.3.

(b) Since $|h| = h^+ + h^-$, this follows from the definition of the integral and 1.6.3.

(c) By 1.5.9(b), $|g|$ is integrable, and the result follows from (b) above. ■

A condition is said to hold *almost everywhere* with respect to the measure μ (written a.e. $[\mu]$ or simply a.e. if μ is understood) iff there is a set $B \in \mathcal{F}$ of μ -measure 0 such that the condition holds outside of B . From the point of view of integration theory, functions that differ only on a set of measure 0 may be identified. This is established by the following result.

1.6.5 Theorem. Let f, g , and h be Borel measurable functions.

(a) If $f = 0$ a.e. $[\mu]$, then $\int_{\Omega} f \, d\mu = 0$.

(b) If $g = h$ a.e. $[\mu]$ and $\int_{\Omega} g \, d\mu$ exists, then so does $\int_{\Omega} h \, d\mu$, and $\int_{\Omega} g \, d\mu = \int_{\Omega} h \, d\mu$.

PROOF. (a) If $f = \sum_{i=1}^n x_i I_{A_i}$ is simple, then $x_i \neq 0$ implies $\mu(A_i) = 0$ by hypothesis, hence $\int_{\Omega} f \, d\mu = 0$. If $f \geq 0$ and $0 \leq s \leq f$, s simple, then $s = 0$ a.e. $[\mu]$, hence $\int_{\Omega} s \, d\mu = 0$; thus $\int_{\Omega} f \, d\mu = 0$. If $f = f^+ - f^-$, then f^+ and f^- , being less than or equal to $|f|$, are 0 a.e. $[\mu]$, and the result follows.

(b) Let $A = \{\omega: g(\omega) = h(\omega)\}$, $B = A^c$. Then $g = gI_A + gI_B$, $h = hI_A + hI_B = gI_A + hI_B$. Since $gI_B = hI_B = 0$ except on B , a set of measure 0, the result follows from part (a) and 1.6.3. ■

Thus in any integration theorem, we may freely use the phrase “almost everywhere.” For example, if $\{h_n\}$ is an increasing sequence of nonnegative Borel measurable functions converging a.e. to the Borel measurable function h , then $\int_{\Omega} h_n d\mu \rightarrow \int_{\Omega} h d\mu$.

Another example: If g and h are Borel measurable and $g \geq h$ a.e., then $\int_{\Omega} g d\mu \geq \int_{\Omega} h d\mu$ [in the sense of 1.5.9(b)].

1.6.6 Theorem.

Let h be Borel measurable.

- (a) If h is integrable, then h is finite a.e.
- (b) If $h \geq 0$ and $\int_{\Omega} h d\mu = 0$, then $h = 0$ a.e.

PROOF. (a) Let $A = \{\omega: |h(\omega)| = \infty\}$. If $\mu(A) > 0$, then $\int_{\Omega} |h| d\mu \geq \int_A |h| d\mu = \infty \mu(A) = \infty$, a contradiction.

(b) Let $B = \{\omega: h(\omega) > 0\}$, $B_n = \{\omega: h(\omega) \geq 1/n\} \uparrow B$. We have $0 \leq hI_{B_n} \leq hI_B = h$; hence by 1.5.9(b), $\int_{B_n} h d\mu = 0$. But $\int_{B_n} h d\mu \geq (1/n)\mu(B_n)$, so that $\mu(B_n) = 0$ for all n , and thus $\mu(B) = 0$. ■

The monotone convergence theorem was proved under the hypothesis that all functions were nonnegative. This assumption can be relaxed considerably, as we now prove.

1.6.7 Extended Monotone Convergence Theorem.

Let g_1, g_2, \dots, g, h be Borel measurable.

- (a) If $g_n \geq h$ for all n , where $\int_{\Omega} h d\mu > -\infty$, and $g_n \uparrow g$, then

$$\int_{\Omega} g_n d\mu \uparrow \int_{\Omega} g d\mu.$$

- (b) If $g_n \leq h$ for all n , where $\int_{\Omega} h d\mu < \infty$, and $g_n \downarrow g$, then

$$\int_{\Omega} g_n d\mu \downarrow \int_{\Omega} h d\mu.$$

PROOF. (a) If $\int_{\Omega} h d\mu = \infty$, then by 1.5.9(b), $\int_{\Omega} g_n d\mu = \infty$ for all n , and $\int_{\Omega} g d\mu = \infty$. Thus assume $\int_{\Omega} h d\mu < \infty$, so that by 1.6.6(a), h is a.e. finite; change h to 0 on the set where it is infinite. Then $0 \leq g_n - h \uparrow g - h$ a.e., hence by 1.6.2, $\int_{\Omega} (g_n - h) d\mu \uparrow \int_{\Omega} (g - h) d\mu$. The result follows from 1.6.3.

(We must check that the additivity theorem actually applies. Since $\int_{\Omega} h \, d\mu > -\infty$, $\int_{\Omega} g_n \, d\mu$ and $\int_{\Omega} g \, d\mu$ exist and are greater than $-\infty$ by 1.5.9(b). Also, $\int_{\Omega} h \, d\mu$ is finite, so that $\int_{\Omega} g_n \, d\mu - \int_{\Omega} h \, d\mu$ and $\int_{\Omega} g \, d\mu - \int_{\Omega} h \, d\mu$ are well-defined.)

(b) $-g_n \geq -h$, $\int_{\Omega} -h \, d\mu > -\infty$, and $-g_n \uparrow -g$. By part (a), $-\int_{\Omega} g_n \, d\mu \uparrow -\int_{\Omega} g \, d\mu$, so $\int_{\Omega} g_n \, d\mu \downarrow \int_{\Omega} g \, d\mu$. ■

The extended monotone convergence theorem asserts that under appropriate conditions, the limit of the integrals of a sequence of functions is the integral of the limit function. More general theorems of this type can be obtained if we replace limits by upper or lower limits. If f_1, f_2, \dots are functions from Ω to \bar{R} , $\liminf_{n \rightarrow \infty} f_n$ and $\limsup_{n \rightarrow \infty} f_n$ are defined pointwise, that is,

$$\begin{aligned}\left(\liminf_{n \rightarrow \infty} f_n\right)(\omega) &= \sup_n \inf_{k \geq n} f_k(\omega), \\ \left(\limsup_{n \rightarrow \infty} f_n\right)(\omega) &= \inf_n \sup_{k \geq n} f_k(\omega).\end{aligned}$$

1.6.8 Fatou's Lemma. Let f_1, f_2, \dots, f be Borel measurable.

(a) If $f_n \geq f$ for all n , where $\int_{\Omega} f \, d\mu > -\infty$, then

$$\liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \geq \int_{\Omega} \left(\liminf_{n \rightarrow \infty} f_n \right) \, d\mu.$$

(b) If $f_n \leq f$ for all n , where $\int_{\Omega} f \, d\mu < \infty$, then

$$\limsup_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} \left(\limsup_{n \rightarrow \infty} f_n \right) \, d\mu.$$

PROOF. (a) Let $g_n = \inf_{k \geq n} f_k$, $g = \liminf f_n$. Then $g_n \geq f$ for all n , $\int_{\Omega} f \, d\mu > -\infty$, and $g_n \uparrow g$. By 1.6.7, $\int_{\Omega} g_n \, d\mu \uparrow \int_{\Omega} (\liminf_{n \rightarrow \infty} f_n) \, d\mu$. But $g_n \leq f_n$, so

$$\lim_{n \rightarrow \infty} \int_{\Omega} g_n \, d\mu = \liminf_{n \rightarrow \infty} \int_{\Omega} g_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

(b) We may write

$$\begin{aligned}\int_{\Omega} \left(\limsup_{n \rightarrow \infty} f_n \right) \, d\mu &= - \int_{\Omega} \liminf_{n \rightarrow \infty} (-f_n) \, d\mu \\ &\geq - \liminf_{n \rightarrow \infty} \int_{\Omega} (-f_n) \, d\mu \quad \text{by (a)} \\ &= \limsup_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu. \blacksquare\end{aligned}$$

The following result is one of the “bread and butter” theorems of analysis; it will be used quite often in later chapters.

1.6.9 Dominated Convergence Theorem. If f_1, f_2, \dots, f, g are Borel measurable, $|f_n| \leq g$ for all n , where g is μ -integrable, and $f_n \rightarrow f$ a.e. [μ], then f is μ -integrable and $\int_{\Omega} f_n d\mu \rightarrow \int_{\Omega} f d\mu$.

PROOF. We have $|f| \leq g$ a.e.; hence f is integrable by 1.6.4(c). By 1.6.8,

$$\begin{aligned} \int_{\Omega} \left(\liminf_{n \rightarrow \infty} f_n \right) d\mu &\leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \leq \limsup_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \\ &\leq \int_{\Omega} \left(\limsup_{n \rightarrow \infty} f_n \right) d\mu. \end{aligned}$$

By hypothesis, $\liminf_{n \rightarrow \infty} f_n = \limsup_{n \rightarrow \infty} f_n = f$ a.e., so all terms of the above inequality are equal to $\int_{\Omega} f d\mu$. ■

1.6.10 Corollary. If f_1, f_2, \dots, f, g are Borel measurable, $|f_n| \leq g$ for all n , where $|g|^p$ is μ -integrable ($p > 0$, fixed), and $f_n \rightarrow f$ a.e. [μ], then $|f|^p$ is μ -integrable and $\int_{\Omega} |f_n - f|^p d\mu \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. We have $|f_n|^p \leq |g|^p$ for all n ; so $|f|^p \leq |g|^p$, and therefore $|f|^p$ is integrable. Also $|f_n - f|^p \leq (|f_n| + |f|)^p \leq (2|g|)^p$, which is integrable, and the result follows from 1.6.9. ■

We have seen in 1.5.9(b) that $g \leq h$ implies $\int_{\Omega} g d\mu \leq \int_{\Omega} h d\mu$, and in fact $\int_A g d\mu \leq \int_A h d\mu$ for all $A \in \mathcal{F}$. There is a converse to this result.

1.6.11 Theorem. If μ is σ -finite on \mathcal{F} , g and h are Borel measurable, $\int_{\Omega} g d\mu$ and $\int_{\Omega} h d\mu$ exist, and $\int_A g d\mu \leq \int_A h d\mu$ for all $A \in \mathcal{F}$, then $g \leq h$ a.e. [μ].

PROOF. It is sufficient to prove this when μ is finite. Let

$$A_n = \left\{ \omega : g(\omega) \geq h(\omega) + \frac{1}{n}, \quad |h(\omega)| \leq n \right\}.$$

Then

$$\int_{A_n} h d\mu \geq \int_{A_n} g d\mu \geq \int_{A_n} h d\mu + \frac{1}{n} \mu(A_n).$$

But

$$\left| \int_{A_n} h d\mu \right| \leq \int_{A_n} |h| d\mu \leq n \mu(A_n) < \infty,$$

and thus we may subtract $\int_{A_n} h \, d\mu$ to obtain $(1/n)\mu(A_n) \leq 0$, hence $\mu(A_n) = 0$. Therefore $\mu(\bigcup_{n=1}^{\infty} A_n) = 0$; hence $\mu\{\omega: g(\omega) > h(\omega), h(\omega) \text{ finite}\} = 0$. Consequently $g \leq h$ a.e. on $\{\omega: h(\omega) \text{ finite}\}$. Clearly, $g \leq h$ everywhere on $\{\omega: h(\omega) = \infty\}$, and by taking $C_n = \{\omega: h(\omega) = -\infty, g(\omega) \geq -n\}$ we obtain

$$-\infty\mu(C_n) = \int_{C_n} h \, d\mu \geq \int_{C_n} g \, d\mu \geq -n\mu(C_n);$$

hence $\mu(C_n) = 0$. Thus $\mu(\bigcup_{n=1}^{\infty} C_n) = 0$, so that

$$\mu\{\omega: g(\omega) > h(\omega), h(\omega) = -\infty\} = 0.$$

Therefore $g \leq h$ a.e. on $\{\omega: h(\omega) = -\infty\}$. ■

If g and h are integrable, the proof is simpler. Let $B = \{\omega: g(\omega) > h(\omega)\}$. Then $\int_B g \, d\mu \leq \int_B h \, d\mu \leq \int_B g \, d\mu$; hence all three integrals are equal. Thus by 1.6.3, $0 = \int_B (g - h) \, d\mu = \int_{\Omega} (g - h)I_B \, d\mu$, with $(g - h)I_B \geq 0$. By 1.6.6(b), $(g - h)I_B = 0$ a.e., so that $g = h$ a.e. on B . But $g \leq h$ on B^c , and the result follows. Note that in this case, μ need not be σ -finite.

The reader may have noticed that several integration theorems in this section were proved by starting with nonnegative simple functions and working up to nonnegative measurable functions and finally to arbitrary measurable functions. This technique is quite basic and will often be useful. A good illustration of the method is the following result, which introduces the notion of a measure-preserving transformation, a key concept in ergodic theory. In fact it is convenient here to start with indicators before proceeding to nonnegative simple functions.

1.6.12 Theorem. Let $T: (\Omega, \mathcal{F}) \rightarrow (\Omega_0, \mathcal{F}_0)$ be a measurable mapping, and let μ be a measure on \mathcal{F} . Define a measure $\mu_0 = \mu T^{-1}$ on \mathcal{F}_0 by

$$\mu_0(A) = \mu(T^{-1}(A)), \quad A \in \mathcal{F}_0.$$

If $\Omega_0 = \Omega$, $\mathcal{F}_0 = \mathcal{F}$, and $\mu_0 = \mu$, T is said to *preserve* the measure μ .

If $f: (\Omega_0, \mathcal{F}_0) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ and $A \in \mathcal{F}_0$, then

$$\int_{T^{-1}A} f(T(\omega)) \, d\mu(\omega) = \int_A f(\omega) \, d\mu_0(\omega)$$

in the sense that if one of the integrals exists, so does the other, and the two integrals are equal.

PROOF. If f is an indicator I_B , the desired formula states that

$$\mu(T^{-1}A \cap T^{-1}B) = \mu_0(A \cap B),$$

which is true by definition of μ_0 . If f is a nonnegative simple function $\sum_{i=1}^n x_i I_{B_i}$, then

$$\begin{aligned}\int_{T^{-1}A} f(T(\omega)) d\mu(\omega) &= \sum_{i=1}^n x_i \int_{T^{-1}A} I_{B_i}(T(\omega)) d\mu(\omega) \quad \text{by 1.6.3} \\ &= \sum_{i=1}^n x_i \int_A I_{B_i}(\omega) d\mu_0(\omega) \\ &\quad \text{by what we have proved for indicators} \\ &= \int_A f(\omega) d\mu_0(\omega) \quad \text{by 1.6.3.}\end{aligned}$$

If f is a nonnegative Borel measurable function, let f_1, f_2, \dots be nonnegative simple functions increasing to f . Then $\int_{T^{-1}A} f_n(T(\omega)) d\mu(\omega) = \int_A f_n(\omega) d\mu_0(\omega)$ by what we have proved for simple functions, and the monotone convergence theorem yields the desired result for f .

Finally, if $f = f^+ - f^-$ is an arbitrary Borel measurable function, we have proved that the result holds for f^+ and f^- . If, say, $\int_A f^+(\omega) d\mu_0(\omega) < \infty$, then $\int_{T^{-1}A} f^+(T(\omega)) d\mu(\omega) < \infty$, and it follows that if one of the integrals exists, so does the other, and the two integrals are equal. ■

If one is having difficulty proving a theorem about measurable functions or integration, it is often helpful to start with indicators and work upward. In fact it is possible to suspect that almost anything can be proved this way, but of course there are exceptions. For example, you will run into trouble trying to prove the proposition “All functions are indicators.”

We shall adopt the following terminology: If μ is Lebesgue measure and A is an interval $[a, b]$, $\int_A f d\mu$, if it exists, will often be denoted by $\int_b^a f(x) dx$ (or $\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \cdots dx_n$ if we are integrating functions on R^n). The endpoints may be deleted from the interval without changing the integral, since the Lebesgue measure of a single point is 0. If f is integrable with respect to μ , then we say that f is *Lebesgue integrable*. A different notation, such as $r_{ab}(f)$, will be used for the Riemann integral of f on $[a, b]$.

Problems

The first three problems give conditions under which some of the most commonly occurring operations in real analysis may be performed: taking a limit under the integral sign, integrating an infinite series term by term, and differentiating under the integral sign.

1. Let $f = f(x, y)$ be a real-valued function of two real variables, defined for $a < y < b$, $c < x < d$. Assume that for each x , $f(x, \cdot)$ is a Borel measurable function of y , and that there is a Borel measurable $g: (a, b) \rightarrow \mathbb{R}$ such that $|f(x, y)| \leq g(y)$ for all x, y , and $\int_a^b g(y) dy < \infty$. If $x_0 \in (c, d)$ and $\lim_{x \rightarrow x_0} f(x, y)$ exists for all $y \in (a, b)$, show that

$$\lim_{x \rightarrow x_0} \int_a^b f(x, y) dy = \int_a^b \left[\lim_{x \rightarrow x_0} f(x, y) \right] dy.$$

2. Let f_1, f_2, \dots be Borel measurable functions on $(\Omega, \mathcal{F}, \mu)$. If

$$\sum_{n=1}^{\infty} \int_{\Omega} |f_n| d\mu < \infty,$$

show that $\sum_{n=1}^{\infty} f_n$ converges a.e. [μ] to a finite-valued function, and $\int_{\Omega} (\sum_{n=1}^{\infty} f_n) d\mu = \sum_{n=1}^{\infty} \int_{\Omega} f_n d\mu$.

3. Let $f = f(x, y)$ be a real-valued function of two real variables, defined for $a < y < b$, $c < x < d$, such that f is a Borel measurable function of y for each fixed x . Assume that for each x , $f(x, \cdot)$ is integrable over (a, b) (with respect to Lebesgue measure). Suppose that the partial derivative $f_1(x, y)$ of f with respect to x exists for all (x, y) , and suppose there is a Borel measurable $h: (a, b) \rightarrow \mathbb{R}$ such that $|f_1(x, y)| \leq h(y)$ for all x, y , where $\int_a^b h(y) dy < \infty$.

Show that $d[\int_a^b f(x, y) dy]/dx$ exists for all $x \in (c, d)$, and equals $\int_a^b f_1(x, y) dy$. (It must be verified that $f_1(x, \cdot)$ is Borel measurable for each x .)

4. If μ is a measure on (Ω, \mathcal{F}) and A_1, A_2, \dots is a sequence of sets in \mathcal{F} , use Fatou's lemma to show that

$$\mu\left(\liminf_n A_n\right) \leq \liminf_{n \rightarrow \infty} \mu(A_n).$$

If μ is finite, show that

$$\mu\left(\limsup_n A_n\right) \geq \limsup_{n \rightarrow \infty} \mu(A_n).$$

Thus if μ is finite and $A = \lim_n A_n$, then $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$. (For another proof of this, see Section 1.2, Problem 10.)

5. Give an example of a sequence of Lebesgue integrable functions f_n converging everywhere to a Lebesgue integrable function f , such that

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx < \int_{-\infty}^{\infty} f(x) dx.$$

Thus the hypotheses of the dominated convergence theorem and Fatou's lemma cannot be dropped.

6. (a) Show that $\int_1^\infty e^{-t} \ln t dt = \lim_{n \rightarrow \infty} \int_1^n [1 - (t/n)]^n \ln t dt$.
 (b) Show that $\int_0^1 e^{-t} \ln t dt = \lim_{n \rightarrow \infty} \int_0^1 [1 - (t/n)]^n \ln t dt$.
7. If $(\Omega, \mathcal{F}, \mu)$ is the completion of $(\Omega, \mathcal{F}_0, \mu)$ and f is a Borel measurable function on (Ω, \mathcal{F}) , show that there is a Borel measurable function g on (Ω, \mathcal{F}_0) such that $f = g$, except on a subset of a set in \mathcal{F}_0 of measure 0. (Start with indicators.)
8. If f is a Borel measurable function from R to R and $a \in R$, show that

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} f(x-a) dx$$

in the sense that if one integral exists, so does the other, and the two are equal. (Start with indicators.)

1.7 Comparison of Lebesgue and Riemann Integrals

In this section we show that integration with respect to Lebesgue measure is more general than Riemann integration, and we obtain a precise criterion for Riemann integrability.

Let $[a, b]$ be a bounded closed interval of reals, and let f be a bounded real-valued function on $[a, b]$, fixed throughout the discussion. If $P: a = x_0 < x_1 < \dots < x_n = b$ is a partition of $[a, b]$, we may construct the upper and lower sums of f relative to P as follows.

Let

$$\begin{aligned} M_i &= \sup\{f(y): x_{i-1} < y \leq x_i\}, & i &= 1, \dots, n, \\ m_i &= \inf\{f(y): x_{i-1} < y \leq x_i\}, & i &= 1, \dots, n, \end{aligned}$$

and define step functions α and β , called the *upper* and *lower* functions corresponding to P , by

$$\begin{aligned} \alpha(x) &= M_i & \text{if } x_{i-1} < x \leq x_i, & i = 1, \dots, n, \\ \beta(x) &= m_i & \text{if } x_{i-1} < x \leq x_i, & i = 1, \dots, n \end{aligned}$$

$[\alpha(a) \text{ and } \beta(a) \text{ may be chosen arbitrarily}].$ The upper and lower sums are given by

$$U(P) = \sum_{i=1}^n M_i(x_i - x_{i-1}),$$

$$L(P) = \sum_{i=1}^n m_i(x_i - x_{i-1}).$$

Now we take as a measure space $\Omega = [a, b]$, $\mathcal{F} = \mathcal{B}[a, b]$, the *Lebesgue measurable* subsets of $[a, b]$, $\mu = \text{Lebesgue measure}$. Since α and β are simple functions, we have

$$U(P) = \int_a^b \alpha \, d\mu, \quad L(P) = \int_a^b \beta \, d\mu.$$

Now let P_1, P_2, \dots be a sequence of partitions of $[a, b]$ such that P_{k+1} is a refinement of P_k for each k , and such that $|P_k|$ (the length of the largest subinterval of P_k) approaches 0 as $k \rightarrow \infty$. If α_k and β_k are the upper and lower functions corresponding to P_k , then

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq f \geq \cdots \geq \beta_2 \geq \beta_1.$$

Thus α_k and β_k approach limit functions α and β . If $|f|$ is bounded by M , then all $|\alpha_k|$ and $|\beta_k|$ are bounded by M as well, and the function that is constant at M is integrable on $[a, b]$ with respect to μ , since

$$\mu[a, b] = b - a < \infty.$$

By the dominated convergence theorem,

$$\lim_{k \rightarrow \infty} U(P_k) = \lim_{k \rightarrow \infty} \int_a^b \alpha_k \, d\mu = \int_a^b \alpha \, d\mu,$$

and

$$\lim_{k \rightarrow \infty} L(P_k) = \lim_{k \rightarrow \infty} \int_a^b \beta_k \, d\mu = \int_a^b \beta \, d\mu.$$

We shall need one other fact, namely that if x is not an endpoint of any of the subintervals of the P_k ,

$$f \text{ is continuous at } x \quad \text{iff} \quad \alpha(x) = f(x) = \beta(x).$$

This follows by a standard ϵ - δ argument.

If $\lim_{k \rightarrow \infty} U(P_k) = \lim_{k \rightarrow \infty} L(P_k) =$ a finite number r , independent of the particular sequence of partitions, f is said to be *Riemann integrable* on $[a, b]$, and $r = r_{ab}(f)$ is said to be the (value of the) *Riemann integral* of f on $[a, b]$. The above argument shows that f is Riemann integrable iff

$$\int_a^b \alpha \, d\mu = \int_a^b \beta \, d\mu = r,$$

independent of the particular sequence of partitions. If f is Riemann integrable,

$$r_{ab}(f) = \int_a^b \alpha \, d\mu = \int_a^b \beta \, d\mu.$$

We are now ready for the main results.

1.7.1 Theorem. Let f be a bounded real-valued function on $[a, b]$.

- (a) The function f is Riemann integrable on $[a, b]$ iff f is continuous almost everywhere on $[a, b]$ (with respect to Lebesgue measure).
- (b) If f is Riemann integrable on $[a, b]$, then f is integrable with respect to Lebesgue measure on $[a, b]$, and the two integrals are equal.

PROOF. (a) If f is Riemann integrable,

$$r_{ab}(f) = \int_a^b \alpha \, d\mu = \int_a^b \beta \, d\mu.$$

Since $\beta \leq f \leq \alpha$, 1.6.6(b) applied to $\alpha - \beta$ yields $\alpha - f = \beta$ a.e.; hence f is continuous a.e. Conversely, assume f continuous a.e.; then $\alpha = f = \beta$ a.e. Now α and β are limits of simple functions, and hence are Borel measurable. Thus f differs from a measurable function on a subset of a set of measure 0, and therefore f is measurable because of the completeness of the measure space. (See Section 1.5, Problem 4.) Since f is bounded, it is integrable with respect to μ , and since $\alpha = f = \beta$ a.e., we have

$$\int_a^b \alpha \, d\mu = \int_a^b \beta \, d\mu = \int_a^b f \, d\mu, \quad (1)$$

independent of the particular sequence of partitions. Therefore f is Riemann integrable.

(b) If f is Riemann integrable, then f is continuous a.e. by part (a). But then Eq. (1) yields $r_{ab}(f) = \int_a^b f \, d\mu$, as desired. ■

Theorem 1.7.1 holds equally well in n dimensions, with $[a, b]$ replaced by a closed bounded interval of R^n ; the proof is essentially the same.

A somewhat more complicated situation arises with *improper integrals*; here the interval of integration is infinite or the function f is unbounded. Some results are given in Problem 3.

We have seen that convenient conditions exist that allow the interchange of limit operations on Lebesgue integrable functions. (For example, see Problems 1–3 of Section 1.6.) The corresponding results for Riemann integrable functions are more complicated, basically because the limit of a sequence of Riemann integrable functions need not be Riemann integrable, even if the entire sequence is uniformly bounded (see Problem 4). Thus Riemann integrability of the limit function must be added as a hypothesis, and this is a serious limitation on the scope of the results.

Problems

- The function defined on $[0, 1]$ by $f(x) = 1$ if x is irrational, and $f(x) = 0$ if x is rational, is the standard example of a function that is Lebesgue integrable (it is 1 a.e.) but not Riemann integrable. But what is wrong with the following reasoning:

If we consider the behavior of f on the irrationals, f assumes the constant value 1 and is therefore continuous. Since the rationals have Lebesgue measure 0, f is therefore continuous almost everywhere and hence is Riemann integrable.

- Let f be a bounded real-valued function on the bounded closed interval $[a, b]$. Let F be an increasing right-continuous function on $[a, b]$ with corresponding Lebesgue–Stieltjes measure μ (defined on the Borel subsets of $[a, b]$).

Define M_i , m_i , α , and β as in Section 1.7, and take

$$U(P) = \sum_{i=1}^n M_i(F(x_i) - F(x_{i-1})) = \int_a^b \alpha \, d\mu,$$

$$L(P) = \sum_{i=1}^n m_i(F(x_i) - F(x_{i-1})) = \int_a^b \beta \, d\mu,$$

where \int_a^b indicates that the integration is over $(a, b]$. If $\{P_k\}$ is a sequence of partitions with $|P_k| \rightarrow 0$ and P_{k+1} refining P_k , with α_k and β_k the upper and lower functions corresponding to P_k ,

$$\lim_{k \rightarrow \infty} U(P_k) = \int_a^b \alpha \, d\mu,$$

$$\lim_{k \rightarrow \infty} L(P_k) = \int_a^b \beta \, d\mu,$$

where $\alpha = \lim_{k \rightarrow \infty} \alpha_k$, $\beta = \lim_{k \rightarrow \infty} \beta_k$. If $U(P_k)$ and $L(P_k)$ approach the same limit $r_{ab}(f; F)$ (independent of the particular sequence of partitions), this number is called the *Riemann–Stieltjes integral* of f with respect to F on $[a, b]$, and f is said to be *Riemann–Stieltjes integrable* with respect to F on $[a, b]$.

- Show that f is Riemann–Stieltjes integrable iff f is continuous a.e. [μ] on $[a, b]$.
- Show that if f is Riemann–Stieltjes integrable, then f is integrable with respect to the completion of the measure μ , and the two integrals are equal.

3. If $f: R \rightarrow R$, the improper Riemann integral of f may be defined as

$$r(f) = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow \infty}} r_{ab}(f)$$

if the limit exists and is finite.

- (a) Show that if f has an improper Riemann integral, it is continuous a.e. [Lebesgue measure] on R , but not conversely.
 - (b) If f is nonnegative and has an improper Riemann integral, show that f is integrable with respect to the completion of Lebesgue measure, and the two integrals are equal. Give a counterexample to this result if the nonnegativity hypothesis is dropped.
4. Give an example of a sequence of functions f_n on $[a, b]$ such that each f_n is Riemann integrable, $|f_n| \leq 1$ for all n , $f_n \rightarrow f$ everywhere, but f is not Riemann integrable.

Note: References on measure and integration will be given at the end of Chapter 2.

2

Further Results in Measure and Integration Theory

2.1 Introduction

This chapter consists of a variety of applications of the basic integration theory developed in Chapter 1. Perhaps the most important result is the Radon–Nikodym theorem, which is fundamental in modern probability theory and other parts of analysis. It will be instructive to consider a special case of this result before proceeding to the general theory. Suppose that F is a distribution function on R , and assume that F has a jump of magnitude a_k at the point x_k , $k = 1, 2, \dots$. Let us subtract out the discontinuities of F ; specifically let μ_1 be a measure concentrated on $\{x_1, x_2, \dots\}$, with $\mu_1\{x_k\} = a_k$ for all k , and let F_1 be a distribution function corresponding to μ_1 . Then $G = F - F_1$ is a continuous distribution function, so that the corresponding Lebesgue–Stieltjes measure λ satisfies $\lambda\{x\} = 0$ for all x . Now in any “practical” case, we can write $G(x) = \int_{-\infty}^x f(t) dt$, $x \in R$, for some nonnegative Borel measurable function f (the way to find f is to differentiate G). It follows that $\lambda(B) = \int_B f(x) dx$ for all $B \in \mathcal{B}(R)$. To see this, observe that if $\lambda'(B) = \int_B f(x) dx$, then λ' is a measure on $\mathcal{B}(R)$ and $\lambda'([a, b]) = G(b) - G(a)$; thus $\lambda' = \lambda$.

It is natural to conjecture that if λ is a measure on $\mathcal{B}(R)$ and $\lambda\{x\} = 0$ for all x , then we can write $\lambda(B) = \int_B f(x) dx$, $B \in \mathcal{B}(R)$, for some nonnegative

Borel measurable f . However, as found by Lebesgue, the conjecture is false unless the hypothesis is strengthened. Not only must we assume that λ assigns measure 0 to singletons, but in fact we must assume that λ is *absolutely continuous* with respect to Lebesgue measure μ , that is, if $\mu(B) = 0$, then $\lambda(B) = 0$. In general, λ may be represented as the sum of two measures λ_1 and λ_2 , where λ_1 is absolutely continuous with respect to μ and λ_2 is *singular* with respect to μ , which means that λ_2 is concentrated on a set of Lebesgue measure 0. A simple example of a measure singular with respect to μ is one that is concentrated on a countable set; however, as we shall see, more complicated examples exist.

The first step in the development of the general Radon–Nikodym theorem is the Jordan–Hahn decomposition, which represents a countably additive set function as the difference of two measures.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, h a Borel measurable function such that $\int_{\Omega} h \, d\mu$ exists. If $\lambda(A) = \int_A h \, d\mu$, $A \in \mathcal{F}$, then by 1.6.1, λ is a countably additive set function on \mathcal{F} . We call λ the *indefinite integral* of h (with respect to μ). If μ is Lebesgue measure and $A = [a, x]$, then $\lambda(A) = \int_a^x h(y) \, dy$, the familiar indefinite integral of calculus. As we noted after the proof of 1.6.1, λ is the difference of two measures, at least one of which is finite. We are going to show that any countably additive set function can be represented in this way. First, a preliminary result.

2.1.1 Theorem. Let λ be a countably additive extended real-valued set function on the σ -field \mathcal{F} of subsets of Ω . Then λ assumes a maximum and a minimum value, that is, there are sets $C, D \in \mathcal{F}$ such that

$$\lambda(C) = \sup \{\lambda(A) : A \in \mathcal{F}\} \quad \text{and} \quad \lambda(D) = \inf \{\lambda(A) : A \in \mathcal{F}\}.$$

Before giving the proof, let us look at some special cases. If λ is a measure, the result is trivial: take $C = \Omega$, $D = \emptyset$. If λ is the indefinite integral of h with respect to μ , we may write

$$\lambda(A) = \int_A h \, d\mu = \int_{A \cap \{\omega : h(\omega) \geq 0\}} h \, d\mu + \int_{A \cap \{\omega : h(\omega) < 0\}} h \, d\mu.$$

Thus

$$\int_{\{\omega : h(\omega) < 0\}} h \, d\mu \leq \lambda(A) \leq \int_{\{\omega : h(\omega) \geq 0\}} h \, d\mu.$$

Therefore we may take $C = \{\omega : h(\omega) \geq 0\}$, $D = \{\omega : h(\omega) < 0\}$.

PROOF. First consider the sup. We may assume that $\lambda < \infty$, for if $\lambda(A_0) = \infty$ we take $C = A_0$. Let $A_n \in \mathcal{F}$ with $\lambda(A_n) \rightarrow \sup \lambda$, and let $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

For each n , we may partition A into 2^n disjoint subsets A_{nm} , where each A_{nm} is of the form $A_1^* \cap A_2^* \cap \cdots \cap A_n^*$, with A_j^* either A_j or $A - A_j$. For example, if $n = 3$, we have (with intersections written as products)

$$\begin{aligned} A &= A_1 A_2 A_3 \cup A_1 A_2 A_3' \cup A_1 A_2' A_3 \cup A_1 A_2' A_3' \cup A_1' A_2 A_3 \\ &\quad \cup A_1' A_2 A_3' \cup A_1' A_2' A_3 \cup A_1' A_2' A_3', \quad \text{where } A_j' = A - A_j. \end{aligned}$$

Let $B_n = \bigcup_m \{A_{nm} : \lambda(A_{nm}) \geq 0\}$; set $B_n = \emptyset$ if $\lambda(A_{nm}) < 0$ for all m . Now each A_n is a finite union of some of the A_{nm} , hence $\lambda(A_n) \leq \lambda(B_n)$ by definition of B_n . Also, if $n' > n$, each $A_{n'm'}$ is either a subset of a given A_{nm} or disjoint from it (for example, $A_1 A_2' A_3' \subset A_1 A_2'$, and $A_1 A_2' A_3'$ is disjoint from $A_1 A_2$, $A_1' A_2$, and $A_1' A_2'$). Thus $\bigcup_{k=n}^r B_k$ can be expressed as a union of B_n and sets E disjoint from B_n such that $\lambda(E) \geq 0$. [Note that, for example, if $\lambda(A_1 A_2') = \lambda(A_1 A_2' A_3) + \lambda(A_1 A_2' A_3') \geq 0$, it may happen that $\lambda(A_1 A_2' A_3) < 0$, $\lambda(A_1 A_2' A_3') > 0$, so the sequence $\{B_n\}$ need not be monotone.] Consequently,

$$\lambda(A_n) \leq \lambda(B_n) \leq \lambda\left(\bigcup_{k=n}^r B_k\right) \rightarrow \lambda\left(\bigcup_{k=n}^{\infty} B_k\right) \quad \text{as } r \rightarrow \infty \quad \text{by 1.2.7(a).}$$

Let $C = \lim_n \sup B_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k$. Now $\bigcup_{k=n}^{\infty} B_k \downarrow C$, and $0 \leq \lambda(\bigcup_{k=n}^{\infty} B_k) < \infty$ for all n . By 1.2.7(b), $\lambda(\bigcup_{k=n}^{\infty} B_k) \rightarrow \lambda(C)$ as $n \rightarrow \infty$. Thus

$$\sup \lambda = \lim_{n \rightarrow \infty} \lambda(A_n) \leq \lim_{n \rightarrow \infty} \lambda\left(\bigcup_{k=n}^{\infty} B_k\right) = \lambda(C) \leq \sup \lambda;$$

hence $\lambda(C) = \sup \lambda$. The above argument applied to $-\lambda$ yields $D \in \mathcal{F}$ with $\lambda(D) = \inf \lambda$. ■

We now prove the main theorem of this section.

2.1.2 Jordan–Hahn Decomposition Theorem. Let λ be a countably additive extended real-valued set function on the σ -field \mathcal{F} . Define

$$\lambda^+(A) = \sup\{\lambda(B) : B \in \mathcal{F}, B \subset A\},$$

$$\lambda^-(A) = -\inf\{\lambda(B) : B \in \mathcal{F}, B \subset A\}.$$

Then λ^+ and λ^- are measures on \mathcal{F} and $\lambda = \lambda^+ - \lambda^-$.

PROOF. We may assume λ never takes on the value $-\infty$. For if $-\infty$ belongs to the range of λ , $+\infty$ does not, by definition of a countably additive set function. Thus $-\lambda$ never takes on the value $-\infty$. But $(-\lambda)^+ = \lambda^-$ and $(-\lambda)^- = \lambda^+$, so that if the theorem is proved for $-\lambda$ it holds for λ as well.

Let D be a set on which λ attains its minimum, as in 2.1.1. Since $\lambda(\emptyset) = 0$, we have $-\infty < \lambda(D) \leq 0$. We claim that

$$\lambda(A \cap D) \leq 0, \quad \lambda(A \cap D^c) \geq 0 \quad \text{for all } A \in \mathcal{F}. \quad (1)$$

For if $\lambda(A \cap D) > 0$, then $\lambda(D) = \lambda(A \cap D) + \lambda(A^c \cap D)$. Since $\lambda(D)$ is finite, so are $\lambda(A \cap D)$ and $\lambda(A^c \cap D)$; hence $\lambda(A^c \cap D) = \lambda(D) - \lambda(A \cap D) < \lambda(D)$, contradicting the fact that $\lambda(D) = \inf \lambda$. If $\lambda(A \cap D^c) < 0$, then $\lambda(D \cup (A \cap D^c)) = \lambda(D) + \lambda(A \cap D^c) < \lambda(D)$, a contradiction.

We now show that

$$\lambda^+(A) = \lambda(A \cap D^c), \quad \lambda^-(A) = -\lambda(A \cap D). \quad (2)$$

The theorem will follow from this. We have, for $B \in \mathcal{F}$, $B \subset A$,

$$\begin{aligned} \lambda(B) &= \lambda(B \cap D) + \lambda(B \cap D^c) \\ &\leq \lambda(B \cap D^c) \quad \text{by (1)} \\ &\leq \lambda(B \cap D^c) + \lambda((A - B) \cap D^c) \\ &= \lambda(A \cap D^c). \end{aligned}$$

Thus $\lambda^+(A) \leq \lambda(A \cap D^c)$. But $\lambda(A \cap D^c) \leq \lambda^+(A)$ by definition of λ^+ , proving the first assertion. Similarly,

$$\begin{aligned} \lambda(B) &= \lambda(B \cap D) + \lambda(B \cap D^c) \\ &\geq \lambda(B \cap D) \\ &\geq \lambda(B \cap D) + \lambda((A - B) \cap D) \\ &= \lambda(A \cap D). \end{aligned}$$

Hence $-\lambda^-(A) \geq \lambda(A \cap D)$. But $\lambda(A \cap D) \geq -\lambda^-(A)$ by definition of λ^- , completing the proof. ■

2.1.3 Corollaries. Let λ be a countably additive extended real-valued set function on the σ -field \mathcal{F} .

- (a) The set function λ is the difference of two measures, at least one of which is finite.
- (b) If λ is finite ($\lambda(A)$ is never $\pm\infty$ for any $A \in \mathcal{F}$), then λ is bounded.
- (c) There is a set $D \in \mathcal{F}$ such that $\lambda(A \cap D) \leq 0$ and $\lambda(A \cap D^c) \geq 0$ for all $A \in \mathcal{F}$.
- (d) If D is any set in \mathcal{F} such that $\lambda(A \cap D) \leq 0$ and $\lambda(A \cap D^c) \geq 0$ for all $A \in \mathcal{F}$, then $\lambda^+(A) = \lambda(A \cap D^c)$ and $\lambda^-(A) = -\lambda(A \cap D)$ for all $A \in \mathcal{F}$.
- (e) If E is another set in \mathcal{F} such that $\lambda(A \cap E) \leq 0$ and $\lambda(A \cap E^c) \geq 0$ for all $A \in \mathcal{F}$, then $|\lambda|(D \Delta E) = 0$, where $|\lambda| = \lambda^+ + \lambda^-$.

PROOF. (a) If $\lambda > -\infty$, then in 2.1.2, λ^- is finite; if $\lambda < +\infty$, λ^+ is finite [see Eq. (2)].

(b) In 2.1.2, λ^+ and λ^- are both finite; hence for any $A \in \mathcal{F}$, $|\lambda(A)| \leq \lambda^+(\Omega) + \lambda^-(\Omega) < \infty$.

(c) This follows from (1) of 2.1.2.

(d) Repeat the part of the proof of 2.1.2 after Eq. (2).

(e) By (d), $\lambda^+(A) = \lambda(A \cap D^c)$, $A \in \mathcal{F}$; take $A = D \cap E^c$ to obtain $\lambda^+(D \cap E^c) = 0$. Also by (d), $\lambda^+(A) = \lambda(A \cap E^c)$, $A \in \mathcal{F}$; take $A = D^c \cap E$ to obtain $\lambda^+(D^c \cap E) = 0$. Therefore $\lambda^+(D \Delta E) = 0$. The same argument using $\lambda^-(A) = -\lambda(A \cap D) = -\lambda(A \cap E)$ shows that $\lambda^-(D \Delta E) = 0$. The result follows. ■

Corollary 2.1.3(d) is often useful in finding the Jordan–Hahn decomposition of a particular set function (see Problems 1 and 2).

2.1.4 Terminology. We call λ^+ the *upper variation* or *positive part* of λ , λ^- the *lower variation* or *negative part*, $|\lambda| = \lambda^+ + \lambda^-$ the *total variation*. Since $\lambda = \lambda^+ - \lambda^-$, it follows that $|\lambda(A)| \leq |\lambda|(A)$, $A \in \mathcal{F}$. For a sharper result, see Problem 4.

Note that if $A \in \mathcal{F}$, then $|\lambda|(A) = 0$ iff $\lambda(B) = 0$ for all $B \in \mathcal{F}$, $B \subset A$.

The phrase *signed measure* is sometimes used for the difference of two measures. By 2.1.3(a), this is synonymous (on a σ -field) with *countably additive set function*.

Problems

- Let P be an arbitrary probability measure on $\mathcal{B}(R)$, and let Q be point mass at 0, that is, $Q(B) = 1$ if $0 \in B$, $Q(B) = 0$ if $0 \notin B$. Find the Jordan–Hahn decomposition of the signed measure $\lambda = P - Q$.
- Let $\lambda(A) = \int_A f d\mu$, A in the σ -field \mathcal{F} , where $\int_\Omega f d\mu$ exists; thus λ is a signed measure on \mathcal{F} . Show that

$$\lambda^+(A) = \int_A f^+ d\mu, \quad \lambda^-(A) = \int_A f^- d\mu, \quad |\lambda|(A) = \int_A |f| d\mu.$$

- If a signed measure λ on the σ -field \mathcal{F} is the difference of two measures λ_1 and λ_2 , show that $\lambda_1 \geq \lambda^+$, $\lambda_2 \geq \lambda^-$.
- Let λ be a signed measure on the σ -field \mathcal{F} . Show that $|\lambda|(A) = \sup\{\sum_{i=1}^n |\lambda(E_i)| : E_1, E_2, \dots, E_n$ disjoint measurable subsets of A , $n = 1, 2, \dots\}$. Consequently, if λ_1 and λ_2 are signed measures on \mathcal{F} , then $|\lambda_1 + \lambda_2| \leq |\lambda_1| + |\lambda_2|$.

2.2 Radon–Nikodym Theorem and Related Results

If $(\Omega, \mathcal{F}, \mu)$ is a measure space, then $\lambda(A) = \int_A g d\mu$, $A \in \mathcal{F}$, defines a signed measure if $\int_\Omega g d\mu$ exists. Furthermore, if $A \in \mathcal{F}$ and $\mu(A) = 0$, then $\lambda(A) = 0$. For $gI_A = 0$ on A^c , so that $gI_A = 0$ a.e. [μ], and the result follows from 1.6.5(a).

If μ is a measure on the σ -field \mathcal{F} , and λ is a signed measure on \mathcal{F} , we say that λ is *absolutely continuous* with respect to μ (notation $\lambda \ll \mu$) iff $\mu(A) = 0$ implies $\lambda(A) = 0$ ($A \in \mathcal{F}$). Thus if λ is an indefinite integral with respect to μ , then $\lambda \ll \mu$. The Radon–Nikodym theorem is an assertion in the converse direction; if $\lambda \ll \mu$ (and μ is σ -finite on \mathcal{F}), then λ is an indefinite integral with respect to μ . As we shall see, large areas of analysis are based on this theorem.

2.2.1 Radon–Nikodym Theorem. Let μ be a σ -finite measure and λ a signed measure on the σ -field \mathcal{F} of subsets of Ω . Assume that λ is absolutely continuous with respect to μ . Then there is a Borel measurable function $g : \Omega \rightarrow \bar{\mathbb{R}}$ such that

$$\lambda(A) = \int_A g d\mu \quad \text{for all } A \in \mathcal{F}.$$

If h is another such function, then $g = h$ a.e. [μ].

PROOF. The uniqueness statement follows from 1.6.11. We break the existence proof into several parts.

(a) Assume λ and μ are finite measures.

Let \mathcal{S} be the set of all nonnegative μ -integrable functions f such that $\int_A f d\mu \leq \lambda(A)$ for all $A \in \mathcal{F}$. Partially order \mathcal{S} by calling $f \leq g$ iff $f \leq g$ a.e. [μ]. Let \mathcal{C} be a chain (totally ordered subset) of \mathcal{S} , and let $s = \sup\{\int_\Omega f d\mu : f \in \mathcal{C}\}$.

We can find functions $f_n \in \mathcal{C}$ with $\int_\Omega f_n d\mu \uparrow s$, and since \mathcal{C} is totally ordered, $f_n \leq f_{n+1}$ a.e. for all n . [If $f_{n+1} \leq f_n$ a.e., then $\int_\Omega f_{n+1} d\mu \leq \int_\Omega f_n d\mu \leq \int_\Omega f_{n+1} d\mu$; hence by 1.6.6(b), $f_n = f_{n+1}$ a.e.]

Thus f_n increases to a limit f a.e., and by the monotone convergence theorem, $\int_\Omega f d\mu = s$.

We show that f is an upper bound of \mathcal{C} . Let h be an element of \mathcal{C} . If $h \leq f_n$ a.e. for some n , then $h \leq f$ a.e. If $h \geq f_n$ a.e. for all n , then $h \geq f$ a.e.; hence $\int_\Omega h d\mu = \int_\Omega f d\mu = s$. But then $h = f$ a.e. and therefore f is an upper bound of \mathcal{C} . (We have $f \in \mathcal{S}$ by definition of \mathcal{S} .)

By Zorn's lemma, there is a maximal element $g \in \mathcal{S}$. We must show that $\lambda(A) = \int_A g d\mu$, $A \in \mathcal{F}$: Let $\lambda_1(A) = \lambda(A) - \int_A g d\mu$, $A \in \mathcal{F}$. Then λ_1 is a

measure, $\lambda_1 \ll \mu$, and $\lambda_1(\Omega) < \infty$. If λ_1 is not identically 0, then $\lambda_1(\Omega) > 0$, hence

$$\mu(\Omega) - k\lambda_1(\Omega) < 0 \quad \text{for some } k > 0. \quad (1)$$

Apply 2.1.3(c) to the signed measure $\mu - k\lambda_1$ to obtain $D \in \mathcal{F}$ such that for all $A \in \mathcal{F}$,

$$\mu(A \cap D) - k\lambda_1(A \cap D) \leq 0, \quad (2)$$

and

$$\mu(A \cap D^c) - k\lambda_1(A \cap D^c) \geq 0. \quad (3)$$

We claim that $\mu(D) > 0$. For if $\mu(D) = 0$, then $\lambda(D) = 0$ by absolute continuity, and therefore $\lambda_1(D) = 0$ by definition of λ_1 . Take $A = \Omega$ in (3) to obtain

$$\begin{aligned} 0 &\leq \mu(D^c) - k\lambda_1(D^c) \\ &= \mu(\Omega) - k\lambda_1(\Omega) \quad \text{since } \mu(D) = \lambda_1(D) = 0 \\ &< 0 \quad \text{by (1),} \end{aligned}$$

a contradiction. Define $h(\omega) = 1/k$, $\omega \in D$; $h(\omega) = 0$, $\omega \notin D$. If $A \in \mathcal{F}$,

$$\begin{aligned} \int_A h \, d\mu &= \frac{1}{k} \mu(A \cap D) \leq \lambda_1(A \cap D) \quad \text{by (2)} \\ &\leq \lambda_1(A) = \lambda(A) - \int_A g \, d\mu. \end{aligned}$$

Thus $\int_A (h + g) \, d\mu \leq \lambda(A)$. But $h + g > g$ on the set D , with $\mu(D) > 0$, contradicting the maximality of g . Thus $\lambda_1 \equiv 0$, and the result follows.

(b) Assume μ is a finite measure, λ a σ -finite measure.

Let Ω be the disjoint union of sets A_n with $\lambda(A_n) < \infty$, and let $\lambda_n(A) = \lambda(A \cap A_n)$, $A \in \mathcal{F}$, $n = 1, 2, \dots$. By part (a) we find a nonnegative Borel measurable g_n with $\lambda_n(A) = \int_A g_n \, d\mu$, $A \in \mathcal{F}$. Thus $\lambda(A) = \int_A g \, d\mu$, where $g = \sum_n g_n$.

(c) Assume μ is a finite measure, λ an arbitrary measure.

Let \mathcal{C} be the class of sets $C \in \mathcal{F}$ such that λ on C (that is, λ restricted to $\mathcal{F}_C = \{A \cap C : A \in \mathcal{F}\}$) is σ -finite; note that $\emptyset \in \mathcal{C}$, so \mathcal{C} is not empty. Let $s = \sup\{\mu(A) : A \in \mathcal{C}\}$ and pick $C_n \in \mathcal{C}$ with $\mu(C_n) \rightarrow s$. If $C = \bigcup_{n=1}^{\infty} C_n$, then $C \in \mathcal{C}$ by definition of \mathcal{C} , and $s \geq \mu(C) \geq \mu(C_n) \rightarrow s$; hence $\mu(C) = s$.

By part (b), there is a nonnegative $g': C \rightarrow \bar{R}$, measurable relative to \mathcal{F}_C and $\mathcal{B}(\bar{R})$, such that

$$\lambda(A \cap C) = \int_{A \cap C} g' d\mu \quad \text{for all } A \in \mathcal{F}.$$

Now consider an arbitrary set $A \in \mathcal{F}$.

Case 1: Let $\mu(A \cap C^c) > 0$. Then $\lambda(A \cap C^c) = \infty$, for if $\lambda(A \cap C^c) < \infty$, then $C \cup (A \cap C^c) \in \mathcal{C}$; hence

$$s \geq \mu(C \cup (A \cap C^c)) = \mu(C) + \mu(A \cap C^c) > \mu(C) = s,$$

a contradiction.

Case 2: Let $\mu(A \cap C^c) = 0$. Then $\lambda(A \cap C^c) = 0$ by absolute continuity.

Thus in either case, $\lambda(A \cap C^c) = \int_{A \cap C^c} \infty d\mu$. It follows that

$$\lambda(A) = \lambda(A \cap C) + \lambda(A \cap C^c) = \int_A g d\mu,$$

where $g = g'$ on C , $g = \infty$ on C^c .

(d) Assume μ is a σ -finite measure, λ an arbitrary measure.

Let Ω be the union of disjoint sets A_n with $\mu(A_n) < \infty$. By part (c), there is a nonnegative function $g_n: A_n \rightarrow \bar{R}$, measurable with respect to \mathcal{F}_{A_n} and $\mathcal{B}(\bar{R})$, such that $\lambda(A \cap A_n) = \int_{A \cap A_n} g_n d\mu$, $A \in \mathcal{F}$. We may write this as $\lambda(A \cap A_n) = \int_A g_n d\mu$ where $g_n(\omega)$ is taken as 0 for $\omega \notin A_n$. Thus $\lambda(A) = \sum_n \lambda(A \cap A_n) = \sum_n \int_A g_n d\mu = \int_A g d\mu$, where $g = \sum_n g_n$.

(e) Assume μ is a σ -finite measure, λ an arbitrary signed measure.

Write $\lambda = \lambda^+ - \lambda^-$ where, say, λ^- is finite. By part (d), there are non-negative Borel measurable functions g_1 and g_2 such that

$$\lambda^+(A) = \int_A g_1 d\mu, \quad \lambda^-(A) = \int_A g_2 d\mu, \quad A \in \mathcal{F}.$$

Since λ^- is finite, g_2 is integrable; hence by 1.6.3 and 1.6.6(a), $\lambda(A) = \int_A (g_1 - g_2) d\mu$. ■

2.2.2 Corollaries.

Under the hypothesis of 2.2.1,

- (a) If λ is finite, then g is μ -integrable, hence finite a.e. $[\mu]$.
- (b) If $|\lambda|$ is σ -finite, so that Ω can be expressed as a countable union of sets A_n such that $|\lambda|(A_n)$ is finite (equivalently $\lambda(A_n)$ is finite), then g is finite a.e. $[\mu]$.
- (c) If λ is a measure, then $g \geq 0$ a.e. $[\mu]$.

PROOF. All results may be obtained by examining the proof of 2.2.1. Alternatively, we may proceed as follows:

- (a) Observe that $\lambda(\Omega) = \int_{\Omega} g d\mu$, finite by hypothesis.
- (b) By (a), g is finite a.e. $[\mu]$ on each A_n , hence finite a.e. $[\mu]$ on Ω .
- (c) Let $A = \{\omega : g(\omega) < 0\}$; then $0 \leq \lambda(A) = \int_A g d\mu \leq 0$. Thus $-gI_A$ is a nonnegative function whose integral is 0, so that $gI_A = 0$ a.e. $[\mu]$ by 1.6.6(b). Since $gI_A < 0$ on A , we must have $\mu(A) = 0$. ■

If $\lambda(A) = \int_A g d\mu$ for each $A \in \mathcal{F}$, g is called the *Radon–Nikodym derivative* or *density* of λ with respect to μ , written $d\lambda/d\mu$. If μ is Lebesgue measure, then g is often called simply the *density* of λ .

There are converse assertions to 2.2.2(a) and (c). Suppose that

$$\lambda(A) = \int_A g d\mu, \quad A \in \mathcal{F},$$

where $\int_{\Omega} g d\mu$ is assumed to exist. If g is μ -integrable, then λ is finite; if $g \geq 0$ a.e. $[\mu]$, then $\lambda \geq 0$, so that λ is a measure. (Note that σ -finiteness of μ is not assumed.) However, the converse to 2.2.2.(b) is false; if g is finite a.e. $[\mu]$, $|\lambda|$ need not be σ -finite (see Problem 1).

We now consider a property that is in a sense opposite to absolute continuity.

2.2.3 Definitions. Let μ_1 and μ_2 be measures on the σ -field \mathcal{F} . We say that μ_1 is *singular* with respect to μ_2 (written $\mu_1 \perp \mu_2$) iff there is a set $A \in \mathcal{F}$ such that $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$; note μ_1 is singular with respect to μ_2 iff μ_2 is singular with respect to μ_1 , so we may say that μ_1 and μ_2 are *mutually singular*. If λ_1 and λ_2 are signed measures on \mathcal{F} , we say that λ_1 and λ_2 are mutually singular iff $|\lambda_1| \perp |\lambda_2|$.

If $\mu_1 \perp \mu_2$, with $\mu_1(A) = \mu_2(A^c) = 0$, then μ_2 only assigns positive measure to subsets of A . Thus μ_2 concentrates its total effect on a set of μ_1 -measure 0; on the other hand, if $\mu_2 \ll \mu_1$, μ_2 can have no effect on sets of μ_1 -measure 0.

If λ is a signed measure with positive part λ^+ and negative part λ^- , we have $\lambda^+ \perp \lambda^-$ by 2.1.3(c) and (d).

Before establishing some facts about absolute continuity and singularity, we need the following lemma. Although the proof is quite simple, the result is applied very often in analysis, especially in probability theory.

2.2.4 Borel–Cantelli Lemma. If $A_1, A_2, \dots \in \mathcal{F}$ and $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then $\mu(\limsup_n A_n) = 0$.

PROOF. Recall that $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$; hence

$$\begin{aligned}\mu(\limsup_n A_n) &\leq \mu\left(\bigcup_{k=n}^{\infty} A_k\right) \quad \text{for all } n \\ &\leq \sum_{k=n}^{\infty} \mu(A_k) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \blacksquare\end{aligned}$$

2.2.5 Lemma. Let μ be a measure, and λ_1 and λ_2 signed measures, on the σ -field \mathcal{F} .

- (a) If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.
- (b) If $\lambda_1 \ll \mu$, then $|\lambda_1| \ll \mu$, and conversely.
- (c) If $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$.
- (d) If $\lambda_1 \ll \mu$ and $\lambda_1 \perp \mu$, then $\lambda_1 \equiv 0$.
- (e) If λ_1 is finite, then $\lambda_1 \ll \mu$ iff $\lim_{\mu(A) \rightarrow 0} \lambda_1(A) = 0$.

PROOF. (a) Let $\mu(A) = \mu(B) = 0$, $|\lambda_1|(A^c) = |\lambda_2|(B^c) = 0$. Then $\mu(A \cup B) = 0$ and $\lambda_1(C) = \lambda_2(C) = 0$ for every $C \in \mathcal{F}$ with $C \subset A^c \cap B^c$; hence $|\lambda_1 + \lambda_2|[(A \cup B)^c] = 0$.

(b) Let $\mu(A) = 0$. If $\lambda_1^+(A) > 0$, then (see 2.1.2) $\lambda_1(B) > 0$ for some $B \subset A$; since $\mu(B) = 0$, this is a contradiction. It follows that λ_1^+ , and similarly λ_1^- , is absolutely continuous with respect to μ ; hence $|\lambda_1| \ll \mu$. (This may also be proved using Section 2.1, Problem 4.) The converse is clear.

(c) Let $\mu(A) = 0$, $|\lambda_2|(A^c) = 0$. By (b), $|\lambda_1|(A) = 0$, so $|\lambda_1| \perp |\lambda_2|$.
(d) By (c), $\lambda_1 \perp \lambda_2$; hence for some $A \in \mathcal{F}$, $|\lambda_1|(A) = |\lambda_1|(A^c) = 0$. Thus $|\lambda_1|(\Omega) = 0$.

(e) If $\mu(A_n) \rightarrow 0$ implies $\lambda_1(A_n) \rightarrow 0$, and $\mu(A) = 0$, set $A_n \equiv A$ to conclude that $\lambda_1(A) = 0$, so $\lambda \ll \mu$.

Conversely, let $\lambda_1 \ll \mu$.

If $\lim_{\mu(A) \rightarrow 0} |\lambda_1|(A) \neq 0$ we can find, for some $\varepsilon > 0$, sets $A_n \in \mathcal{F}$ with $\mu(A_n) < 2^{-n}$ and $|\lambda_1|(A_n) \geq \varepsilon$ for all n . Let $A = \limsup_n A_n$; by 2.2.4, $\mu(A) = 0$. But $|\lambda_1|(\bigcup_{k=n}^{\infty} A_k) \geq |\lambda_1|(A_n) \geq \varepsilon$ for all n ; hence by 1.2.7(b), $|\lambda_1|(A) \geq \varepsilon$, contradicting (b). Thus $\lim_{\mu(A) \rightarrow 0} |\lambda_1|(A) = 0$, and the result follows since $|\lambda_1(A)| \leq |\lambda_1|(A)$. \blacksquare

If λ_1 is an indefinite integral with respect to μ (hence $\lambda_1 \ll \mu$), then 2.2.5(e) has an easier proof. If $\lambda_1(A) = \int_A f d\mu$, $A \in \mathcal{F}$, then

$$\begin{aligned}\int_A |f| d\mu &= \int_{A \cap \{|f| \leq n\}} |f| d\mu + \int_{A \cap \{|f| > n\}} |f| d\mu \\ &\leq n\mu(A) + \int_{\{|f| > n\}} |f| d\mu.\end{aligned}$$

By 1.6.1 and 1.2.7(b), $\int_{\{|f| > n\}} |f| d\mu$ may be made less than $\varepsilon/2$ for large n , say $n \geq N$. Fix $n = N$ and take $\mu(A) < \varepsilon/2N$, so that $\int_A |f| d\mu < \varepsilon$.

If μ is a σ -finite measure and λ a signed measure on the σ -field \mathcal{F} , λ may be neither absolutely continuous nor singular with respect to μ . However, if $|\lambda|$ is σ -finite, the two concepts of absolute continuity and singularity are adequate to describe the relation between λ and μ , in the sense that λ can be written as the sum of two signed measures, one absolutely continuous and the other singular with respect to μ .

2.2.6 Lebesgue Decomposition Theorem. Let μ be a σ -finite measure on the σ -field \mathcal{F} , λ a σ -finite signed measure (that is, $|\lambda|$ is σ -finite). Then λ has a unique decomposition as $\lambda_1 + \lambda_2$, where λ_1 and λ_2 are signed measures such that $\lambda_1 \ll \mu$, $\lambda_2 \perp \mu$.

PROOF. First assume λ is a σ -finite measure. Let $m = \mu + \lambda$, also a σ -finite measure. Then μ and λ are each absolutely continuous with respect to m ; hence by 2.2.1 and 2.2.2(c) there are nonnegative Borel measurable functions f and g such that $\mu(A) = \int_A f dm$, $\lambda(A) = \int_A g dm$, $A \in \mathcal{F}$.

Let $B = \{\omega : f(\omega) > 0\}$, $C = B^c = \{\omega : f(\omega) = 0\}$, and define, for each $A \in \mathcal{F}$,

$$\lambda_1(A) = \lambda(A \cap B), \quad \lambda_2(A) = \lambda(A \cap C).$$

Thus $\lambda_1 + \lambda_2 = \lambda$. In fact, $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$. To prove $\lambda_1 \ll \mu$, assume $\mu(A) = 0$. Then $\int_A f dm = 0$, hence $f = 0$ a.e. [m] on A . But $f > 0$ on $A \cap B$; hence $m(A \cap B) = 0$, and consequently $\lambda(A \cap B) = 0$; in other words, $\lambda_1(A) = 0$. Thus $\lambda_1 \ll \mu$.

To prove $\lambda_2 \perp \mu$, observe that $\lambda_2(B) = 0$ and $\mu(B^c) = \mu(C) = \int_C 0 dm = 0$.

Now if λ is a σ -finite signed measure, the above argument applied to λ^+ and λ^- proves the existence of the desired decomposition.

To prove uniqueness, first assume λ finite. If $\lambda = \lambda_1 + \lambda_2 = \lambda'_1 + \lambda'_2$, where $\lambda_1, \lambda'_1 \ll \mu$, $\lambda_2, \lambda'_2 \perp \mu$, then $\lambda_1 - \lambda'_1 = \lambda'_2 - \lambda_2$ is both absolutely continuous and singular with respect to μ ; hence is identically 0 by 2.2.5(d). If λ is σ -finite and Ω is the disjoint union of sets A_n with $|\lambda|(A_n) < \infty$, apply the above argument to each A_n and put the results together to obtain uniqueness of λ_1 and λ_2 . ■

Problems

- Give an example of a measure μ and a nonnegative finite-valued Borel measurable function g such that the measure λ defined by $\lambda(A) = \int_A g d\mu$ is not σ -finite.

2. If $\lambda(A) = \int_A g d\mu$, $A \in \mathcal{F}$, and g is μ -integrable, we know that λ is finite; in particular, $A = \{\omega : g(\omega) \neq 0\}$ has finite λ -measure. Show that A has σ -finite μ -measure, that is, it is a countable union of sets of finite μ -measure. Give an example to show that $\mu(A)$ need not be finite.
3. Give an example in which the conclusion of the Radon–Nikodym theorem fails; in other words, $\lambda \ll \mu$ but there is no Borel measurable g such that $\lambda(A) = \int_A g d\mu$ for all $A \in \mathcal{F}$. Of course μ cannot be σ -finite.
4. (*A chain rule*) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and g a nonnegative Borel measurable function on Ω . Define a measure λ on \mathcal{F} by

$$\lambda(A) = \int_A g d\mu, \quad A \in \mathcal{F}.$$

Show that if f is a Borel measurable function on Ω ,

$$\int_{\Omega} f d\lambda = \int_{\Omega} fg d\mu$$

in the sense that if one of the integrals exists, so does the other, and the two integrals are equal. (Intuitively, $d\lambda/d\mu = g$, so that $d\lambda = g d\mu$.)

5. Show that Theorem 2.2.5(e) fails if λ_1 is not finite.
6. (*Complex measures*) If (Ω, \mathcal{F}) is a measurable space, a *complex measure* λ on \mathcal{F} is a countably additive complex-valued set function; that is, $\lambda = \lambda_1 + i\lambda_2$, where λ_1 and λ_2 are finite signed measures.

- (a) Define the *total variation* of λ as

$$|\lambda|(A) = \sup \left\{ \sum_{i=1}^n |\lambda(E_i)| : E_1, \dots, E_n \text{ disjoint measurable subsets of } A, n = 1, 2, \dots \right\}.$$

Show that $|\lambda|$ is a measure on \mathcal{F} . (The definition is consistent with the earlier notion of total variation of a signed measure; see Section 2.1, Problem 4.)

In the discussion below, λ 's, with various subscripts, denote arbitrary measures (real signed measures or complex measures), and μ denotes a nonnegative real measure. We define $\lambda \ll \mu$ in the usual way; if $A \in \mathcal{F}$ and $\mu(A) = 0$, then $\lambda(A) = 0$. Define $\lambda_1 \perp \lambda_2$ iff $|\lambda_1| \perp |\lambda_2|$. Establish the following results.

- (b) $|\lambda_1 + \lambda_2| \leq |\lambda_1| + |\lambda_2|$; $|a\lambda| = |a| |\lambda|$ for any complex number a . In particular if $\lambda = \lambda_1 + i\lambda_2$ is a complex measure, then

$$|\lambda| \leq |\lambda_1| + |\lambda_2|; \quad \text{hence} \quad |\lambda|(\Omega) < \infty \quad \text{by 2.1.3(b).}$$

- (c) If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.
- (d) If $\lambda \ll \mu$, then $|\lambda| \ll \mu$, and conversely.
- (e) If $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$.
- (f) If $\lambda \ll \mu$ and $\lambda \perp \mu$, then $\lambda = 0$.
- (g) If λ is finite, then $\lambda \ll \mu$ iff $\lim_{\mu(A) \rightarrow 0} \lambda(A) = 0$.

2.3 Applications to Real Analysis

We are going to apply the concepts of the previous section to some problems involving functions of a real variable. If $[a, b]$ is a closed bounded interval of reals and $f: [a, b] \rightarrow \mathbb{R}$, f is said to be *absolutely continuous* iff for each $\varepsilon > 0$ there is a $\delta > 0$ such that for all positive integers n and all families $(a_1, b_1), \dots, (a_n, b_n)$ of disjoint open subintervals of $[a, b]$ of total length at most δ , we have

$$\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \varepsilon.$$

It is immediate that this property holds also for countably infinite families of disjoint open intervals of total length at most δ . It also follows from the definition that f is continuous.

We can connect absolute continuity of functions with the earlier notion of absolute continuity of measures, as follows:

2.3.1 Theorem. Suppose that F and G are distribution functions on $[a, b]$, with corresponding (finite) Lebesgue-Stieltjes measures μ_1 and μ_2 . Let $f = F - G$, $\mu = \mu_1 - \mu_2$, so that μ is a finite signed measure on $\mathcal{B}[a, b]$, with $\mu(x, y] = f(y) - f(x)$, $x < y$. If m is Lebesgue measure on $\mathcal{B}[a, b]$, then $\mu \ll m$ iff f is absolutely continuous.

PROOF. Assume $\mu \ll m$. If $\varepsilon > 0$, by 2.2.5(b) and (e), there is a $\delta > 0$ such that $m(A) \leq \delta$ implies $|\mu|(A) \leq \varepsilon$. Thus if $(a_1, b_1), \dots, (a_n, b_n)$ are disjoint open intervals of total length at most δ ,

$$\begin{aligned} \sum_{i=1}^n |f(b_i) - f(a_i)| &= \sum_{i=1}^n |\mu(a_i, b_i)| \\ &= \sum_{i=1}^n |\mu(a_i, b_i)| \leq \varepsilon. \end{aligned}$$

(Note that $\mu\{b_i\} = 0$ since $\mu \ll m$.) Therefore f is absolutely continuous.

Now assume f absolutely continuous; if $\varepsilon > 0$, choose $\delta > 0$ as in the definition of absolute continuity. If $m(A) = 0$, we must show that $\mu(A) = 0$. We use Problem 12, Section 1.4:

$$\begin{aligned} m(A) &= \inf\{m(V): V \supset A, \quad V \text{ open}\}, \\ \mu_i(A) &= \inf\{\mu_i(V): V \supset A, \quad V \text{ open}\}, \quad i = 1, 2. \end{aligned}$$

(This problem assumes that the measures are defined on $\mathcal{B}(R)$ rather than $\mathcal{B}[a, b]$. The easiest way out is to extend all measures to $\mathcal{B}(R)$ by assigning measure 0 to $R - [a, b]$.) Since a finite intersection of open sets is open, we can find a decreasing sequence $\{V_n\}$ of open sets such that $\mu(V_n) \rightarrow \mu(A)$ and $m(V_n) \rightarrow m(A) = 0$.

Choose n large enough so that $m(V_n) < \delta$; if V_n is the disjoint union of the open intervals (a_i, b_i) , $i = 1, 2, \dots$, then $|\mu(V_n)| \leq \sum_i |\mu(a_i, b_i)|$. But f is continuous, hence

$$\mu\{b_i\} = \lim_{n \rightarrow \infty} \mu(b_i - 1/n, b_i] = \lim_{n \rightarrow \infty} [f(b_i) - f(b_i - 1/n)] = 0.$$

Therefore

$$|\mu(V_n)| \leq \sum_i |\mu(a_i, b_i)| = \sum_i |f(b_i) - f(a_i)| \leq \varepsilon.$$

Since ε is arbitrary and $\mu(V_n) \rightarrow \mu(A)$, we have $\mu(A) = 0$. ■

If $f: R \rightarrow R$, absolute continuity of f is defined exactly as above. If F and G are bounded distribution functions on R with corresponding Lebesgue-Stieltjes measures μ_1 and μ_2 , and $f = F - G$, $\mu = \mu_1 - \mu_2$ [a finite signed measure on $\mathcal{B}(R)$], then f is absolutely continuous iff μ is absolutely continuous with respect to Lebesgue measure; the proof is the same as in 2.3.1.

Any absolutely continuous function on $[a, b]$ can be represented as the difference of two absolutely continuous increasing functions. We prove this in a sequence of steps.

If $f: [a, b] \rightarrow R$ and $P: a = x_0 < x_1 < \dots < x_n = b$ is a partition of $[a, b]$, define

$$V(P) = \sum_{i=1}^n |f(x_i) - f(x_{i-1})|.$$

The sup of $V(P)$ over all partitions of $[a, b]$ is called the *variation* of f on $[a, b]$, written $V_f(a, b)$, or simply $V(a, b)$ if f is understood. We say that f is of *bounded variation* on $[a, b]$ iff $V(a, b) < \infty$. If $a < c < b$, a brief argument shows that $V(a, b) = V(a, c) + V(c, b)$.

2.3.2 Lemma. If $f: [a, b] \rightarrow R$ and f is absolutely continuous on $[a, b]$, then f is of bounded variation on $[a, b]$.

PROOF. Pick any $\varepsilon > 0$, and let $\delta > 0$ be chosen as in the definition of absolute continuity. If P is any partition of $[a, b]$, there is a refinement Q of P consisting of subintervals of length less than $\delta/2$. If $Q: a = x_0 < x_1 < \dots < x_n = b$, let $i_0 = 0$, and let i_1 be the largest integer such that $x_{i_1} - x_{i_0} < \delta$; let i_2 be the largest integer greater than i_1 such that $x_{i_2} - x_{i_1} < \delta$, and continue in this fashion until the process terminates, say with $i_r = n$. Now $x_{i_k} - x_{i_{k-1}} \geq \delta/2$, $k = 1, 2, \dots, r-1$, by construction of Q ; hence

$$r \leq 1 + \frac{2(b-a)}{\delta} = M.$$

By absolute continuity, $V(Q) \leq M\varepsilon$. But $V(P) \leq V(Q)$ since the refining process can never decrease V ; the result follows. ■

It is immediate that a monotone function F on $[a, b]$ is of bounded variation: $V_F(a, b) = |F(b) - F(a)|$. Thus if $f = F - G$, where F and G are increasing, then f is of bounded variation. The converse is also true.

2.3.3 Lemma. If $f: [a, b] \rightarrow R$ and f is of bounded variation on $[a, b]$, then there are increasing functions F and G on $[a, b]$ such that $f = F - G$. If f is absolutely continuous, F and G may also be taken as absolutely continuous.

PROOF. Let $F(x) = V_f(a, x)$, $a \leq x \leq b$; F is increasing, for if $h \geq 0$, $V(a, x+h) - V(a, x) = V(x, x+h) \geq 0$. If $G(x) = F(x) - f(x)$, then G is also increasing. For if $x_1 < x_2$, then

$$\begin{aligned} G(x_2) - G(x_1) &= F(x_2) - F(x_1) - (f(x_2) - f(x_1)) \\ &= V(x_1, x_2) - (f(x_2) - f(x_1)) \\ &\geq V(x_1, x_2) - |f(x_2) - f(x_1)| \\ &\geq 0 \quad \text{by definition of } V(x_1, x_2). \end{aligned}$$

Now assume f absolutely continuous. If $\varepsilon > 0$, choose $\delta > 0$ as in the definition of absolute continuity. Let $(a_1, b_1), \dots, (a_n, b_n)$ be disjoint open intervals with total length at most δ . If P_i is a partition of $[a_i, b_i]$, $i = 1, 2, \dots, n$, then

$$\sum_{i=1}^n V(P_i) \leq \varepsilon \quad \text{by absolute continuity of } f.$$

Take the sup successively over P_1, \dots, P_n to obtain

$$\sum_{i=1}^n V(a_i, b_i) \leq \varepsilon;$$

in other words,

$$\sum_{i=1}^n [F(b_i) - F(a_i)] \leq \varepsilon.$$

Therefore F is absolutely continuous. Since sums and differences of absolutely continuous functions are absolutely continuous, G is also absolutely continuous. ■

We have seen that there is a close connection between absolute continuity and indefinite integrals, via the Radon–Nikodym theorem. The connection carries over to real analysis, as follows:

2.3.4 Theorem. Let $f: [a, b] \rightarrow R$. Then f is absolutely continuous on $[a, b]$ iff f is an indefinite integral, that is, iff

$$f(x) - f(a) = \int_a^x g(t) dt, \quad a \leq x \leq b,$$

where $g: [a, b] \rightarrow R$ is Borel measurable and integrable with respect to Lebesgue measure.

PROOF. First assume f absolutely continuous. By 2.3.3, it is sufficient to assume f increasing. If μ is the Lebesgue–Stieltjes measure corresponding to f , and m is Lebesgue measure, then $\mu \ll m$ by 2.3.1. By the Radon–Nikodym theorem, there is an m -integrable function g such that $\mu(A) = \int_A g dm$ for all Borel subsets A of $[a, b]$. Take $A = [a, x]$ to obtain $f(x) - f(a) = \int_a^x g(t) dt$.

Conversely, assume $f(x) - f(a) = \int_a^x g(t) dt$. It is sufficient to assume $g \geq 0$ (if not, consider g^+ and g^- separately). Define $\mu(A) = \int_A g dm$, $A \in \mathcal{B}[a, b]$; then $\mu \ll m$, and if F is a distribution function corresponding to μ , F is absolutely continuous by 2.3.1. But

$$F(x) - F(a) = \mu(a, x] = \int_a^x g(t) dt = f(x) - f(a).$$

Therefore f is absolutely continuous. ■

If g is Lebesgue integrable on R , the “if” part of the proof of 2.3.4 shows that the function defined by $\int_{-\infty}^x g(t) dt$, $x \in R$, is absolutely continuous, hence continuous, on R . Another way of proving continuity is to observe that

$$\int_{-\infty}^{x+h} g(t) dt - \int_{-\infty}^x g(t) dt = \int_{-\infty}^{\infty} g(t) I_{(x, x+h)}(t) dt$$

if $h > 0$, and this approaches 0 as $h \rightarrow 0$, by the dominated convergence theorem.

If $f(x) - f(a) = \int_a^x g(t) dt$, $a \leq x \leq b$, and g is continuous at x , then f is differentiable at x and $f'(x) = g(x)$; the proof given in calculus carries over.

If the continuity hypothesis is dropped, we can prove that $f'(x) = g(x)$ for *almost* every $x \in [a, b]$. One approach to this result is via the theory of differentiation of measures, which we now describe.

2.3.5 Definition. For the remainder of this section, μ is a signed measure on the Borel sets of R^k , assumed finite on bounded sets; thus if μ is nonnegative, it is a Lebesgue–Stieltjes measure. If m is Lebesgue measure, we define, for each $x \in R^k$,

$$(\bar{D}\mu)(x) = \limsup_{r \rightarrow 0} \frac{\mu(C_r)}{m(C_r)}, \quad (\underline{D}\mu)(x) = \liminf_{r \rightarrow 0} \frac{\mu(C_r)}{m(C_r)},$$

where the C_r range over all open cubes of diameter less than r that contain x . It will be convenient (although not essential) to assume that all cubes have edges parallel to the coordinate axes.

We say that μ is *differentiable* at x iff $\bar{D}\mu$ and $\underline{D}\mu$ are equal and finite at x ; we write $(D\mu)(x)$ for the common value. Thus μ is differentiable at x iff for every sequence $\{C_n\}$ of open cubes containing x , with the diameter of C_n approaching 0, $\mu(C_n)/m(C_n)$ approaches a finite limit, independent of the particular sequence.

The following result will play an important role:

2.3.6 Lemma. If $\{C_1, \dots, C_n\}$ is a family of open cubes in R^k , there is a disjoint subfamily $\{C_{i_1}, \dots, C_{i_s}\}$ such that $m(\bigcup_{j=1}^n C_j) \leq 3^k \sum_{p=1}^s m(C_{i_p})$.

PROOF. Assume that the diameter of C_i decreases with i . Set $i_1 = 1$, and take i_2 to be the smallest index greater than i_1 such that C_{i_2} is disjoint from C_{i_1} ; let i_3 be the smallest index greater than i_2 such that C_{i_3} is disjoint from $C_{i_1} \cup C_{i_2}$. Continue in this fashion to obtain disjoint sets C_{i_1}, \dots, C_{i_s} . Now for any $j = 1, \dots, n$, we have $C_j \cap C_{i_p} \neq \emptyset$ for some $i_p \leq j$, for if not, j is not one of the i_p , hence $i_p < j < i_{p+1}$ for some p (or $i_s < j$). But $C_j \cap (C_{i_1} \cup \dots \cup C_{i_p})$ is assumed empty, contradicting the definition of i_{p+1} .

If B_p is the open cube with the same center as C_{i_p} and diameter three times as large, then since $C_j \cap C_{i_p} \neq \emptyset$ and diameter $C_j \leq$ diameter C_{i_p} , we have $C_j \subset B_p$. Therefore,

$$m\left(\bigcup_{j=1}^n C_j\right) \leq m\left(\bigcup_{p=1}^s B_p\right) \leq \sum_{p=1}^s m(B_p) = 3^k \sum_{p=1}^s m(C_{i_p}). \blacksquare$$

We now prove the first differentiation result.

2.3.7 Lemma. Let μ be a Lebesgue–Stieltjes measure on the Borel sets of R^k . If $\mu(A) = 0$, then $D\mu = 0$ a.e. [m] on A .

PROOF. If $a > 0$, let $B = \{x \in A : (\bar{D}\mu)(x) > a\}$. (Note that $\{x : \sup_{C_r} \mu(C_r)/m(C_r) > a\}$ is open, and it follows that B is a Borel set.) Fix $r > 0$, and let K be a compact subset of B . If $x \in K$, there is an open cube C_r of diameter less than r with $x \in C_r$, and $\mu(C_r) > am(C_r)$. By compactness, K is covered by finitely many of the cubes, say C_1, \dots, C_n . If $\{C_{i_1}, \dots, C_{i_s}\}$ is the subcollection of 2.3.6, we have

$$m(K) \leq m\left(\bigcup_{j=1}^n C_j\right) \leq 3^k \sum_{p=1}^s m(C_{i_p}) \leq \frac{3^k}{a} \sum_{p=1}^s \mu(C_{i_p}) = \frac{3^k}{a} \mu\left(\bigcup_{p=1}^s C_{i_p}\right) \leq \frac{3^k}{a} \mu(K_r),$$

where $K_r = \{x \in R^k : \text{dist}(x, K) < r\}$. Since r is arbitrary, we have $m(K) \leq 3^k \mu(K)/a \leq 3^k \mu(A)/a = 0$. Take the sup over K to obtain, by Problem 12, Section 1.4, $m(B) = 0$, and since a is arbitrary, it follows that $\bar{D}\mu \leq 0$ a.e. [m] on A . But $\mu \geq 0$; hence $0 \leq \underline{D}\mu \leq \bar{D}\mu$, so that $D\mu = 0$ a.e. [m] on A . ■

We are going to show that $D\mu$ exists a.e. [m], and to do this the Lebesgue decomposition theorem is helpful. We write $\mu = \mu_1 + \mu_2$, where $\mu_1 \ll m$, $\mu_2 \perp m$. If $|\mu_2|(A) = 0$ and $m(A^c) = 0$, then by 2.3.7, $D\mu_2^+ = D\mu_2^- = 0$ a.e. [m] on A ; hence a.e. [m] on R^k . Thus $D\mu_2 = 0$ a.e. [m] on R^k .

By the Radon–Nikodym theorem, we have $\mu_1(E) = \int_E g dm$, $E \in \mathcal{B}(R^k)$, for some Borel measurable function g . As might be expected intuitively, g is (a.e.) the derivative of μ_1 ; hence $D\mu = g$ a.e. [m].

2.3.8 Theorem. Let μ be a signed measure on $\mathcal{B}(R^k)$ that is finite on bounded sets, and let $\mu = \mu_1 + \mu_2$, where $\mu_1 \ll m$ and $\mu_2 \perp m$. Then $D\mu$ exists a.e. [m] and coincides a.e. [m] with the Radon–Nikodym derivative $g = d\mu_1/dm$.

PROOF. If $a \in R$ and C is an open cube of diameter less than r ,

$$\mu_1(C) - am(C) = \int_C (g - a) dm \leq \int_{C \cap \{g \geq a\}} (g - a) dm.$$

If $\lambda(E) = \int_{E \cap \{g \geq a\}} (g - a) dm$, $E \in \mathcal{B}(R^k)$, and $A = \{g < a\}$, then $\lambda(A) = 0$; so by 2.3.7, $D\lambda = 0$ a.e. [m] on A . But

$$\frac{\mu_1(C)}{m(C)} \leq a + \frac{\lambda(C)}{m(C)};$$

hence $\bar{D}\mu_1 \leq a$ a.e. [m] on A . Therefore, if $E_a = \{x \in R^k : g(x) < a < (\bar{D}\mu_1)(x)\}$, then $m(E_a) = 0$. Since $\{\bar{D}\mu_1 > g\} \subset \bigcup \{E_a : a \text{ rational}\}$, we have $\bar{D}\mu_1 \leq g$ a.e. [m]. Replace μ_1 by $-\mu_1$ and g by $-g$ to obtain $\underline{D}\mu_1 \geq g$ a.e. [m]. By 2.2.2(b), g is finite a.e. [m], and the result follows. ■

We now return to functions on the real line.

2.3.9 Theorem. Let $f: [a, b] \rightarrow R$ be an increasing function. Then the derivative of f exists at almost every point of $[a, b]$ (with respect to Lebesgue measure). Thus by 2.3.3, a function of bounded variation is differentiable almost everywhere.

PROOF. Let μ be the Lebesgue–Stieltjes measure corresponding to f ; by 2.3.8, $D\mu$ exists a.e. [m]; we show that $D\mu = f'$ a.e. [m]. If $a \leq x \leq b$ and μ is differentiable at x , then f is continuous at x by definition of μ . If $\lim_{h \rightarrow 0} [f(x + h) - f(x)]/h \neq (D\mu)(x) = c$, there is an $\varepsilon > 0$ and a sequence $h_n \rightarrow 0$ with all h_n of the same sign and $|[f(x + h_n) - f(x)]/h_n - c| \geq \varepsilon$ for all n . Assuming all $h_n > 0$, we can find numbers $k_n > 0$ such that

$$\left| \frac{f(x + h_n) - f(x - k_n)}{h_n + k_n} - c \right| \geq \frac{\varepsilon}{2}$$

for all n , and since f has only countably many discontinuities, it may be assumed that f is continuous at $x + h_n$ and $x - k_n$. Thus we conclude that $\mu(x - k_n, x + h_n)/(h_n + k_n) \nrightarrow c$, a contradiction. ■

We now prove the main theorem on absolutely continuous functions.

2.3.10 Theorem. Let f be absolutely continuous on $[a, b]$, with $f(x) - f(a) = \int_a^x g(t) dt$, as in 2.3.4. Then $f' = g$ almost everywhere on $[a, b]$ (Lebesgue measure). Thus by 2.3.4, f is absolutely continuous iff f is the integral of its derivative, that is,

$$f(x) - f(a) = \int_a^x f'(t) dt, \quad a \leq x \leq b.$$

PROOF. We may assume $g \geq 0$ (if not, consider g^+ and g^-). If $\mu_1(A) = \int_A g dm$, $A \in \mathcal{B}(R^k)$, then $D\mu_1 = g$ a.e. [m] by 2.3.8. But if $a \leq x \leq y \leq b$, then $\mu_1(x, y) = f(y) - f(x)$, so that μ_1 is the Lebesgue–Stieltjes measure corresponding to f . Thus by the proof of 2.3.9, $D\mu_1 = f'$ a.e. [m]. ■

Problems

- Let F be a bounded distribution function on R . Use the Lebesgue decomposition theorem to show that F may be represented uniquely (up to additive constants) as $F_1 + F_2 + F_3$, where the distribution functions F_j , $j = 1, 2, 3$ (and the corresponding Lebesgue–Stieltjes measures μ_j) have the following properties:
 - F_1 is discrete (that is, μ_1 is concentrated on a countable set of points).

- (b) F_2 is absolutely continuous (μ_2 is absolutely continuous with respect to Lebesgue measure; see 2.3.1).
(c) F_3 is continuous and singular (that is, μ_3 is singular with respect to Lebesgue measure).
2. If f is an increasing function from $[a, b]$ to R , show that $\int_a^b f'(x) dx \leq f(b) - f(a)$. The inequality may be strict, as Problem 3 shows. (Note that by 2.3.9, f' exists a.e.; for integration purposes, f' may be defined arbitrarily on the exceptional set of Lebesgue measure 0.)
3. (*The Cantor function*) Let E_1, E_2, \dots be the sets removed from $[0, 1]$ to form the Cantor ternary set (see Problem 7, Section 1.4). Define functions $F_n : [0, 1] \rightarrow [0, 1]$ as follows: Let $A_1, A_2, \dots, A_{2^n-1}$ be the subintervals of $\bigcup_{i=1}^n E_i$, arranged in increasing order. For example, if $n = 3$,

$$\begin{aligned} E_1 \cup E_2 \cup E_3 &= \left(\frac{1}{27}, \frac{2}{27}\right) \cup \left(\frac{1}{9}, \frac{2}{9}\right) \cup \left(\frac{7}{27}, \frac{8}{27}\right) \cup \left(\frac{1}{3}, \frac{2}{3}\right) \\ &\quad \cup \left(\frac{19}{27}, \frac{20}{27}\right) \cup \left(\frac{7}{9}, \frac{8}{9}\right) \cup \left(\frac{25}{27}, \frac{26}{27}\right) \\ &= A_1 \cup A_2 \cup \cdots \cup A_7. \end{aligned}$$

Define

$$\begin{aligned} F_n(0) &= 0, \\ F_n(x) &= k/2^n \quad \text{if } x \in A_k, \quad k = 1, 2, \dots, 2^n - 1, \\ F_n(1) &= 1. \end{aligned}$$

Complete the specification of F_n by interpolating linearly. For $n = 2$, see Fig. 2.1; in this case,

$$\begin{aligned} E_1 \cup E_2 &= \left(\frac{1}{9}, \frac{2}{9}\right) \cup \left(\frac{1}{3}, \frac{2}{3}\right) \cup \left(\frac{7}{9}, \frac{8}{9}\right) \\ &= A_1 \cup A_2 \cup A_3. \end{aligned}$$

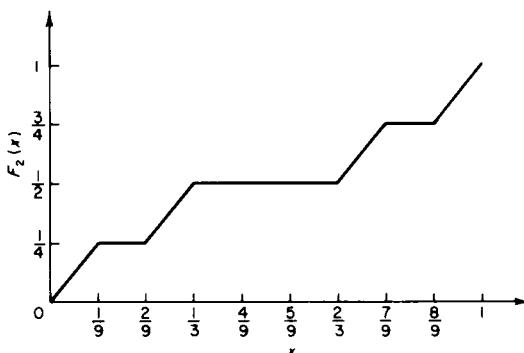


Figure 2.1. Approximation to the Cantor function.

Show that $F_n(x) \rightarrow F(x)$ for each x , where F , the Cantor function, has the following properties:

- (a) F is continuous and increasing.
- (b) $F' = 0$ almost everywhere (Lebesgue measure).
- (c) F is not absolutely continuous.

In fact

- (d) F is singular; that is, the corresponding Lebesgue–Stieltjes measure μ is singular with respect to Lebesgue measure.

4. Let f be a Lebesgue integrable real-valued function on R^k (or on an open subset of R^k). If $\mu(E) = \int_E f(x) dx$, $E \in \mathcal{B}(R^k)$, we know that $D\mu = f$ a.e. (Lebesgue measure). If $D\mu = f$ at x_0 , then if C is an open cube containing x_0 and $\text{diam } C \rightarrow 0$, we have $\mu(C)/m(C) \rightarrow f(x_0)$; that is,

$$\frac{1}{m(C)} \int_C [f(x) - f(x_0)] dx \rightarrow 0 \quad \text{as } \text{diam } C \rightarrow 0.$$

In fact, show that

$$\frac{1}{m(C)} \int_C |f(x) - f(x_0)| dx \rightarrow 0 \quad \text{as } \text{diam } C \rightarrow 0$$

for almost every x_0 . The set of favorable x_0 is called the *Lebesgue set* of f .

5. This problem relates various concepts discussed in Section 2.3. In all cases, f is a real-valued function defined on the closed bounded interval $[a, b]$. Establish the following:
 - (a) If f is continuous, f need not be of bounded variation.
 - (b) If f is continuous and increasing (hence of bounded variation), f need not be absolutely continuous.
 - (c) If f satisfies a Lipschitz condition, that is, $|f(x) - f(y)| \leq L|x - y|$ for some fixed positive number L and all $x, y \in [a, b]$, then f is absolutely continuous.
 - (d) If f' exists everywhere and is bounded, f is absolutely continuous. [It can also be shown that if f' exists everywhere and is Lebesgue integrable on $[a, b]$, then f is absolutely continuous; see Titchmarsh (1939, p. 368).]
 - (e) If f is continuous and f' exists everywhere, f need not be absolutely continuous [consider $f(x) = x^2 \sin(1/x^2)$, $0 < x \leq 1$, $f(0) = 0$].
6. The following problem considers the change of variable formula in a multiple integral. Throughout the problem, T will be a map from V onto W , where V and W are open subsets of R^k , T is assumed one-to-one, continuously differentiable, with a nonzero Jacobian. Thus T has a continuously differentiable inverse, by the inverse function theorem of

advanced calculus [see, for example, Apostol (1957, p. 144)]. It also follows from standard advanced calculus results that for all $x \in V$,

$$\frac{1}{|h|} [|T(x+h) - T(x) - A(x)h|] \rightarrow 0 \quad \text{as } h \rightarrow 0, \quad (1)$$

where $A(x)$ is the linear transformation on R^k represented by the Jacobian matrix of T , evaluated at x . [See Apostol (1957, p. 118).]

- (a) Let A be a nonsingular linear transformation on R^k , and define a measure λ on $\mathcal{B}(R^k)$ by $\lambda(E) = m(A(E))$ where m is Lebesgue measure. Show that $\lambda = c(A)m$ for some constant $c(A)$, and in fact $c(A)$ is the absolute value of the determinant of A . [Use translation-invariance of Lebesgue measure (Problem 5, Section 1.4) and the fact that any matrix can be represented as a product of matrices corresponding to elementary row operations.]

Now define a measure μ on $\mathcal{B}(V)$ by $\mu(E) = m(T(E))$. By continuity of T , if $\varepsilon > 0$, $x \in V$, and C is a sufficiently small open cube containing x , then $T(C)$ has diameter less than ε , in particular, $m(T(C)) < \infty$. It follows by a brief compactness argument that μ is a Lebesgue–Stieltjes measure on $\mathcal{B}(V)$.

Our objective is to show that μ is differentiable and $(D\mu)(x) = |J(x)|$ for every $x \in V$, where $J(x) = \det A(x)$, the Jacobian of the transformation T .

- (b) Show that it suffices to prove that if $0 \in V$ and $T(0) = 0$, then $(D\mu)(0) = |\det A(0)|$.
- (c) Show that it may be assumed without loss of generality that $A(0)$ is the identity transformation; hence $\det A(0) = 1$.

Now given $\varepsilon > 0$, choose $\alpha \in (0, \frac{1}{4})$ such that

$$1 - \varepsilon < (1 - 2\alpha)^k < (1 + 2\alpha)^k < 1 + \varepsilon.$$

Under the assumptions of (b) and (c), by Eq. (1), there is a $\delta > 0$ such that if $|x| < \delta$, then $|T(x) - x| \leq \alpha|x|/\sqrt{k}$.

- (d) If C is an open cube containing 0 with edge length β and diameter $\sqrt{k}\beta < \delta$, take C_1, C_2 as open cubes concentric with C , with edge lengths $\beta_1 = (1 - 2\alpha)\beta$ and $\beta_2 = (1 + 2\alpha)\beta$. Establish the following:

- (i) If $x \in \bar{C}$, then $T(x) \in C_2$.
- (ii) If x belongs to the boundary of C , then $T(x) \notin C_1$.
- (iii) If x is the center of C , then $T(x) \in C_1$.
- (iv) $C_1 - T(C) = C_1 - T(\bar{C})$.

Use a connectedness argument to conclude that $C_1 \subset T(C) \subset C_2$, and complete the proof that $(D\mu)(0) = 1$.

- (e) If λ is any measure on $\mathcal{B}(V)$ and $\bar{D}\lambda < \infty$ on V , show that λ is absolutely continuous with respect to Lebesgue measure. It therefore follows from Theorem 2.3.8 that

$$m(T(E)) = \int_E |J(x)| dx, \quad E \in \mathcal{B}(V).$$

[If this is false, find a compact set K and positive integers n and j such that $m(K) = 0$, $\lambda(K) > 0$, and $\lambda(C) < nm(C)$ for all open cubes C containing a point of K and having diameter less than $1/j$. Essentially, the idea is to cover K by such cubes and conclude that $\lambda(K) = 0$, a contradiction.]

- (f) If f is a real-valued Borel measurable function on W , show that

$$\int_W f(y) dy = \int_V f(T(x))|J(x)| dx$$

in the sense that if one of the two integrals exists, so does the other, and the two integrals are equal.

7. (*Fubini's differentiation theorem*) Let f_1, f_2, \dots be increasing functions from R to R , and assume that for each x , $\sum_{n=1}^{\infty} f_n(x)$ converges to a finite number $f(x)$. Show that $\sum_{n=1}^{\infty} f'_n(x) = f'(x)$ almost everywhere (Lebesgue measure).

Outline:

- (a) It suffices to restrict the domain of all functions to $[0,1]$ and to assume all functions nonnegative. Use Fatou's lemma to show that $\sum_{n=1}^{\infty} f'_n(x) \leq f'(x)$ a.e.; hence $f'_n(x) \rightarrow 0$ a.e.
- (b) Choose n_1, n_2, \dots such that $\sum_{j > n_k} f_j(1) \leq 2^{-k}$, $k = 1, 2, \dots$, and apply part (a) to the functions $g_k(x) = f(x) - \sum_{j=1}^{n_k} f_j(x) = \sum_{j > n_k} f_j(x)$.

2.4 L^p Spaces

If $(\Omega, \mathcal{F}, \mu)$ is a measure space and p is a real number with $p \geq 1$, the set of all Borel measurable functions f such that $|f|^p$ is μ -integrable has many important properties. In order to fully develop these properties, it will be convenient to work with complex-valued functions.

2.4.1 Definitions. If (Ω, \mathcal{F}) is a measurable space, a *complex-valued Borel measurable function* on (Ω, \mathcal{F}) is a mapping $f: (\Omega, \mathcal{F}) \rightarrow (R^2, \mathcal{B}(R^2))$. If $p_1(x, y) = x$ and $p_2(x, y) = y$, $x, y \in R$, we may identify $p_1 \circ f$ and $p_2 \circ f$ with the real and imaginary parts of f . If μ is a measure on \mathcal{F} , we define

$$\int_{\Omega} f d\mu = \int_{\Omega} \operatorname{Re} f d\mu + i \int_{\Omega} \operatorname{Im} f d\mu,$$

provided $\int_{\Omega} \operatorname{Re} f d\mu$ and $\int_{\Omega} \operatorname{Im} f d\mu$ are both finite. In this case we say that f is μ -integrable. Thus in working with complex-valued functions, we do not consider any cases in which integrals exist but are not finite.

The following result was established earlier for real-valued f [see 1.5.9(c)]; it is still valid in the complex case, but the proof must be modified.

2.4.2 Lemma. If f is μ -integrable,

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu.$$

PROOF. If $\int_{\Omega} f d\mu = re^{i\theta}$, $r \geq 0$, then $\int_{\Omega} e^{-i\theta} f d\mu = r = |\int_{\Omega} f d\mu|$. But if $f(\omega) = \rho(\omega)e^{i\varphi(\omega)}$ (taking $\rho \geq 0$), then

$$\begin{aligned} \int_{\Omega} e^{-i\theta} f d\mu &= \int_{\Omega} \rho e^{i(\varphi-\theta)} d\mu \\ &= \int_{\Omega} \rho \cos(\varphi - \theta) d\mu \quad \text{since } r \text{ is real} \\ &\leq \int_{\Omega} \rho d\mu = \int_{\Omega} |f| d\mu. \blacksquare \end{aligned}$$

Many other standard properties of the integral carry over to the complex case, in particular 1.5.5(b), 1.5.9(a) and (e), 1.6.1, 1.6.3, 1.6.4(b) and (c), 1.6.5, 1.6.9, 1.6.10, and 1.7.1. In almost all cases, the result is an immediate consequence of the fact that integrating a complex-valued function is equivalent to integrating the real and imaginary parts separately. Only two theorems require additional comment. To prove that h is integrable iff $|h|$ is integrable [1.6.4(b)], use the fact that $|\operatorname{Re} h|, |\operatorname{Im} h| \leq |h| \leq |\operatorname{Re} h| + |\operatorname{Im} h|$. Finally, to prove the dominated convergence theorem (1.6.9), apply the real version of the theorem to $|f_n - f|$, and note that $|f_n - f| \leq |f_n| + |f| \leq 2g$.

If $p > 0$, we define the space $L^p = L^p(\Omega, \mathcal{F}, \mu)$ as the collection of all complex-valued Borel measurable functions f such that $\int_{\Omega} |f|^p d\mu < \infty$. We set

$$\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{1/p}, \quad f \in L^p.$$

It follows that for any complex number a , $\|af\|_p = |a| \|f\|_p$, $f \in L^p$.

We are going to show that L^p forms a linear space over the complex field. The key steps in the proof are the Hölder and Minkowski inequalities, which we now develop.

2.4.3 Lemma. If $a, b, \alpha, \beta > 0$, $\alpha + \beta = 1$, then $a^\alpha b^\beta \leq \alpha a + \beta b$.

PROOF. The statement to be proved is equivalent to $-\log(\alpha a + \beta b) \leq \alpha(-\log a) + \beta(-\log b)$, which holds since $-\log$ is convex. [If g has a non-negative second derivative on the interval $I \subset R$, then g is convex on I , that is $g(\alpha x + \beta y) \leq \alpha g(x) + \beta g(y)$, $x, y \in I$, $\alpha, \beta > 0$, $\alpha + \beta = 1$. To see this, assume $x < y$ and write

$$\begin{aligned} g(\alpha x + \beta y) - \alpha g(x) - \beta g(y) &= \alpha[g(\alpha x + \beta y) - g(x)] + \beta[g(\alpha x + \beta y) - g(y)] \\ &= \alpha\beta(y - x)[g'(u) - g'(v)] \quad \text{for some } u, v \end{aligned}$$

with $x \leq u \leq \alpha x + \beta y$, $\alpha x + \beta y \leq v \leq y$. But $g'(u) - g'(v) \leq 0$ since g' is increasing on I .] ■

2.4.4 Corollary. If $c, d > 0$, $p, q > 1$, $(1/p) + (1/q) = 1$, then $cd \leq (c^p/p) + (d^q/q)$.

PROOF. In 2.4.3, let $\alpha = 1/p$, $\beta = 1/q$, $a = c^p$, $b = d^q$. ■

2.4.5 Hölder Inequality. Let $1 < p < \infty$, $1 < q < \infty$, $(1/p) + (1/q) = 1$. If $f \in L^p$ and $g \in L_q$, then $fg \in L^1$ and $\|fg\|_1 \leq \|f\|_p \|g\|_q$. (the inequality is immediate if $\|f\|_p$ or $\|g\|_q = 0$). Then

$$\frac{|f(\omega)g(\omega)|}{\|f\|_p \|g\|_q} \leq \frac{|f(\omega)|^p}{p\|f\|_p^p} + \frac{|g(\omega)|^q}{q\|g\|_q^q};$$

integrate to obtain

$$\frac{\int_{\Omega} |fg| \, d\mu}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1. \blacksquare$$

When $p = q = 2$, we obtain

$$\int_{\Omega} |fg| \, d\mu \leq \left[\int_{\Omega} |f|^2 \, d\mu \int_{\Omega} |g|^2 \, d\mu \right]^{1/2}$$

and thus, using 2.4.2, we have the *Cauchy–Schwarz inequality*: If f and $g \in L^2$, then $fg \in L^1$ and

$$\left| \int_{\Omega} f\bar{g} \, d\mu \right| \leq \left[\int_{\Omega} |f|^2 \, d\mu \int_{\Omega} |g|^2 \, d\mu \right]^{1/2},$$

where \bar{g} is the complex conjugate of g . (The reason for replacing g by \bar{g} is to make the inequality agree with the Hilbert space result to be discussed in Chapter 3.)

2.4.6 Lemma. If $a, b \geq 0, p \geq 1$, then $(a + b)^p \leq 2^{p-1}(a^p + b^p)$.

PROOF. Let $h(x) = d[(a + x)^p - 2^{p-1}(a^p + x^p)]/dx = p(a + x)^{p-1} - 2^{p-1}px^{p-1}$; since $p \geq 1$,

$$\begin{aligned} h(x) &> 0 & \text{for } a + x > 2x, & \text{that is, } x < a, \\ h(x) &= 0 & \text{at } & x = a, \\ h(x) &< 0 & \text{for } & x > a. \end{aligned}$$

The maximum therefore occurs at $x = a$; hence

$$(a + b)^p - 2^{p-1}(a^p + b^p) \leq (a + a)^p - 2^{p-1}(a^p + a^p) = 0. \blacksquare$$

2.4.7 Minkowski Inequality. If $f, g \in L^p$ ($1 \leq p < \infty$), then $f + g \in L^p$ and $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

PROOF. By 2.4.6, $|f + g|^p \leq (|f| + |g|)^p \leq 2^{p-1}(|f|^p + |g|^p)$, hence $f, g \in L^p$ implies $f + g \in L^p$. Now the inequality is clear when $p = 1$, so assume $p > 1$ and choose q such that $(1/p) + (1/q) = 1$. Then

$$|f + g|^p = |f + g| |f + g|^{p-1} \leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}. \quad (1)$$

Now $|f + g|^{p-1} \in L^q$; for

$$(p-1)q = \frac{p-1}{1/q} = \frac{p-1}{1-1/p} = p;$$

hence

$$\int_{\Omega} [|f + g|^{p-1}]^q d\mu = \int_{\Omega} |f + g|^p d\mu < \infty.$$

Since f and g belong to L^p and $|f + g|^{p-1} \in L^q$, Hölder's inequality implies that $|f| |f + g|^{p-1}$ and $|g| |f + g|^{p-1} \in L^1$, and

$$\int_{\Omega} |f| |f + g|^{p-1} d\mu \leq \|f\|_p \left[\int_{\Omega} (|f + g|^{p-1})^q d\mu \right]^{1/q} = \|f\|_p \|f + g\|_p^{p/q}, \quad (2)$$

$$\int_{\Omega} |g| |f + g|^{p-1} d\mu \leq \|g\|_p \|f + g\|_p^{p/q}. \quad (3)$$

By Eq. (1), $\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p)(\|f + g\|_p^{p/q})$. Since $p - (p/q) = 1$, the result follows. \blacksquare

By Minkowski's inequality and the fact that $\|af\|_p = |a| \|f\|_p$ for $f \in L^p$, L^p ($1 \leq p < \infty$) is a vector space over the complex field. Furthermore, there is a natural notion of distance in L^p , by virtue of the fact that $\|\cdot\|_p$ is a semi-norm.

2.4.8 Definitions and Comments. A *seminorm* on a vector space L (over the real or complex field) is a real-valued function $\| \cdot \|$ on L , with the following properties.

$$\|f\| \geq 0,$$

$$\|af\| = |a| \|f\| \quad \text{for each scalar } a;$$

consequently, if $f = 0$, then $\|f\| = 0$.

$$\|f + g\| \leq \|f\| + \|g\|$$

(f and g are arbitrary elements of L). If $\| \cdot \|$ is a seminorm with the additional property that $\|f\| = 0$ implies $f = 0$, $\| \cdot \|$ is said to be a *norm*.

Now $\| \cdot \|_p$ is a seminorm on L^p ; the first two properties follow from the definition of $\| \cdot \|_p$, and the last property is a consequence of Minkowski's inequality.

We can, in effect, change $\| \cdot \|_p$ into a norm by passing to equivalence classes as follows.

If $f, g \in L^p(\Omega, \mathcal{F}, \mu)$, define $f \sim g$ iff $f = g$ a.e. $[\mu]$. Then $\|f\|_p$ is the same for all f in a given equivalence class, by 1.6.5(b). Thus if \mathbf{L}^p is the collection of equivalence classes, \mathbf{L}^p becomes a linear space, and $\| \cdot \|_p$ is a seminorm on \mathbf{L}^p . In fact $\| \cdot \|_p$ is a norm, since $\|f\|_p = 0$ implies $f = 0$ a.e. $[\mu]$, by 1.6.6(b).

If $\| \cdot \|$ is a seminorm on a vector space, we have a natural notion of distance: $d(f, g) = \|f - g\|$. By definition of seminorm we have

$$d(f, g) \geq 0,$$

$$d(f, g) = 0 \quad \text{if } f = g,$$

$$d(f, g) = d(g, f),$$

$$d(f, h) \leq d(f, g) + d(g, h).$$

Thus d has all the properties of a metric, except that $d(f, g) = 0$ does not necessarily imply $f = g$; we call d a *pseudometric*. If $\| \cdot \|$ is a norm, d is actually a metric. [There is an asymmetry of terminology between seminorm and pseudometric, but these terms seem to be most popular, although "pseudonorm" is sometimes used, as is "semimetric."]

One of the first questions that arises in any metric space is the problem of completeness; we ask whether or not Cauchy sequences converge. We are going to show that the L^p spaces are complete. The following result will be needed; students of probability are likely to recognize it immediately, but it appears in other parts of analysis as well.

2.4.9 Chebyshev's Inequality. Let f be a nonnegative, extended real-valued, Borel measurable function on $(\Omega, \mathcal{F}, \mu)$. If $0 < p < \infty$ and $0 < \varepsilon < \infty$,

$$\mu\{\omega: f(\omega) \geq \varepsilon\} \leq \frac{1}{\varepsilon^p} \int_{\Omega} f^p d\mu.$$

The following version is often applied in probability. If g is an extended real-valued Borel measurable function on (Ω, \mathcal{F}) and P is a probability measure on \mathcal{F} , define

$$m = \int_{\Omega} g \, dP \quad (\text{assumed finite, so that } g \text{ is finite a.e. [P]}),$$

$$\sigma^2 = \int_{\Omega} (g - m)^2 \, dP.$$

If $0 < k < \infty$,

$$P\{\omega : |g(\omega) - m| \geq k\sigma\} \leq \frac{1}{k^2}.$$

This follows from the first version with $f = |g - m|$, $\varepsilon = k\sigma$, $p = 2$.

PROOF.

$$\int_{\Omega} f^p \, d\mu \geq \int_{\{\omega : f(\omega) \geq \varepsilon\}} f^p \, d\mu \geq \varepsilon^p \mu\{\omega : f(\omega) \geq \varepsilon\}. \blacksquare$$

One more auxiliary result will be needed.

2.4.10 Lemma. If $g_1, g_2, \dots \in L^p$ ($p > 0$) and $\|g_k - g_{k+1}\|_p < (\frac{1}{4})^k$, $k = 1, 2, \dots$, then $\{g_k\}$ converges a.e.

PROOF. Let $A_k = \{\omega : |g_k(\omega) - g_{k+1}(\omega)| \geq 2^{-k}\}$. Then by 2.4.9,

$$\mu(A_k) \leq 2^{kp} \|g_k - g_{k+1}\|_p^p < 2^{-kp}.$$

By 2.2.4, $\mu(\limsup_n A_n) = 0$. But if $\omega \notin \limsup_n A_n$, then $|g_k(\omega) - g_{k+1}(\omega)| < 2^{-k}$ for large k , so $\{g_k(\omega)\}$ is a Cauchy sequence of complex numbers, and therefore converges. \blacksquare

Now, the main result:

2.4.11 Completeness of L^p , $1 \leq p < \infty$. If f_1, f_2, \dots form a Cauchy sequence in L^p , that is, $\|f_n - f_m\|_p \rightarrow 0$ as $n, m \rightarrow \infty$, there is an $f \in L^p$ such that $\|f_n - f\|_p \rightarrow 0$.

PROOF. Let n_1 be such that $\|f_n - f_m\|_p < \frac{1}{4}$ for $n, m \geq n_1$, and let $g_1 = f_{n_1}$. In general, having chosen g_1, \dots, g_k and n_1, \dots, n_k , let $n_{k+1} > n_k$ be such that $\|f_n - f_m\|_p < (\frac{1}{4})^{k+1}$ for $n, m \geq n_{k+1}$, and let $g_{k+1} = f_{n_{k+1}}$. By 2.4.10, g_k converges a.e. to a limit function f .

Given $\varepsilon > 0$, choose N such that $\|f_n - f_m\|_p^p < \varepsilon$ for $n, m \geq N$. Fix $n \geq N$ and let $m \rightarrow \infty$ through values in the subsequence, that is, let $m = n_k$, $k \rightarrow \infty$. Then

$$\begin{aligned}
 \varepsilon &\geq \liminf_{k \rightarrow \infty} \|f_n - f_{n_k}\|_p^p = \liminf_{k \rightarrow \infty} \int_{\Omega} |f_n - f_{n_k}|^p d\mu \\
 &\geq \int_{\Omega} \liminf_{k \rightarrow \infty} |f_n - g_k|^p d\mu \quad \text{by Fatou's lemma} \\
 &= \|f_n - f\|_p^p.
 \end{aligned}$$

Thus $\|f_n - f\|_p \rightarrow 0$. Since $f = f - f_n + f_n$, we have $f \in L^p$. ■

2.4.12 Examples and Comments. Let Ω be the positive integers; take \mathcal{F} as all subsets of Ω , and let μ be counting measure. A real-valued function on Ω may be represented as a sequence of real numbers; we write $f = \{a_n, n = 1, 2, \dots\}$. An integral on this space is really a sum [see Problem 1(a)]:

$$\int_{\Omega} f d\mu = \sum_{n=1}^{\infty} a_n,$$

where the series is interpreted as $\sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-$ if this is not of the form $+\infty - \infty$ (if it is, the integral does not exist). Thus the following cases occur:

(1) $\sum_{n=1}^{\infty} a_n^+ = \infty, \sum_{n=1}^{\infty} a_n^- < \infty$. The series diverges to ∞ and the integral is ∞ .

(2) $\sum_{n=1}^{\infty} a_n^+ < \infty, \sum_{n=1}^{\infty} a_n^- = \infty$. The series diverges to $-\infty$ and the integral is $-\infty$.

(3) $\sum_{n=1}^{\infty} a_n^+ < \infty, \sum_{n=1}^{\infty} a_n^- < \infty$. The series is absolutely convergent and the integral equals the sum of the series.

(4) $\sum_{n=1}^{\infty} a_n^+ = \infty, \sum_{n=1}^{\infty} a_n^- = \infty$. The series is not absolutely convergent; it may or may not converge conditionally. Whether it does or not, the integral does not exist. Thus when summation is considered from the point of view of Lebesgue integration theory, series that converge conditionally but not absolutely are ignored.

If μ is changed so that $\mu\{n\}$ is a nonnegative number p_n , not necessarily 1 as in the case of counting measure, the same analysis shows that

$$\int_{\Omega} f d\mu = \sum_{n=1}^{\infty} p_n a_n,$$

where the series is interpreted as $\sum_{n=1}^{\infty} p_n a_n^+ - \sum_{n=1}^{\infty} p_n a_n^-$.

If $f = \{a_n, n = 1, 2, \dots\}$ is a sequence of complex numbers and μ is counting measure,

$$\int_{\Omega} f d\mu = \sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \operatorname{Re} a_n + i \sum_{n=1}^{\infty} \operatorname{Im} a_n;$$

the integral is defined provided $\sum_{n=1}^{\infty} |a_n| < \infty$.

Now let Ω be an arbitrary set, and take \mathcal{F} as all subsets of Ω and μ as counting measure. If $f = (f(\alpha), \alpha \in \Omega)$ is a nonnegative real-valued function on Ω , then [Problem 1(b)]

$$\int_{\Omega} f \, d\mu = \sum_{\alpha} f(\alpha), \quad (1)$$

where the series is defined as $\sup\{\sum_{\alpha \in F} f(\alpha) : F \subset \Omega, F \text{ finite}\}$. If $f(\alpha) > 0$ for uncountably many α , then for some $\delta > 0$ we have $f(\alpha) \geq \delta$ for infinitely many α , so that $\sum_{\alpha} f(\alpha) = \infty$.

If the nonnegativity hypothesis is dropped, we apply the above results to f^+ and f^- to again obtain Eq. (1), where the series is interpreted as $\sum_{\alpha} f^+(\alpha) - \sum_{\alpha} f^-(\alpha)$. If f is complex-valued, Eq. (1) still applies, with the series interpreted as $\sum_{\alpha} \operatorname{Re} f(\alpha) + i \sum_{\alpha} \operatorname{Im} f(\alpha)$. The integral is defined provided $\sum_{\alpha} |f(\alpha)| < \infty$.

The space $L^p(\Omega, \mathcal{F}, \mu)$ will be denoted by $l^p(\Omega)$; it consists of all complex-valued functions $(f(\alpha), \alpha \in \Omega)$ such that $f(\alpha) = 0$ for all but countably many α , and

$$\|f\|_p^p = \sum_{\alpha} |f(\alpha)|^p < \infty.$$

If Ω is the set of positive integers, the space $l^p(\Omega)$ will be denoted simply by l^p ; it consists of all sequences $f = \{a_n\}$ of complex numbers such that

$$\|f\|_p^p = \sum_{n=1}^{\infty} |a_n|^p < \infty.$$

It will be useful to state the Hölder and Minkowski inequalities for sums. If $f \in l^p(\Omega)$ and $g \in l^q(\Omega)$, where $1 < p < \infty$, $1 < q < \infty$, $(1/p) + (1/q) = 1$, then $fg \in l^1(\Omega)$ and

$$\sum_{\alpha} |f(\alpha)g(\alpha)| \leq \left(\sum_{\alpha} |f(\alpha)|^p \right)^{1/p} \left(\sum_{\alpha} |g(\alpha)|^q \right)^{1/q}.$$

If $f, g \in l^p(\Omega)$, $1 \leq p < \infty$, then $f + g \in l^p(\Omega)$ and

$$\left(\sum_{\alpha} |f(\alpha) + g(\alpha)|^p \right)^{1/p} \leq \left(\sum_{\alpha} |f(\alpha)|^p \right)^{1/p} + \left(\sum_{\alpha} |g(\alpha)|^p \right)^{1/p}.$$

As in 2.4.5, we obtain the Cauchy-Schwarz inequality for sums from the Hölder inequality. If $f, g \in l^2(\Omega)$, then $fg \in l^1(\Omega)$ and

$$\left| \sum_{\alpha} f(\alpha) \overline{g(\alpha)} \right| \leq \left(\sum_{\alpha} |f(\alpha)|^2 \right)^{1/2} \left(\sum_{\alpha} |g(\alpha)|^2 \right)^{1/2}.$$

If in the above discussion we replace Ω by $\{1, 2, \dots, n\}$, all convergence difficulties are eliminated, and all the spaces $l^p(\Omega)$ coincide with C^n .

If $0 < p < 1$, $\|\cdot\|_p$ is not a seminorm on $L^p(\Omega, \mathcal{F}, \mu)$. For let A and B be disjoint sets with $a = \mu(A)$ and $b = \mu(B)$ assumed finite and positive. If $f = I_A$, $g = I_B$, then

$$\begin{aligned}\|f + g\|_p &= \left(\int_{\Omega} |f + g|^p d\mu \right)^{1/p} = \left(\int_{\Omega} (I_A + I_B) d\mu \right)^{1/p} = (a + b)^{1/p}, \\ \|f\|_p &= a^{1/p}, \quad \|g\|_p = b^{1/p}.\end{aligned}$$

But $(a + b)^{1/p} > a^{1/p} + b^{1/p}$ if $a, b > 0$, $0 < p < 1$, since $(a + x)^r - a^r - x^r$ is strictly increasing for $r > 1$, and has the value 0 when $x = 0$. Thus the triangle inequality fails. We can, however, describe convergence in L^p , $0 < p < 1$, in the following way. We use the inequality

$$(a + b)^p \leq a^p + b^p, \quad a, b \geq 0, \quad 0 < p < 1,$$

which is proved by considering $(a + x)^p - a^p - x^p$. It follows that

$$\int_{\Omega} |f + g|^p d\mu \leq \int_{\Omega} |f|^p d\mu + \int_{\Omega} |g|^p d\mu, \quad f, g \in L^p, \quad (2)$$

and therefore $d(f, g) = \int_{\Omega} |f - g|^p d\mu$ defines a pseudometric on L^p . In fact the pseudometric is complete (every Cauchy sequence converges); for Eq. (2) implies that if $f, g \in L^p$, then $f + g \in L^p$, so that the proof of 2.4.11 goes through.

If Ω is an interval of reals, \mathcal{F} is the class of Borel sets of Ω , and μ is Lebesgue measure, the space $L^p(\Omega, \mathcal{F}, \mu)$ will be denoted by $L^p(\Omega)$. Thus, for example, $L^p[a, b]$ is the set of all complex-valued Borel measurable functions f on $[a, b]$ such that

$$\|f\|_p^p = \int_a^b |f(x)|^p dx < \infty.$$

If f is a complex valued Borel measurable function on $(\Omega, \mathcal{F}, \mu)$ and $f_1, f_2, \dots \in L^p(\Omega, \mathcal{F}, \mu)$, we say that the sequence $\{f_n\}$ converges to f in L^p iff $\|f_n - f\|_p \rightarrow 0$, that is, iff $\int_{\Omega} |f_n - f|^p d\mu \rightarrow 0$ as $n \rightarrow \infty$. We use the notation $f_n \xrightarrow{L^p} f$. In Section 2.5, we shall compare various types of convergence of sequences of measurable functions. We show now that any $f \in L^p$ is an L^p -limit of simple functions.

2.4.13 Theorem. Let $f \in L^p$, $0 < p < \infty$. If $\varepsilon > 0$, there is a simple function $g \in L^p$ such that $\|f - g\|_p < \varepsilon$; g can be chosen to be finite-valued and to satisfy $|g| \leq |f|$. Thus the finite-valued simple functions are dense in L^p .

PROOF. This follows from 1.5.5(b) and 1.6.10. ■

If we specialize to functions on R^n and Lebesgue–Stieltjes measures, we may obtain another basic approximation theorem.

2.4.14 Theorem. Let $f \in L^p(\Omega, \mathcal{F}, \mu)$, $0 < p < \infty$, where $\Omega = R^n$, $\mathcal{F} = \mathcal{B}(R^n)$, and μ is a Lebesgue–Stieltjes measure. If $\varepsilon > 0$, there is a continuous function $g \in L^p(\Omega, \mathcal{F}, \mu)$ such that $\|f - g\|_p < \varepsilon$; furthermore, g can be chosen so that $\sup |g| \leq \sup |f|$. Thus the continuous functions are dense in L^p .

PROOF. By 2.4.13, it suffices to show that an indicator I_A in L^p can be approximated in the L^p sense by a continuous function with absolute value at most 1. Now $I_A \in L^p$ means that $\mu(A) < \infty$; hence by Problem 12, Section 1.4, there is a closed set $C \subset A$ and an open set $V \supset A$ such that $\mu(V - C) < \varepsilon^p 2^{-p}$. Let g be a continuous map of Ω into $[0, 1]$ with $g = 1$ on C and $g = 0$ on V^c (g exists by Urysohn's lemma). Then

$$\int_{\Omega} |I_A - g|^p d\mu = \int_{\{I_A \neq g\}} |I_A - g|^p d\mu.$$

But $\{I_A \neq g\} \subset V - C$ and $|I_A - g| \leq 2$; hence

$$\|I_A - g\|_p^p \leq 2^p \mu(V - C) < \varepsilon^p.$$

Since $g = g - I_A + I_A$, we have $g \in L^p$. ■

Theorem 2.4.14 shows that the continuous functions in L^p do not form a closed subset, for if they did, every function in L^p would be continuous. Equivalently, the continuous functions in L^p are not complete, in other words, there are Cauchy sequences of continuous functions in L^p that do not converge in L^p to a continuous limit. Explicit examples may be given without making use of 2.4.14; see Problem 2.

2.4.15 The Space L^∞ . If we wish to define L^p spaces for $p = \infty$, we must proceed differently. We define the *essential supremum* of the real-valued Borel measurable function g on $(\Omega, \mathcal{F}, \mu)$ as

$$\text{ess sup } g = \inf\{c \in \bar{\mathbb{R}} : \mu\{\omega : g(\omega) > c\} = 0\}$$

that is, the smallest number c such that $g \leq c$ a.e. [μ].

If f is a complex-valued Borel measurable function on $(\Omega, \mathcal{F}, \mu)$, we define

$$\|f\|_\infty = \text{ess sup } |f|.$$

The space $L^\infty(\Omega, \mathcal{F}, \mu)$ is the collection of all f such that $\|f\|_\infty < \infty$. Thus $f \in L^\infty$ iff f is essentially bounded, that is, bounded outside a set of measure 0.

Now $|f + g| \leq |f| + |g| \leq \|f\|_\infty + \|g\|_\infty$ a.e.; hence

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.$$

In particular, $f, g \in L^\infty$ implies $f + g \in L^\infty$. The other properties of a seminorm are easily checked. Thus L^∞ is a vector space over the complex field, $\|\cdot\|_\infty$ is a seminorm on L^∞ , and becomes a norm if we pass to equivalence classes as before.

If $f, f_1, f_2, \dots \in L^\infty$ and $\|f_n - f\|_\infty \rightarrow 0$, we write $f_n \xrightarrow{L^\infty} f$; we claim that:

$\|f_n - f\|_\infty \rightarrow 0$ iff there is a set $A \in \mathcal{F}$ with $\mu(A) = 0$ such that $f_n \rightarrow f$ uniformly on A^c .

For, assume $\|f_n - f\|_\infty \rightarrow 0$. Given a positive integer m , $\|f_n - f\|_\infty \leq 1/m$ for sufficiently large n ; hence $|f_n(\omega) - f(\omega)| \leq 1/m$ for almost every ω , say for $\omega \notin A_m$, where $\mu(A_m) = 0$. If $A = \bigcup_{m=1}^{\infty} A_m$, then $\mu(A) = 0$ and $f_n \rightarrow f$ uniformly on A^c . Conversely, assume $\mu(A) = 0$ and $f_n \rightarrow f$ uniformly on A^c . Given $\varepsilon > 0$, $|f_n - f| \leq \varepsilon$ on A^c for sufficiently large n , so that $|f_n - f| \leq \varepsilon$ a.e. Thus $\|f_n - f\|_\infty \leq \varepsilon$ for large enough n , and the result follows.

An identical argument shows that $\{f_n\}$ is a Cauchy sequence in L^∞ ($\|f_n - f_m\|_\infty \rightarrow 0$ as $n, m \rightarrow \infty$) iff there is a set $A \in \mathcal{F}$ with $\mu(A) = 0$ and $f_n - f_m \rightarrow 0$ uniformly on A^c .

It is immediate that the Hölder inequality still holds when $p = 1, q = \infty$, and we have shown above that the Minkowski inequality holds when $p = \infty$.

To show that L^∞ is complete, let $\{f_n\}$ be a Cauchy sequence in L^∞ , and let A be a set of measure 0 such that $f_n(\omega) - f_m(\omega) \rightarrow 0$ uniformly for $\omega \in A^c$. But then $f_n(\omega)$ converges to a limit $f(\omega)$ for each $\omega \in A^c$, and the convergence is uniform on A^c . If we define $f(\omega) = 0$ for $\omega \in A$, we have $f \in L^\infty$ and $f_n \xrightarrow{L^\infty} f$.

Theorem 2.4.13 holds also when $p = \infty$. For if f is a function in L^∞ , the standard approximating sequence $\{f_n\}$ of simple functions (see 1.5.5) converges to f uniformly, outside a set of measure 0. However, Theorem 2.4.14 fails when $p = \infty$ (see Problem 12).

If Ω is an arbitrary set, \mathcal{F} consists of all subsets of Ω , and μ is counting measure, then $L^\infty(\Omega, \mathcal{F}, \mu)$ is the set of all bounded complex-valued functions $f = (f(\alpha), \alpha \in \Omega)$, denoted by $l^\infty(\Omega)$. The essential supremum is simply the supremum; in other words, $\|f\|_\infty = \sup\{|f(\alpha)| : \alpha \in \Omega\}$. If Ω is the set of positive integers, $l^\infty(\Omega)$ is the space of bounded sequences of complex numbers, denoted simply by l^∞ .

Problems

1. (a) If $f = \{a_n, n = 1, 2, \dots\}$, the a_n are real or complex numbers, and μ is counting measure on subsets of the positive integers, show that $\int_{\Omega} f d\mu = \sum_{n=1}^{\infty} a_n$, where the sum is interpreted as in 2.4.12.
 (b) If $f = (f(\alpha), \alpha \in \Omega)$ is a real- or complex-valued function on the arbitrary set Ω , and μ is counting measure on subsets of Ω , show that $\int_{\Omega} f d\mu = \sum_{\alpha} f(\alpha)$, where the sum is interpreted as in 2.4.12.
2. Give an example of functions f, f_1, f_2, \dots from R to $[0, 1]$ such that
 - (a) each f_n is continuous on R ,
 - (b) $f_n(x)$ converges to $f(x)$ for all x , $\int_{-\infty}^{\infty} |f_n(x) - f(x)|^p dx \rightarrow 0$ for every $p \in (0, \infty)$, and
 - (c) f is discontinuous at some point of R .
3. For each $n = 1, 2, \dots$, let $f_n = \{a_1^{(n)}, a_2^{(n)}, \dots\}$ be a sequence of complex numbers.

- (a) If the $a_k^{(n)}$ are real and $0 \leq a_k^{(n)} \leq a_k^{(n+1)}$ for all k and n , show that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_k^{(n)} = \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} a_k^{(n)}.$$

Show that the same conclusion holds if the $a_k^{(n)}$ are complex and $|a_k^{(n)}| \leq b_k$ for all k and n , where $\sum_{k=1}^{\infty} b_k < \infty$.

- (b) If the $a_k^{(n)}$ are real and nonnegative, show that

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_k^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_k^{(n)}.$$

- (c) If the $a_k^{(n)}$ are complex and $\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} |a_k^{(n)}| < \infty$, show that $\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_k^{(n)}$ and $\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_k^{(n)}$ both converge to the same finite number.

4. Show that there is equality in the Hölder inequality iff $|f|^p$ and $|g|^q$ are linearly dependent, that is, iff $A|f|^p = B|g|^q$ a.e. for some constants A and B , not both 0.
5. If f is a complex-valued μ -integrable function, show that $|\int_{\Omega} f d\mu| = \int_{\Omega} |f| d\mu$ iff $\arg f$ is a.e. constant on $\{\omega : f(\omega) \neq 0\}$.
6. Show that equality holds in the Cauchy–Schwarz inequality iff f and g are linearly dependent.
7. (a) If $1 < p < \infty$, show that equality holds in the Minkowski inequality iff $Af = Bg$ a.e. for some nonnegative constants A and B , not both 0.
(b) What are the conditions for equality if $p = 1$?
8. If $1 \leq r < s < \infty$, and $f \in L^s(\Omega, \mathcal{F}, \mu)$, μ finite, show that $\|f\|_r \leq k \|f\|_s$ for some finite positive constant k . Thus $L^s \subset L^r$ and L^s convergence implies L^r convergence. (We may take $k = 1$ if μ is a probability measure.) Note that finiteness of μ is essential here; if μ is Lebesgue measure on $\mathcal{B}(R)$ and $f(x) = 1/x$ for $x \geq 1$, $f(x) = 0$ for $x < 1$, then $f \in L^2$ but $f \notin L^r$.
9. If μ is finite, show that $\|f\|_p \rightarrow \|f\|_{\infty}$ as $p \rightarrow \infty$. Give an example to show that this fails if $\mu(\Omega) = \infty$.
10. (*Radon–Nikodym theorem, complex case*) If μ is a σ -finite (nonnegative, real) measure, λ a complex measure on (Ω, \mathcal{F}) , and $\lambda \ll \mu$, show that there is a complex-valued μ -integrable function g such that $\lambda(A) = \int_A g d\mu$ for all $A \in \mathcal{F}$. If h is another such function, $g = h$ a.e.
Show also that the Lebesgue decomposition theorem holds if λ is a complex measure and μ is a σ -finite measure. (See Problem 6, Section 2.2, for properties of complex measures.)
11. (a) Let f be a complex-valued μ -integrable function, where μ is a nonnegative real measure. If S is a closed set of complex numbers

and $[1/\mu(E)] \int_E f d\mu \in S$ for all measurable sets E such that $\mu(E) > 0$, show that $f(\omega) \in S$ for almost every ω . [If D is a closed disk with center at z and radius r , and $D \subset S^c$, take $E = f^{-1}(D)$. Show that $|\int_E (f - z) d\mu| \leq r\mu(E)$, and conclude that $\mu(E) = 0$.]

- (b) If λ is a complex measure, then $\lambda \ll |\lambda|$ by definition of $|\lambda|$; hence by the Radon–Nikodym theorem, there is a $|\lambda|$ -integrable complex-valued function h such that $\lambda(E) = \int_E h d|\lambda|$ for all $E \in \mathcal{F}$. Show that $|h| = 1$ a.e. [$|\lambda|$]. [Let $A_r = \{\omega : |h(\omega)| < r\}$, $0 < r < 1$, and use the definition of $|\lambda|$ to show that $|h| \geq 1$ a.e. Use part (a) to show $|h| \leq 1$ a.e.]
 - (c) Let μ be a nonnegative real measure, g a complex-valued μ -integrable function, and $\lambda(E) = \int_E g d\mu$, $E \in \mathcal{F}$. If $h = d\lambda/d|\lambda|$ as in part (b), show that $|\lambda|(E) = \int_E \bar{h}g d\mu$. (Intuitively, $\bar{h}g d\mu = \bar{h} d\lambda = \bar{h}h d|\lambda| = |h|^2 d|\lambda| = d|\lambda|$. Formally, show that $\int_{\Omega} f h d|\lambda| = \int_{\Omega} fg d\mu$ if f is a bounded, complex-valued, Borel measurable function, and set $f = \bar{h}I_E$.)
 - (d) Under the hypothesis of (c), show that
- $$|\lambda|(E) = \int_E |g| d\mu \quad \text{for all } E \in \mathcal{F}.$$
12. Give an example of a bounded real-valued function f on R such that there is no sequence of continuous functions f_n such that $\|f - f_n\|_{\infty} \rightarrow 0$. Thus the continuous functions are not dense in $L^{\infty}(R)$.

2.5 Convergence of Sequences of Measurable Functions

In the previous section we introduced the notion of L^p convergence; we are also familiar with convergence almost everywhere. We now consider other types of convergence and make comparisons.

Let f, f_1, f_2, \dots be complex-valued Borel measurable functions on $(\Omega, \mathcal{F}, \mu)$. We say that $f_n \rightarrow f$ in measure (or in μ -measure if we wish to emphasize the dependence on μ) iff for every $\varepsilon > 0$, $\mu\{\omega : |f_n(\omega) - f(\omega)| \geq \varepsilon\} \rightarrow 0$ as $n \rightarrow \infty$. (Notation: $f_n \xrightarrow{\mu} f$.) When μ is a probability measure, the convergence is called *convergence in probability*.

The first result shows that L^p convergence is stronger than convergence in measure.

2.5.1 Theorem. If $f, f_1, f_2, \dots \in L^p$ ($0 < p < \infty$), then $f_n \xrightarrow{L^p} f$ implies $f_n \xrightarrow{\mu} f$.

PROOF. Apply Chebyshev's inequality (2.4.9) to $|f_n - f|$. ■

The same argument shows that if $\{f_n\}$ is a Cauchy sequence in L^p , then $\{f_n\}$ is *Cauchy in measure*, that is, given $\varepsilon > 0$, $\mu\{\omega: |f_n(\omega) - f_m(\omega)| \geq \varepsilon\} \rightarrow 0$ as $n, m \rightarrow \infty$.

If f, f_1, f_2, \dots are complex-valued Borel measurable functions on $(\Omega, \mathcal{F}, \mu)$, we say that $f_n \rightarrow f$ *almost uniformly* iff, given $\varepsilon > 0$, there is a set $A \in \mathcal{F}$ such that $\mu(A) < \varepsilon$ and $f_n \rightarrow f$ uniformly on A^c .

Almost uniform convergence is stronger than both a.e. convergence and convergence in measure, as we now prove.

2.5.2 Theorem. If $f_n \rightarrow f$ almost uniformly, then $f_n \rightarrow f$ in measure and almost everywhere.

PROOF. If $\varepsilon > 0$, let $f_n \rightarrow f$ uniformly on A^c , with $\mu(A) < \varepsilon$. If $\delta > 0$, then eventually $|f_n - f| < \delta$ on A^c , so $\{|f_n - f| \geq \delta\} \subset A$. Thus $\mu\{|f_n - f| \geq \delta\} \leq \mu(A) < \varepsilon$, proving convergence in measure.

To prove almost everywhere convergence, choose, for each positive integer k , a set A_k with $\mu(A_k) < 1/k$ and $f_n \rightarrow f$ uniformly on A_k^c . If $B = \bigcup_{k=1}^{\infty} A_k^c$, then $f_n \rightarrow f$ on B and $\mu(B^c) = \mu(\bigcap_{k=1}^{\infty} A_k) \leq \mu(A_k) \rightarrow 0$ as $k \rightarrow \infty$. Thus $\mu(B^c) = 0$ and the result follows. ■

The converse to 2.5.2 does not hold in general, as we shall see in 2.5.6(c), but we do have the following result.

2.5.3 Theorem. If $\{f_n\}$ is convergent in measure, there is a subsequence converging almost uniformly (in particular, a.e. and in measure) to the same limit function.

PROOF. First note that $\{f_n\}$ is Cauchy in measure, because if $|f_n - f_m| \geq \varepsilon$, then either $|f_n - f| \geq \varepsilon/2$ or $|f - f_m| \geq \varepsilon/2$. Thus

$$\mu\{|f_n - f_m| \geq \varepsilon\} \leq \mu\left\{|f_n - f| \geq \frac{\varepsilon}{2}\right\} + \mu\left\{|f - f_m| \geq \frac{\varepsilon}{2}\right\} \rightarrow 0 \quad \text{as } n, m \rightarrow \infty.$$

Now for each positive integer k , choose a positive integer N_k such that $N_{k+1} > N_k$ for all k and

$$\mu\{\omega: |f_n(\omega) - f_m(\omega)| \geq 2^{-k}\} \leq 2^{-k} \quad \text{for } n, m \geq N_k.$$

Pick integers $n_k \geq N_k$, $k = 1, 2, \dots$; then if $g_k = f_{n_k}$,

$$\mu\{\omega: |g_k(\omega) - g_{k+1}(\omega)| \geq 2^{-k}\} \leq 2^{-k}.$$

Let $A_k = \{|g_k - g_{k+1}| \geq 2^{-k}\}$, $A = \limsup_k A_k$. Then $\mu(A) = 0$ by 2.2.4; but if $\omega \notin A$, then $\omega \in A_k$ for only finitely many k ; hence $|g_k(\omega) - g_{k+1}(\omega)| < 2^{-k}$

for large k , and it follows that $g_k(\omega)$ converges to a limit $g(\omega)$. Since $\mu(A) = 0$ we have $g_k \rightarrow g$ a.e.

If $B_r = \bigcup_{k=r}^{\infty} A_k$, then $\mu(B_r) \leq \sum_{k=r}^{\infty} \mu(A_k) < \varepsilon$ for large r . If $\omega \notin B_r$, then $|g_k(\omega) - g_{k+1}(\omega)| < 2^{-k}$, $k = r, r+1, r+2, \dots$. By the Weierstrass M-test, $g_k \rightarrow g$ uniformly on B_r , which proves almost uniform convergence.

Now by hypothesis, we have $f_n \xrightarrow{\mu} f$ for some f , hence $f_{n_k} \xrightarrow{\mu} f$. But by 2.5.2, $f_{n_k} \xrightarrow{\mu} g$ as well, hence $f = g$ a.e. (see Problem 1). Thus f_{n_k} converges almost uniformly to f , completing the proof. ■

There is a partial converse to 2.5.2, but before discussing this it will be convenient to look at a condition equivalent to a.e. convergence:

2.5.4 Lemma. If μ is finite, then $f_n \rightarrow f$ a.e. iff for every $\delta > 0$,

$$\mu\left(\bigcup_{k=n}^{\infty} \{\omega : |f_k(\omega) - f(\omega)| \geq \delta\}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. Let $B_{n\delta} = \{\omega : |f_n(\omega) - f(\omega)| \geq \delta\}$, $B_\delta = \limsup_n B_{n\delta} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_{k\delta}$. Now $\bigcup_{k=n}^{\infty} B_{k\delta} \downarrow B_\delta$; hence $\mu\left(\bigcup_{k=n}^{\infty} B_{k\delta}\right) \rightarrow \mu(B_\delta)$ as $n \rightarrow \infty$ by 1.2.7(b). Now

$$\begin{aligned} \{\omega : f_n(\omega) \rightarrow f(\omega)\} &= \bigcup_{\delta > 0} B_\delta \\ &= \bigcup_{m=1}^{\infty} B_{1/m} \quad \text{since } B_{\delta_1} \subset B_{\delta_2} \text{ for } \delta_1 > \delta_2. \end{aligned}$$

Therefore,

$$\begin{aligned} f_n \rightarrow f \quad \text{a.e.} \quad \text{iff} \quad \mu(B_\delta) = 0 \quad \text{for all } \delta > 0 \\ \text{iff} \quad \mu\left(\bigcup_{k=n}^{\infty} B_{k\delta}\right) \rightarrow 0 \quad \text{for all } \delta > 0. \blacksquare \end{aligned}$$

2.5.5 Egoroff's Theorem. If μ is finite and $f_n \rightarrow f$ a.e., then $f_n \rightarrow f$ almost uniformly. Hence by 2.5.2, if μ is finite, then almost everywhere convergence implies convergence in measure.

PROOF. It follows from 2.5.4 that given $\varepsilon > 0$ and a positive integer j , for sufficiently large $n = n(j)$, the set $A_j = \bigcup_{k=n(j)}^{\infty} \{|f_k - f| \geq 1/j\}$ has measure less than $\varepsilon/2^j$. If $A = \bigcup_{j=1}^{\infty} A_j$, then $\mu(A) \leq \sum_{j=1}^{\infty} \mu(A_j) < \varepsilon$. Also, if $\delta > 0$ and j is chosen so that $1/j < \delta$, we have, for any $k \geq n(j)$ and $\omega \in A^c$ (hence $\omega \notin A_j$), $|f_k(\omega) - f(\omega)| < 1/j < \delta$. Thus $f_n \rightarrow f$ uniformly on A^c . ■

We now give some examples to illustrate the relations between the various types of convergence. In all cases, we assume that \mathcal{F} is the class of Borel sets and μ is Lebesgue measure.

2.5.6 Examples. (a) Let $\Omega = [0, 1]$ and define

$$f_n(x) = \begin{cases} e^n & \text{if } 0 \leq x \leq \frac{1}{n}, \\ 0 & \text{elsewhere.} \end{cases}$$

Then $f_n \rightarrow 0$ a.e., hence in measure by 2.5.5. But for each $p \in (0, \infty]$, f_n fails to converge in L^p . For if $p < \infty$,

$$\|f_n\|_p^p = \int_0^1 |f_n(x)|^p dx = \frac{1}{n} e^{np} \rightarrow \infty,$$

and

$$\|f_n\|_\infty = e^n \rightarrow \infty.$$

(b) Let $\Omega = R$, and define

$$f_n(x) = \begin{cases} \frac{1}{n} & \text{if } 0 \leq x \leq e^n, \\ 0 & \text{elsewhere.} \end{cases}$$

Then $f_n \rightarrow 0$ uniformly on R , so that $f_n \xrightarrow{L^\infty} 0$. It follows quickly that $f_n \rightarrow 0$ a.e. and in measure. But for each $p \in (0, \infty)$, f_n fails to converge in L^p , since $\|f_n\|_p^p = n^{-p} e^n \rightarrow \infty$.

(c) Let $\Omega = [0, \infty)$ and define

$$f_n(x) = \begin{cases} 1 & \text{if } n \leq x \leq n + \frac{1}{n}, \\ 0 & \text{elsewhere.} \end{cases}$$

Then $f_n \rightarrow 0$ a.e. and in measure (as well as in L^p , $0 < p < \infty$), but does not converge almost uniformly. For, if $f_n \rightarrow 0$ uniformly on A and $\mu(A^c) < \varepsilon$, then eventually $f_n < 1$ on A ; hence if $A_n = [n, n + (1/n)]$ we have $A \cap \bigcup_{k \geq n} A_k = \emptyset$ for sufficiently large n . Therefore, $A^c \supset \bigcup_{k \geq n} A_k$, and consequently $\mu(A^c) \geq \sum_{k=n}^{\infty} \mu(A_k) = \infty$, a contradiction.

(d) Let $\Omega = [0, 1]$, and define

$$f_{nm}(x) = \begin{cases} 1 & \text{if } \frac{m-1}{n} < x \leq \frac{m}{n}, \quad m = 1, \dots, n, \quad n = 1, 2, \dots, \\ 0 & \text{elsewhere.} \end{cases}$$

Then $\|f_{nm}\|_p^p = 1/n \rightarrow 0$, so for each $p \in (0, \infty)$, the sequence $f_{11}, f_{21}, f_{22}, f_{31}, f_{32}, f_{33}, \dots$ converges to 0 in L^p (hence converges in measure by 2.5.1). But the sequence does not converge a.e., hence by 2.5.2, does not converge almost uniformly. To see this, observe that for any $x \neq 0$, the sequence $\{f_{nm}(x)\}$ has infinitely many zeros and infinitely many ones. Thus the set on which f_{nm} converges has measure 0. Also, f_{nm} does not converge in L^∞ , for if $f_{nm} \xrightarrow{L^\infty} f$, then $f_{nm} \xrightarrow{\mu} f$, hence $f = 0$ a.e. (see Problem 1). But $\|f_{nm}\|_\infty \equiv 1$, a contradiction.

Problems

1. If f_n converges to both f and g in measure, show that $f = g$ a.e.
2. Show that a sequence is Cauchy in measure iff it is convergent in measure.
3. (a) If μ is finite, show that L^∞ convergence implies L^p convergence for all $p \in (0, \infty)$.
- (b) Show that any real-valued function in $L^p[a, b]$, $-\infty < a < b < \infty$, $0 < p < \infty$, can be approximated in L^p by a polynomial, in fact by a polynomial with rational coefficients.
4. If μ is finite, show that $\{f_n\}$ is Cauchy a.e. (for almost every ω , $\{f_n(\omega)\}$ is a Cauchy sequence) iff for every $\delta > 0$,

$$\mu\left(\bigcup_{j, k=n}^{\infty} \{\omega : |f_j(\omega) - f_k(\omega)| \geq \delta\}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

5. (*Extension of the dominated convergence theorem*) If $|f_n| \leq g$ for all $n = 1, 2, \dots$, where g is μ -integrable, and $f_n \xrightarrow{\mu} f$, show that f is μ -integrable and $\int_{\Omega} f_n d\mu \rightarrow \int_{\Omega} f d\mu$.

2.6 Product Measures and Fubini's Theorem

Lebesgue measure on R^n is in a sense the product of n copies of one-dimensional Lebesgue measure, since the volume of an n -dimensional rectangular box is the product of the lengths of the sides. In this section we develop this idea in a general setting. We shall be interested in two constructions. First, suppose that $(\Omega_j, \mathcal{F}_j, \mu_j)$ is a measure space for $j = 1, 2, \dots, n$. We wish to construct a measure on subsets of $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ such that the measure of the “rectangle” $A_1 \times A_2 \times \dots \times A_n$ [with each $A_j \in \mathcal{F}_j$] is $\mu_1(A_1)\mu_2(A_2) \dots \mu_n(A_n)$. The second construction involves compound experiments in probability. Suppose that two observations are made, with the first observation resulting in a point $\omega_1 \in \Omega_1$, the second in a point $\omega_2 \in \Omega_2$. The probability that the first observation falls into the set A is, say, $\mu_1(A)$. Furthermore, if the first observation is ω_1 , the probability that the second observation

falls into B is, say, $\mu(\omega_1, B)$, where $\mu(\omega_1, \cdot)$ is a probability measure defined on \mathcal{F}_2 for each $\omega_1 \in \Omega_1$. The probability that the first observation will belong to A and the second will belong to B should be given by

$$\mu(A \times B) = \int_A \mu(\omega_1, B) \mu_1(d\omega_1),$$

and we would like to construct a probability measure on subsets of $\Omega_1 \times \Omega_2$ such that $\mu(A \times B)$ is given by this formula for each $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. [Intuitively, the probability that the first observation will fall near ω_1 is $\mu_1(d\omega_1)$; given that the first observation is ω_1 , the second observation will fall in B with probability $\mu(\omega_1, B)$. Thus $\mu(\omega_1, B)\mu_1(d\omega_1)$ represents the probability of one possible favorable outcome of the experiment. The total probability is found by adding the probabilities of favorable outcomes, in other words, by integrating over A . Reasoning of this type may not appear natural at this point, since we have not yet talked in detail about probability theory. However, it may serve to indicate the motivation behind the theorems of this section.]

2.6.1 Definitions. Let \mathcal{F}_j be a σ -field of subsets of Ω_j , $j = 1, 2, \dots, n$, and let $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$. A *measurable rectangle* in Ω is a set $A = A_1 \times A_2 \times \dots \times A_n$, where $A_j \in \mathcal{F}_j$ for each $j = 1, 2, \dots, n$. The smallest σ -field containing the measurable rectangles is called the *product σ -field*, written $\mathcal{F}_1 \times \mathcal{F}_2 \times \dots \times \mathcal{F}_n$. If all \mathcal{F}_j coincide with a fixed σ -field \mathcal{F} , the product σ -field is denoted by \mathcal{F}^n . Note that in spite of the notation, $\mathcal{F}_1 \times \mathcal{F}_2 \times \dots \times \mathcal{F}_n$ is not the Cartesian product of the \mathcal{F}_j ; the Cartesian product is the set of measurable rectangles, while the product σ -field is the minimal σ -field over the measurable rectangles. Note also that the collection of finite disjoint unions of measurable rectangles forms a field (see Problem 1).

The next theorem is stated in such a way that both constructions described above become special cases.

2.6.2 Product Measure Theorem. Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ be a measure space, with μ_1 σ -finite on \mathcal{F}_1 , and let Ω_2 be a set with σ -field \mathcal{F}_2 . Assume that for each $\omega_1 \in \Omega_1$ we are given a measure $\mu(\omega_1, \cdot)$ on \mathcal{F}_2 . Assume that $\mu(\omega_1, B)$, besides being a measure in B for each fixed $\omega_1 \in \Omega_1$, is Borel measurable in ω_1 for each fixed $B \in \mathcal{F}_2$. Assume that the $\mu(\omega_1, \cdot)$ are *uniformly σ -finite*; that is, Ω_2 can be written as $\bigcup_{n=1}^{\infty} B_n$, where for some positive (finite) constants k_n we have $\mu(\omega_1, B_n) \leq k_n$ for all $\omega_1 \in \Omega_1$. [The case in which the $\mu(\omega_1, \cdot)$ are uniformly bounded, that is, $\mu(\omega_1, \Omega_2) \leq k < \infty$ for all ω_1 , is of course included.]

Then there is a unique measure μ on $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ such that

$$\mu(A \times B) = \int_A \mu(\omega_1, B) \mu_1(d\omega_1) \quad \text{for all } A \in \mathcal{F}_1, B \in \mathcal{F}_2,$$

namely,

$$\mu(F) = \int_{\Omega_1} \mu(\omega_1, F(\omega_1)) \mu_1(d\omega_1), \quad F \in \mathcal{F},$$

where $F(\omega_1)$ denotes the *section* of F at ω_1 :

$$F(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in F\}.$$

Furthermore, μ is σ -finite on \mathcal{F} ; if μ_1 and all the $\mu(\omega_1, \cdot)$ are probability measures, so is μ .

PROOF. First assume that the $\mu(\omega_1, \cdot)$ are finite.

(1) If $C \in \mathcal{F}$, then $C(\omega_1) \in \mathcal{F}_2$ for each $\omega_1 \in \Omega_1$.

To prove this, let $\mathcal{C} = \{C \in \mathcal{F} : C(\omega_1) \in \mathcal{F}_2\}$. Then \mathcal{C} is a σ -field since

$$\left(\bigcup_{n=1}^{\infty} C_n \right)(\omega_1) = \bigcup_{n=1}^{\infty} C_n(\omega_1), \quad C^c(\omega_1) = (C(\omega_1))^c.$$

If $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$, then $(A \times B)(\omega_1) = B$ if $\omega_1 \in A$ and \emptyset if $\omega_1 \notin A$. Thus \mathcal{C} contains all measurable rectangles; hence $\mathcal{C} = \mathcal{F}$.

(2) If $C \in \mathcal{F}$, then $\mu(\omega_1, C(\omega_1))$ is Borel measurable in ω_1 .

To prove this, let \mathcal{C} be the class of sets in \mathcal{F} for which the conclusion of (2) holds. If $C = A \times B$, $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$, then

$$\mu(\omega_1, C(\omega_1)) = \begin{cases} \mu(\omega_1, B) & \text{if } \omega_1 \in A, \\ \mu(\omega_1, \emptyset) = 0 & \text{if } \omega_1 \notin A. \end{cases}$$

Thus $\mu(\omega_1, C(\omega_1)) = \mu(\omega_1, B)I_A(\omega_1)$, and is Borel measurable by hypothesis. Therefore measurable rectangles belong to \mathcal{C} . If C_1, \dots, C_n are disjoint measurable rectangles,

$$\begin{aligned} \mu\left(\omega_1, \left(\bigcup_{i=1}^n C_i\right)(\omega_1)\right) &= \mu\left(\omega_1, \bigcup_{i=1}^n C_i(\omega_1)\right) \\ &= \sum_{i=1}^n \mu(\omega_1, C_i(\omega_1)) \end{aligned}$$

is a finite sum of Borel measurable functions, and hence is Borel measurable in ω_1 . Thus \mathcal{C} contains the field of finite disjoint unions of measurable rect-

angles. But \mathcal{C} is a monotone class, for if $C_n \in \mathcal{C}$, $n = 1, 2, \dots$, and $C_n \uparrow C$, then $C_n(\omega_1) \uparrow C(\omega_1)$; hence $\mu(\omega_1, C_n(\omega_1)) \rightarrow \mu(\omega_1, C(\omega_1))$. Thus $\mu(\omega_1, C(\omega_1))$, a limit of measurable functions, is measurable in ω_1 . If $C_1 \downarrow C$, the same conclusion holds since the $\mu(\omega_1, \cdot)$ are finite. Thus $\mathcal{C} = \mathcal{F}$.

(3) Define

$$\mu(F) = \int_{\Omega_1} \mu(\omega_1, F(\omega_1)) \mu_1(d\omega_1), \quad F \in \mathcal{F}$$

[the integral exists by (2)]. Then μ is a measure on \mathcal{F} , and

$$\mu(A \times B) = \int_A \mu(\omega_1, B) \mu_1(d\omega_1) \quad \text{for all } A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

To prove this, let F_1, F_2, \dots be disjoint sets in \mathcal{F} . Then

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} F_n\right) &= \int_{\Omega_1} \mu\left(\omega_1, \bigcup_{n=1}^{\infty} F_n(\omega_1)\right) \mu_1(d\omega_1) \\ &= \int_{\Omega_1} \sum_{n=1}^{\infty} \mu(\omega_1, F_n(\omega_1)) \mu_1(d\omega_1) \\ &= \sum_{n=1}^{\infty} \int_{\Omega_1} \mu(\omega_1, F_n(\omega_1)) \mu_1(d\omega_1) \\ &= \sum_{n=1}^{\infty} \mu(F_n), \end{aligned}$$

proving that μ is a measure. Now

$$\begin{aligned} \mu(A \times B) &= \int_{\Omega_1} \mu(\omega_1, (A \times B)(\omega_1)) \mu_1(d\omega_1) \\ &= \int_{\Omega_1} \mu(\omega_1, B) I_A(\omega_1) \mu_1(d\omega_1) \quad [\text{see (2)}] \\ &= \int_A \mu(\omega_1, B) \mu_1(d\omega_1) \end{aligned}$$

as desired.

Now assume the $\mu(\omega_1, \cdot)$ uniformly σ -finite. Let $\Omega_2 = \bigcup_{n=1}^{\infty} B_n$, where the B_n are disjoint sets in \mathcal{F}_2 and $\mu(\omega_1, B_n) \leq k_n < \infty$ for all $\omega_1 \in \Omega_1$. If we set

$$\mu_n'(\omega_1, B) = \mu(\omega_1, B \cap B_n), \quad B \in \mathcal{F}_2,$$

the $\mu_n'(\omega_1, \cdot)$ are finite, and the above construction gives a measure μ_n' on \mathcal{F} such that

$$\begin{aligned}\mu_n'(A \times B) &= \int_A \mu_n'(\omega_1, B)\mu_1(d\omega_1), \quad A \in \mathcal{F}_1, \quad B \in \mathcal{F}_2 \\ &= \int_A \mu(\omega_1, B \cap B_n)\mu_1(d\omega_1),\end{aligned}$$

namely,

$$\mu_n'(F) = \int_{\Omega_1} \mu_n'(\omega_1, F(\omega_1))\mu_1(d\omega_1) = \int_{\Omega_1} \mu(\omega_1, F(\omega_1) \cap B_n)\mu_1(d\omega_1).$$

Let $\mu = \sum_{n=1}^{\infty} \mu_n'$; μ has the desired properties.

For the uniqueness proof, assume the $\mu(\omega_1, \cdot)$ to be uniformly σ -finite. If λ is a measure on \mathcal{F} such that $\lambda(A \times B) = \int_A \mu(\omega_1, B)\mu_1(d\omega_1)$ for all $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$, then $\lambda = \mu$ on the field \mathcal{F}_0 of finite disjoint unions of measurable rectangles. Now μ is σ -finite on \mathcal{F}_0 , for if $\Omega_2 = \bigcup_{n=1}^{\infty} B_n$ with $B_n \in \mathcal{F}_2$ and $\mu(\omega_1, B_n) \leq k_n < \infty$ for all ω_1 , and $\Omega_1 = \bigcup_{m=1}^{\infty} A_m$, where the A_m belong to \mathcal{F}_1 and $\mu_1(A_m) < \infty$, then $\Omega_1 \times \Omega_2 = \bigcup_{m,n=1}^{\infty} (A_m \times B_n)$ and

$$\mu(A_m \times B_n) = \int_{A_m} \mu(\omega_1, B_n)\mu_1(d\omega_1) \leq k_n \mu_1(A_m) < \infty.$$

Thus $\lambda = \mu$ on \mathcal{F} by the Carathéodory extension theorem.

We have just seen that μ is σ -finite on \mathcal{F}_0 , hence on \mathcal{F} . If μ_1 and all the $\mu(\omega_1, \cdot)$ are probability measures, it is immediate that μ is also. ■

2.6.3 Corollary: Classical Product Measure Theorem. Let $(\Omega_j, \mathcal{F}_j, \mu_j)$ be a measure space for $j = 1, 2$, with μ_j σ -finite on \mathcal{F}_j . If $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, the set function given by

$$\mu(F) = \int_{\Omega_1} \mu_2(F(\omega_1)) d\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(F(\omega_2)) d\mu_2(\omega_2)$$

is the unique measure on \mathcal{F} such that $\mu(A \times B) = \mu_1(A)\mu_2(B)$ for all $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$. Furthermore, μ is σ -finite on \mathcal{F} , and is a probability measure if μ_1 and μ_2 are. The measure μ is called the *product* of μ_1 and μ_2 , written $\mu = \mu_1 \times \mu_2$.

PROOF. In 2.6.2, take $\mu(\omega_1, \cdot) = \mu_2$ for all ω_1 . The second formula for $\mu(F)$ is obtained by interchanging μ_1 and μ_2 . ■

As a special case, Let $\Omega_1 = \Omega_2 = R$, $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}(R)$, $\mu_1 = \mu_2 =$ Lebesgue measure. Then $\mathcal{F}_1 \times \mathcal{F}_2 = \mathcal{B}(R^2)$ (Problem 2), and $\mu = \mu_1 \times \mu_2$ agrees with Lebesgue measure on intervals $(a, b] = (a_1, b_1] \times (a_2, b_2]$. By the Carathéodory extension theorem, μ is Lebesgue measure on $\mathcal{B}(R^2)$, so we have another method of constructing two-dimensional Lebesgue measure. We shall generalize to n dimensions later in the section.

The integration theory we have developed thus far includes the notion of a multiple integral on R^n ; this is simply an integral with respect to n -dimensional Lebesgue measure. However, in calculus, integrals of this type are evaluated by computing iterated integrals. The general theorem which justifies this process is Fubini's theorem, which is a direct consequence of the product measure theorem.

2.6.4 Fubini's Theorem. Assume the hypothesis of the product measure theorem 2.6.2. Let $f: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$.

(a) If f is nonnegative, then $\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ exists and defines a Borel measurable function of ω_1 . Also

$$\int_{\Omega} f \, d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \right) \mu_1(d\omega_1).$$

(b) If $\int_{\Omega} f \, d\mu$ exists (respectively, is finite), then $\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ exists (respectively, is finite) for μ_1 -almost every ω_1 , and defines a Borel measurable function of ω_1 if it is taken as 0 (or as any Borel measurable function of ω_1) on the exceptional set. Also,

$$\int_{\Omega} f \, d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \right) \mu_1(d\omega_1).$$

[The notation $\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ indicates that for a fixed ω_1 , the function given by $g(\omega_2) = f(\omega_1, \omega_2)$ is to be integrated with respect to the measure $\mu(\omega_1, \cdot)$.]

PROOF. (a) First note that: .

(1) For each fixed ω_1 we have $f(\omega_1, \cdot): (\Omega_2, \mathcal{F}_2) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$. In other words if f is jointly measurable, that is, measurable relative to the product σ -field $\mathcal{F}_1 \times \mathcal{F}_2$, it is measurable in each variable separately. For if $B \in \mathcal{B}(\bar{R})$, $\{\omega_2 : f(\omega_1, \omega_2) \in B\} = \{\omega_2 : (\omega_1, \omega_2) \in f^{-1}(B)\} = f^{-1}(B)(\omega_1) \in \mathcal{F}_2$ by part (1) of the proof of 2.6.2. Thus $\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ exists.

Now let I_F , $F \in \mathcal{F}$, be an indicator. Then

$$\int_{\Omega_2} I_F(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) = \int_{\Omega_2} I_{F(\omega_1)}(\omega_2) \mu(\omega_1, d\omega_2) = \mu(\omega_1, F(\omega_1)),$$

and this is Borel measurable in ω_1 by part (2) of the proof of 2.6.2. Also

$$\begin{aligned} \int_{\Omega} I_F \, d\mu &= \mu(F) = \int_{\Omega_1} \mu(\omega_1, F(\omega_1)) \mu_1(d\omega_1) \quad \text{by 2.6.2} \\ &= \int_{\Omega_1} \int_{\Omega_2} I_F(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1). \end{aligned}$$

Now if $f = \sum_{j=1}^n x_j I_{F_j}$, the F_j disjoint sets in \mathcal{F} , is a nonnegative simple function, then

$$\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) = \sum_{j=1}^n x_j \mu(\omega_1, F_j(\omega_1)),$$

Borel measurable in ω_1 ,

and

$$\begin{aligned} \int_{\Omega} f d\mu &= \sum_{j=1}^n x_j \int_{\Omega} I_{F_j} d\mu = \sum_{j=1}^n x_j \int_{\Omega_1} \int_{\Omega_2} I_{F_j}(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) \\ &\quad \text{by what we have proved for indicators} \\ &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1). \end{aligned}$$

Finally, if $f: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$, $f \geq 0$, let $0 \leq f_n \uparrow f$, f_n simple. Then

$$\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) = \lim_{n \rightarrow \infty} \int_{\Omega_2} f_n(\omega_1, \omega_2) \mu(\omega_1, d\omega_2),$$

which is Borel measurable in ω_1 , and

$$\begin{aligned} \int_{\Omega} f d\mu &= \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \lim_{n \rightarrow \infty} \int_{\Omega_1} \int_{\Omega_2} f_n(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) \\ &\quad \text{by what we have proved for simple functions} \\ &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) \\ &\quad \text{using the monotone convergence theorem twice.} \end{aligned}$$

This proves (a).

(b) Suppose that $\int_{\Omega} f^- d\mu < \infty$. By (a),

$$\int_{\Omega_1} \int_{\Omega_2} f^-(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) = \int_{\Omega} f^- d\mu < \infty$$

so that $\int_{\Omega_2} f^-(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ is μ_1 -integrable; hence finite a.e. [μ_1]; thus:

(2) For μ_1 -almost every ω_1 we may write:

$$\begin{aligned} \int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) &= \int_{\Omega_2} f^+(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \\ &\quad - \int_{\Omega_2} f^-(\omega_1, \omega_2) \mu(\omega_1, d\omega_2). \end{aligned}$$

If $\int_{\Omega} f d\mu$ is finite, both integrals on the right side of (2) are finite a.e. [μ_1]. In any event, we may define all integrals in (2) to be 0 (or any other Borel measurable function of ω_1) on the exceptional set, and (2) will then be valid.

for all ω_1 , and will define a Borel measurable function of ω_1 . If we integrate (2) with respect to μ_1 , we obtain, by (a) and the additivity theorem for integrals,

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) = \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu = \int_{\Omega} f d\mu. \blacksquare$$

2.6.5 Corollary. If $f: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ and the iterated integral $\int_{\Omega_1} \int_{\Omega_2} |f(\omega_1, \omega_2)| \mu(\omega_1, d\omega_2) \mu_1(d\omega_1) < \infty$, then $\int_{\Omega} f d\mu$ is finite, and thus Fubini's theorem applies.

PROOF. By 2.6.4(a), $\int_{\Omega} |f| d\mu < \infty$, and thus the hypothesis of 2.6.4(b) is satisfied. \blacksquare

As a special case, we obtain the following classical result.

2.6.6 Classical Fubini Theorem. Let $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $\mu = \mu_1 \times \mu_2$, where μ_j is a σ -finite measure on \mathcal{F}_j , $j = 1, 2$. If f is a Borel measurable function on (Ω, \mathcal{F}) such that $\int_{\Omega} f d\mu$ exists, then

$$\begin{aligned} \int_{\Omega} f d\mu &= \int_{\Omega_1} \int_{\Omega_2} f d\mu_2 d\mu_1 \\ &= \int_{\Omega_2} \int_{\Omega_1} f d\mu_1 d\mu_2 \quad \text{by symmetry.} \end{aligned}$$

PROOF. Apply 2.6.4 with $\mu(\omega_1, \cdot) = \mu_2$ for all ω_1 . \blacksquare

Note that by 2.6.5, if $\int_{\Omega_1} \int_{\Omega_2} |f| d\mu_2 d\mu_1 < \infty$ (or $\int_{\Omega_2} \int_{\Omega_1} |f| d\mu_1 d\mu_2 < \infty$), the iterated integration formula 2.6.6 holds.

In 2.6.4(b), if we wish to define $\int_{\Omega_2} f(\omega_1, \omega_2) \mu(\omega_1, d\omega_2)$ in a completely arbitrary fashion on the exceptional set where the integral does not exist, and still produce a Borel measurable function of ω_1 , we should assume that $(\Omega_1, \mathcal{F}_1, \mu_1)$ is a complete measure space. The situation is as follows. We have $h: (\Omega_1, \mathcal{F}_1) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$, where h is the above integral, taken as 0 on the exceptional set A . We set $g(\omega_1) = h(\omega_1)$, $\omega_1 \notin A$; $g(\omega_1) = q(\omega_1)$ arbitrary, $\omega_1 \in A$ (q not necessarily Borel measurable). If B is a Borel subset of \bar{R} , then

$$\{\omega_1 : g(\omega_1) \in B\} = [A^c \cap \{\omega_1 : h(\omega_1) \in B\}] \cup [A \cap \{\omega_1 : q(\omega_1) \in B\}].$$

The first set of the union belongs to \mathcal{F}_1 , and the second is a subset of A , with $\mu_1(A) = 0$, and hence belongs to \mathcal{F}_1 by completeness. Thus g is Borel measurable.

In the classical Fubini theorem, if we want to define $\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2)$ and $\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$ in a completely arbitrary fashion on the exceptional sets, we should assume completeness of both spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$.

The product measure theorem and Fubini's theorem may be extended to n factors, as follows:

2.6.7 Theorem. Let \mathcal{F}_j be a σ -field of subsets of Ω_j , $j = 1, \dots, n$. Let μ_1 be a σ -finite measure on \mathcal{F}_1 , and, for each $(\omega_1, \dots, \omega_j) \in \Omega_1 \times \dots \times \Omega_j$, let $\mu(\omega_1, \dots, \omega_j, B)$, $B \in \mathcal{F}_{j+1}$, be a measure on \mathcal{F}_{j+1} ($j = 1, 2, \dots, n-1$). Assume the $\mu(\omega_1, \dots, \omega_j, \cdot)$ to be uniformly σ -finite, and assume that $\mu(\omega_1, \dots, \omega_j, C)$ is measurable: $(\Omega_1 \times \dots \times \Omega_j, \mathcal{F}_1 \times \dots \times \mathcal{F}_j) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ for each fixed $C \in \mathcal{F}_{j+1}$.

Let $\Omega = \Omega_1 \times \dots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \times \dots \times \mathcal{F}_n$.

(a) There is a unique measure μ on \mathcal{F} such that for each measurable rectangle $A_1 \times \dots \times A_n \in \mathcal{F}$,

$$\begin{aligned} \mu(A_1 \times \dots \times A_n) &= \int_{A_1} \mu_1(d\omega_1) \int_{A_2} \mu(\omega_1, d\omega_2) \\ &\quad \cdots \int_{A_{n-1}} \mu(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1}) \int_{A_n} \mu(\omega_1, \dots, \omega_{n-1}, d\omega_n). \end{aligned}$$

[Note that the last factor on the right is $\mu(\omega_1, \dots, \omega_{n-1}, A_n)$.] The measure μ is σ -finite on \mathcal{F} , and is a probability measure if μ_1 and all the $\mu(\omega_1, \dots, \omega_j, \cdot)$ are probability measures.

(b) Let $f: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$. If $f \geq 0$, then

$$\begin{aligned} \int_{\Omega} f d\mu &= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu(\omega_1, d\omega_2) \cdots \int_{\Omega_{n-1}} \mu(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1}) \\ &\quad \int_{\Omega_n} f(\omega_1, \dots, \omega_n) \mu(\omega_1, \dots, \omega_{n-1}, d\omega_n), \quad (1) \end{aligned}$$

where, after the integration with respect to $\mu(\omega_1, \dots, \omega_j, \cdot)$ is performed ($j = n-1, n-2, \dots, 1$), the result is a Borel measurable function of $(\omega_1, \dots, \omega_j)$.

If $\int_{\Omega} f d\mu$ exists (respectively, is finite), then Eq. (1) holds in the sense that for each $j = n-1, n-2, \dots, 1$, the integral with respect to $\mu(\omega_1, \dots, \omega_j, \cdot)$ exists (respectively, is finite) except for $(\omega_1, \dots, \omega_j)$ in a set of λ_j -measure 0, where λ_j is the measure determined [see (a)] by μ_1 and the measures $\mu(\omega_1, \cdot), \dots, \mu(\omega_1, \dots, \omega_{j-1}, \cdot)$. If the integral is defined on the exceptional

set as 0 [or any Borel measurable function on the space $(\Omega_1 \times \cdots \times \Omega_j, \mathcal{F}_1 \times \cdots \times \mathcal{F}_j)$], it becomes Borel measurable in $(\omega_1, \dots, \omega_j)$.

PROOF. By 2.6.2 and 2.6.4, the result holds for $n = 2$. Assuming that (a) and (b) hold up to $n - 1$ factors, we consider the n -dimensional case. By the induction hypothesis, there is a unique measure λ_{n-1} on $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_{n-1}$ such that for all $A_1 \in \mathcal{F}_1, \dots, A_{n-1} \in \mathcal{F}_{n-1}$,

$$\begin{aligned}\lambda_{n-1}(A_1 \times \cdots \times A_{n-1}) &= \int_{A_1} \mu_1(d\omega_1) \int_{A_2} \mu(\omega_1, d\omega_2) \\ &\quad \cdots \int_{A_{n-1}} \mu(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1})\end{aligned}$$

and λ_{n-1} is σ -finite. By the $n = 2$ case, there is a unique measure μ on $(\mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1}) \times \mathcal{F}_n$ (which equals $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$; see Problem 3) such that for each $A \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1}, A_n \in \mathcal{F}_n$,

$$\begin{aligned}\mu(A \times A_n) &= \int_A \mu(\omega_1, \dots, \omega_{n-1}, A_n) d\lambda_{n-1}(\omega_1, \dots, \omega_{n-1}) \\ &= \int_{\Omega_1 \times \cdots \times \Omega_{n-1}} I_A(\omega_1, \dots, \omega_{n-1}) \mu(\omega_1, \dots, \omega_{n-1}, A_n) \\ &\quad d\lambda_{n-1}(\omega_1, \dots, \omega_{n-1}).\end{aligned}\tag{2}$$

If A is a measurable rectangle $A_1 \times \cdots \times A_{n-1}$, then $I_A(\omega_1, \dots, \omega_{n-1}) = I_{A_1}(\omega_1) \cdots I_{A_{n-1}}(\omega_{n-1})$; thus (2) becomes, with the aid of the induction hypothesis on (b),

$$\begin{aligned}\mu(A_1 \times \cdots \times A_n) &= \int_{A_1} \mu_1(d\omega_1) \\ &\quad \cdots \int_{A_{n-1}} \mu(\omega_1, \dots, \omega_{n-1}, A_n) \mu(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1})\end{aligned}$$

which proves the existence of the desired measure μ on \mathcal{F} . To show that μ is σ -finite on \mathcal{F}_0 , the field of finite disjoint unions of measurable rectangles, and consequently μ is unique, let $\Omega_j = \bigcup_{r=1}^{\infty} A_{jr}, j = 1, \dots, n$, where $\mu(\omega_1, \dots, \omega_{j-1}, A_{jr}) \leq k_{jr} < \infty$ for all $\omega_1, \dots, \omega_{j-1}, j = 2, \dots, n$, and $\mu_1(A_{1r}) = k_{1r} < \infty$. Then

$$\Omega = \bigcup_{i_1, \dots, i_n=1}^{\infty} (A_{1i_1} \times A_{2i_2} \times \cdots \times A_{ni_n}),$$

with

$$\mu(A_{1i_1} \times A_{2i_2} \times \cdots \times A_{ni_n}) \leq k_{1i_1} k_{2i_2} \cdots k_{ni_n} < \infty.$$

This proves (a).

To prove (b), note that the measure μ constructed in (a) is determined by λ_{n-1} and the measures $\mu(\omega_1, \dots, \omega_{n-1}, \cdot)$. Thus by the $n = 2$ case,

$$\int_{\Omega} f d\mu = \int_{\Omega_1 \times \dots \times \Omega_{n-1}} \int_{\Omega_n} f(\omega_1, \dots, \omega_n) \mu(\omega_1, \dots, \omega_{n-1}, d\omega_n) d\lambda_{n-1}(\omega_1, \dots, \omega_{n-1})$$

where the inner integral is Borel measurable in $(\omega_1, \dots, \omega_{n-1})$, or becomes so after adjustment on a set of λ_{n-1} -measure 0. The desired result now follows by the induction hypothesis. ■

2.6.8 Comments. (a) If we take $f = I_F$ in formula (1) of 2.6.7(b), we obtain an explicit formula for $\mu(F)$, $F \in \mathcal{F}$, namely,

$$\begin{aligned} \mu(F) &= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu(\omega_1, d\omega_2) \cdots \int_{\Omega_{n-1}} \mu(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1}) \\ &\quad \int_{\Omega_n} I_F(\omega_1, \dots, \omega_n) \mu(\omega_1, \dots, \omega_{n-1}, d\omega_n) \end{aligned}$$

(b) We obtain the classical product measure and Fubini theorems by taking $\mu(\omega_1, \dots, \omega_j, \cdot) \equiv \mu_{j+1}$, $j = 1, 2, \dots, n - 1$ (with μ_{j+1} σ -finite). We obtain a unique measure μ on \mathcal{F} such that on measurable rectangles,

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1)\mu_2(A_2) \cdots \mu_n(A_n).$$

If $f: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ and $f \geq 0$ or $\int_{\Omega} f d\mu$ exists, then

$$\int_{\Omega} f d\mu = \int_{\Omega_1} d\mu_1 \int_{\Omega_2} d\mu_2 \cdots \int_{\Omega_n} f d\mu_n$$

and by symmetry, the integration may be performed in any order. The measure μ is called the *product* of μ_1, \dots, μ_n , written $\mu = \mu_1 \times \cdots \times \mu_n$. In particular, if each μ_j is Lebesgue measure on $\mathcal{B}(R)$, then $\mu_1 \times \cdots \times \mu_n$ is Lebesgue measure on $\mathcal{B}(R^n)$, just as in the discussion after 2.6.3.

Problems

1. Show that the collection of finite disjoint unions of measurable rectangles in $\Omega_1 \times \cdots \times \Omega_n$ forms a field.
2. Show that $\mathcal{B}(R^n) = \mathcal{B}(R) \times \cdots \times \mathcal{B}(R)$ (n times).
3. If $\mathcal{F}_1, \dots, \mathcal{F}_n$ are arbitrary σ -fields, show that

$$(\mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1}) \times \mathcal{F}_n = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n.$$

4. Let μ be the product of the σ -finite measures μ_1 and μ_2 . If $C \in \mathcal{F}_1 \times \mathcal{F}_2$, show that the following are equivalent:
- $\mu(C) = 0$,
 - $\mu_2(C(\omega_1)) = 0$ for μ_1 -almost all $\omega_1 \in \Omega_1$,
 - $\mu_1(C(\omega_2)) = 0$ for μ_2 -almost all $\omega_2 \in \Omega_2$.
5. In Problem 4, let $(\Omega', \mathcal{F}', \mu')$ be the completion of $(\Omega, \mathcal{F}, \mu)$, and assume μ_1, μ_2 complete. If $B \in \mathcal{F}'$, show that $B(\omega_1) \in \mathcal{F}_2$ for μ_1 -almost all $\omega_1 \in \Omega_1$ [and $B(\omega_2) \in \mathcal{F}_1$ for μ_2 -almost all $\omega_2 \in \Omega_2$]. Give an example in which $B(\omega_1) \notin \mathcal{F}_2$ for some $\omega_1 \in \Omega_1$.
6. (a) Let $\Omega_1 = \Omega_2 =$ the set of positive integers, $\mathcal{F}_1 = \mathcal{F}_2 =$ all subsets, $\mu_1 = \mu_2 =$ counting measure, $f(n, n) = n, f(n, n+1) = -n, n = 1, 2, \dots, f(i, j) = 0$ if $j \neq i$ or $i+1$. Show that $\int_{\Omega_1} \int_{\Omega_2} f \, d\mu_2 \, d\mu_1 = 0, \int_{\Omega_2} \int_{\Omega_1} f \, d\mu_1 \, d\mu_2 = \infty$. (Fubini's theorem fails since the integral of f with respect to $\mu_1 \times \mu_2$ does not exist.)
- (b) Let $\Omega_1 = \Omega_2 = R, \mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}(R), \mu_1 =$ Lebesgue measure, $\mu_2 =$ counting measure. Let $A = \{(\omega_1, \omega_2) : \omega_1 = \omega_2\} \in \mathcal{F}_1 \times \mathcal{F}_2$. Show that

$$\int_{\Omega_1} \int_{\Omega_2} I_A \, d\mu_2 \, d\mu_1 = \int_{\Omega_1} \mu_2(A(\omega_1)) \, d\mu_1(\omega_1) = \infty,$$

but

$$\int_{\Omega_2} \int_{\Omega_1} I_A \, d\mu_1 \, d\mu_2 = \int_{\Omega_2} \mu_1(A(\omega_2)) \, d\mu_2(\omega_2) = 0.$$

[Fubini's theorem fails since μ_2 is not σ -finite; the product measure theorem fails also since $\int_{\Omega_1} \mu_2(F(\omega_1)) \, d\mu_1(\omega_1)$ and

$\int_{\Omega_2} \mu_1(F(\omega_2)) \, d\mu_2(\omega_2)$ do not agree on $\mathcal{F}_1 \times \mathcal{F}_2$.]

- 7.[†] Let $\Omega_1 = \Omega_2 =$ the first uncountable ordinal, $\mathcal{F}_1 = \mathcal{F}_2 =$ all subsets, $\Omega = \Omega_1 \times \Omega_2, \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. Assume the continuum hypothesis, which identifies Ω_1 and Ω_2 with $[0, 1]$.
- If f is any function from Ω_1 (or from a subset of Ω_1) to $[0, 1]$ and $G = \{(x, y) : x \in \Omega_1, y = f(x)\}$ is the graph of f , show that $G \in \mathcal{F}$.
 - Let $C_1 = \{(x, y) \in \Omega : y \leq x\}, C_2 = \{(x, y) \in \Omega : y > x\}$. If $B \subset C_1$ or $B \subset C_2$, show that $B \in \mathcal{F}$. (The relation $y \leq x$ refers to the ordering of y and x as *ordinals*, not as real numbers.)
 - Show that \mathcal{F} consists of all subsets of Ω .
8. Show that a measurable function of one variable is jointly measurable. Specifically, if $g : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega', \mathcal{F}')$ and we define $f : \Omega_1 \times \Omega_2 \rightarrow \Omega'$ by $f(\omega_1, \omega_2) = g(\omega_1)$, then f is measurable relative to $\mathcal{F}_1 \times \mathcal{F}_2$ and \mathcal{F}' , regardless of the nature of \mathcal{F}_2 .

[†] Rao, B. V., *Bull. Amer. Math. Soc.* **75**, 614 (1969).

9. Give an example of a function $f: [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that
- $f(x, y)$ is Borel measurable in y for each fixed x and Borel measurable in x for each fixed y ,
 - f is not jointly measurable, that is, f is not measurable relative to the product σ -field $\mathcal{B}[0, 1] \times \mathcal{B}[0, 1]$, and
 - $\int_0^1 (\int_0^1 f(x, y) dy) dx$ and $\int_0^1 (\int_0^1 f(x, y) dx) dy$ exist but are unequal. (One example is suggested by Problem 7.)

2.7 Measures on Infinite Product Spaces

The n -dimensional product measure theorem formalizes the notion of an n -stage random experiment, where the probability of an event associated with the n th stage depends on the result of the first $n - 1$ trials. It will be convenient later to have a single probability space which is adequate to handle n -stage experiments for n arbitrarily large (not fixed in advance). Such a space can be constructed if the product measure theorem can be extended to infinitely many dimensions. Our first task is to construct the product of infinitely many σ -fields.

2.7.1 Definitions. For each $j = 1, 2, \dots$, let $(\Omega_j, \mathcal{F}_j)$ be a measurable space. Let $\Omega = \prod_{j=1}^{\infty} \Omega_j$, the set of all sequences $(\omega_1, \omega_2, \dots)$ such that $\omega_j \in \Omega_j$, $j = 1, 2, \dots$. If $B^n \subset \prod_{j=1}^n \Omega_j$, we define

$$B_n = \{\omega \in \Omega: (\omega_1, \dots, \omega_n) \in B^n\}.$$

The set B_n is called the *cylinder* with base B^n ; the cylinder is said to be *measurable* if $B^n \in \prod_{j=1}^n \mathcal{F}_j$. If $B^n = A_1 \times \dots \times A_n$, where $A_i \subset \Omega_i$ for each i , B_n is called a *rectangle*, a *measurable rectangle* if $A_i \in \mathcal{F}_i$ for each i .

A cylinder with an n -dimensional base may always be regarded as having a higher dimensional base. For example, if

$$B = \{\omega \in \Omega: (\omega_1, \omega_2, \omega_3) \in B^3\},$$

then

$$\begin{aligned} B &= \{\omega \in \Omega: (\omega_1, \omega_2, \omega_3) \in B^3, \quad \omega_4 \in \Omega_4\} \\ &= \{\omega \in \Omega: (\omega_1, \omega_2, \omega_3, \omega_4) \in B^3 \times \Omega_4\}. \end{aligned}$$

It follows that the measurable cylinders form a field. It is also true that finite disjoint unions of measurable rectangles form a field; the argument is the same as in Problem 1 of Section 2.6.

The minimal σ -field over the measurable cylinders is called the *product* of the σ -fields \mathcal{F}_j , written $\prod_{j=1}^{\infty} \mathcal{F}_j$; $\prod_{j=1}^{\infty} \mathcal{F}_j$ is also the minimal σ -field over

the measurable rectangles (see Problem 1). If all \mathcal{F}_j coincide with a fixed σ -field \mathcal{F} , then $\prod_{j=1}^{\infty} \mathcal{F}_j$ is denoted by \mathcal{F}^{∞} , and if all Ω_j coincide with a fixed set S , $\prod_{j=1}^{\infty} \Omega_j$ is denoted by S^{∞} .

The infinite-dimensional version of the product measure theorem will be used only for probability measures, and is therefore stated in that context. (In fact the construction to be described below runs into trouble for non-probability measures.)

2.7.2 Theorem. Let $(\Omega_j, \mathcal{F}_j)$, $j = 1, 2, \dots$, be arbitrary measurable spaces; let $\Omega = \prod_{j=1}^{\infty} \Omega_j$, $\mathcal{F} = \prod_{j=1}^{\infty} \mathcal{F}_j$. Suppose that we are given an arbitrary probability measure P_1 on \mathcal{F}_1 , and for each $j = 1, 2, \dots$ and each $(\omega_1, \dots, \omega_j) \in \Omega_1 \times \dots \times \Omega_j$ we are given a probability measure $P(\omega_1, \dots, \omega_j, \cdot)$ on \mathcal{F}_{j+1} . Assume that $P(\omega_1, \dots, \omega_j, C)$ is measurable: $(\prod_{i=1}^j \Omega_i, \prod_{i=1}^j \mathcal{F}_i) \rightarrow (R, \mathcal{B}(R))$ for each fixed $C \in \mathcal{F}_{j+1}$.

If $B^n \in \prod_{j=1}^n \mathcal{F}_j$, define

$$\begin{aligned} P_n(B^n) &= \int_{\Omega_1} P_1(d\omega_1) \int_{\Omega_2} P(\omega_1, d\omega_2) \cdots \int_{\Omega_{n-1}} P(\omega_1, \dots, \omega_{n-2}, d\omega_{n-1}) \\ &\quad \int_{\Omega_n} I_{B^n}(\omega_1, \dots, \omega_n) P(\omega_1, \dots, \omega_{n-1}, d\omega_n). \end{aligned}$$

Note that P_n is a probability measure on $\prod_{j=1}^n \mathcal{F}_j$ by 2.6.7 and 2.6.8(a).

There is a unique probability measure P on \mathcal{F} such that for all n , P agrees with P_n on n -dimensional cylinders, that is, $P\{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B^n\} = P_n(B^n)$ for all $n = 1, 2, \dots$ and all $B^n \in \prod_{j=1}^n \mathcal{F}_j$.

PROOF. Any measurable cylinder can be represented in the form $B_n = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B^n\}$ for some n and some $B^n \in \prod_{j=1}^n \mathcal{F}_j$; define $P(B_n) = P_n(B^n)$. We must show that P is well-defined on measurable cylinders. For suppose that B_n can also be expressed as $\{\omega \in \Omega : (\omega_1, \dots, \omega_m) \in C^m\}$ where $C^m \in \prod_{j=1}^m \mathcal{F}_j$; we must show that $P_n(B^n) = P_m(C^m)$. Say $m < n$; then $(\omega_1, \dots, \omega_m) \in C^m$ iff $(\omega_1, \dots, \omega_n) \in B^n$, hence $B^n = C^m \times \Omega_{m+1} \times \dots \times \Omega_n$. It follows from the definition of P_n that $P_n(B^n) = P_m(C^m)$. (The fact that the $P(\omega_1, \dots, \omega_j, \cdot)$ are probability measures is used here.)

Since P_n is a measure on $\prod_{j=1}^n \mathcal{F}_j$, it is immediate that P is finitely additive on the field \mathcal{F}_0 of measurable cylinders. If we can show that P is continuous from above at the empty set, 1.2.8(b) implies that P is countably additive on \mathcal{F}_0 , and the Carathéodory extension theorem extends P to a probability measure on $\prod_{j=1}^{\infty} \mathcal{F}_j$; by construction, P agrees with P_n on n -dimensional cylinders.

Let $\{B_n, n = n_1, n_2, \dots\}$ be a sequence of measurable cylinders decreasing to \emptyset (we may assume $n_1 < n_2 < \dots$, and in fact nothing is lost if we take $n_i = i$ for all i). Assume $\lim_{n \rightarrow \infty} P(B_n) > 0$. Then for each $n > 1$,

$$P(B_n) = \int_{\Omega_1} g_n^{(1)}(\omega_1) P_1(d\omega_1),$$

where

$$g_n^{(1)}(\omega_1) = \int_{\Omega_2} P(\omega_1, d\omega_2) \cdots \int_{\Omega_n} I_{B^n}(\omega_1, \dots, \omega_n) P(\omega_1, \dots, \omega_{n-1}, d\omega_n).$$

Since $B_{n+1} \subset B_n$, it follows that $B^{n+1} \subset B^n \times \Omega_{n+1}$; hence

$$I_{B^{n+1}}(\omega_1, \dots, \omega_{n+1}) \leq I_{B^n}(\omega_1, \dots, \omega_n).$$

Therefore $g_n^{(1)}(\omega_1)$ decreases as n increases (ω_1 fixed); say $g_n^{(1)}(\omega_1) \rightarrow h_1(\omega_1)$. By the extended monotone convergence theorem (or the dominated convergence theorem), $P(B_n) \rightarrow \int_{\Omega_1} h_1(\omega_1) P_1(d\omega_1)$. If $\lim_{n \rightarrow \infty} P(B_n) > 0$, then $h_1(\omega_1') > 0$ for some $\omega_1' \in \Omega_1$. In fact $\omega_1' \in B^1$, for if not, $I_{B^n}(\omega_1', \omega_2, \dots, \omega_n) = 0$ for all n ; hence $g_n^{(1)}(\omega_1') = 0$ for all n , and $h_1(\omega_1') = 0$, a contradiction.

Now for each $n > 2$,

$$g_n^{(1)}(\omega_1') = \int_{\Omega_2} g_n^{(2)}(\omega_2) P(\omega_1', d\omega_2),$$

where

$$\begin{aligned} g_n^{(2)}(\omega_2) &= \int_{\Omega_3} P(\omega_1', \omega_2, d\omega_3) \\ &\cdots \int_{\Omega_n} I_{B^n}(\omega_1', \omega_2, \dots, \omega_n) P(\omega_1', \dots, \omega_{n-1}, d\omega_n). \end{aligned}$$

As above, $g_n^{(2)}(\omega_2) \downarrow h_2(\omega_2)$; hence

$$g_n^{(1)}(\omega_1') \rightarrow \int_{\Omega_2} h_2(\omega_2) P(\omega_1', d\omega_2).$$

Since $g_n^{(1)}(\omega_1') \rightarrow h_1(\omega_1') > 0$, we have $h_2(\omega_2') > 0$ for some $\omega_2' \in \Omega_2$, and as above we have $(\omega_1', \omega_2') \in B^2$.

The process may be repeated inductively to obtain points $\omega_1', \omega_2', \dots$ such that for each n , $(\omega_1', \dots, \omega_n') \in B^n$. But then $(\omega_1', \omega_2', \dots) \in \bigcap_{n=1}^{\infty} B_n = \emptyset$, a contradiction. This proves the existence of the desired probability measure P . If Q is another such probability measure, then $P = Q$ on measurable cylinders, hence $P = Q$ on \mathcal{F} by the uniqueness part of the Carathéodory extension theorem. ■

The classical product measure theorem extends as follows:

2.7.3 Corollary. For each $j = 1, 2, \dots$, let $(\Omega_j, \mathcal{F}_j, P_j)$ be an arbitrary probability space. Let $\Omega = \prod_{j=1}^{\infty} \Omega_j$, $\mathcal{F} = \prod_{j=1}^{\infty} \mathcal{F}_j$. There is a unique probability measure P on \mathcal{F} such that

$$P\{\omega \in \Omega : \omega_1 \in A_1, \dots, \omega_n \in A_n\} = \prod_{j=1}^n P_j(A_j)$$

for all $n = 1, 2, \dots$ and all $A_j \in \mathcal{F}_j$, $j = 1, 2, \dots$. We call P the *product* of the P_j , and write $P = \prod_{j=1}^{\infty} P_j$.

PROOF. In 2.7.2, take $P(\omega_1, \dots, \omega_j, B) = P_{j+1}(B)$, $B \in \mathcal{F}_{j+1}$. Then $P_n(A_1 \times \dots \times A_n) = \prod_{j=1}^n P_j(A_j)$, and thus the probability measure P of 2.7.2 has the desired properties. If Q is another such probability measure, then $P = Q$ on the field of finite disjoint unions of measurable rectangles; hence $P = Q$ on \mathcal{F} by the Carathéodory extension theorem. ■

Problems

1. Show that $\prod_{j=1}^{\infty} \mathcal{F}_j$ is the minimal σ -field over the measurable rectangles.
2. Let $\mathcal{F} = \mathcal{B}(R)$; show that the following sets belong to \mathcal{F}^∞ :
 - (a) $\{x \in R^\infty : \sup_n x_n < a\}$,
 - (b) $\{x \in R^\infty : \sum_{n=1}^{\infty} |x_n| < a\}$,
 - (c) $\{x \in R^\infty : \lim_{n \rightarrow \infty} x_n \text{ exists and is finite}\}$,
 - (d) $\{x \in R^\infty : \limsup_{n \rightarrow \infty} x_n \leq a\}$,
 - (e) $\{x \in R^\infty : \sum_{k=1}^n x_k = 0 \text{ for at least one } n > 0\}$.
3. Let \mathcal{F} be a σ -field of subsets of a set S , and assume \mathcal{F} is countably generated, that is, there is a sequence of sets A_1, A_2, \dots in \mathcal{F} such that the smallest σ -field containing the A_j is \mathcal{F} . Show that \mathcal{F}^∞ is also countably generated. In particular, $\mathcal{B}(R)^\infty$ is countably generated; take the A_j as intervals with rational endpoints.
4. How many sets are there in $\mathcal{B}(R)^\infty$?
5. Define $f: R^\infty \rightarrow \bar{R}$ as follows:

$$f(x_1, x_2, \dots) = \begin{cases} \text{the smallest positive integer } n \\ \text{such that } x_1 + \dots + x_n \geq 1, \\ \text{if such an } n \text{ exists,} \\ \infty \text{ if } x_1 + \dots + x_n < 1 \text{ for all } n. \end{cases}$$

Show that $f: (R^\infty, \mathcal{B}(R)^\infty) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$.

2.8 References

The presentation in Chapters 1 and 2 has been strongly influenced by several sources. The first systematic presentation of measure theory appeared in Halmos (1950). Halmos achieves slightly greater generality at the expense of technical complications by replacing σ -fields by σ -rings. (A σ -ring is a class of sets closed under differences and countable unions.) However, σ -fields will be completely adequate for our purposes. The first account of measure theory specifically oriented toward probability was given by Loève (1955). Several useful refinements were made by Royden (1963), Neveu (1965), and Rudin (1966). Neveu's book emphasizes probability while Rudin's book is particularly helpful as a preparation for work in harmonic analysis.

For further properties of finitely additive set functions, and a development of integration theory for functions with values in a Banach space, see Dunford and Schwartz (1958).

3

Introduction to Functional Analysis

3.1 Introduction

An important part of analysis consists of the study of vector spaces endowed with an additional structure of some kind. In Chapter 2, for example, we studied the vector space $L^p(\Omega, \mathcal{F}, \mu)$. If $1 \leq p \leq \infty$, the seminorm $\| \cdot \|_p$ allowed us to talk about such notions as distance, convergence, and completeness.

In this chapter, we look at various structures that can be defined on vector spaces. The most general concept studied is that of a topological vector space, which is a vector space endowed with a topology compatible with the algebraic operations, that is, the topology makes vector addition and scalar multiplication continuous. Special cases are Banach and Hilbert spaces. In a Banach space there is a notion of length of a vector, and in a Hilbert space, length is in turn determined by a “dot product” of vectors. Hilbert spaces are a natural generalization of finite-dimensional Euclidean spaces.

We now list the spaces we are going to study. The term “vector space” will always mean vector space over the complex field C ; “real vector space” indicates that the scalar field is R ; no other fields will be considered.

3.1.1 Definitions. Let L be a vector space. A *seminorm* on L is a function $\| \cdot \|$ from L to the nonnegative reals satisfying

$$\begin{aligned} \|ax\| &= |a| \|x\| && \text{for all } a \in C, x \in L, \\ \|x + y\| &\leq \|x\| + \|y\| && \text{for all } x, y \in L. \end{aligned}$$

The first property is called *absolute homogeneity*, the second *subadditivity*. Note that absolute homogeneity implies that $\|0\| = 0$. (We use the same symbol for the zero vector and the zero scalar.) If, in addition, $\|x\| = 0$ implies that $x = 0$, the seminorm is called a *norm* on L and L is said to be a *normed linear space*.

If $\|\cdot\|$ is a seminorm on L , and $d(x, y) = \|x - y\|$, $x, y \in L$, d is a pseudo-metric on L ; a metric if $\|\cdot\|$ is a norm. A *Banach space* is a complete normed linear space, that is, relative to the metric d induced by the norm, every Cauchy sequence converges.

An *inner product* on L is a function from $L \times L$ to C , denoted by $(x, y) \rightarrow \langle x, y \rangle$, satisfying

$$\begin{aligned}\langle ax + by, z \rangle &= a\langle x, z \rangle + b\langle y, z \rangle \quad \text{for all } a, b \in C, \quad x, y, z \in L, \\ \langle x, y \rangle &= \overline{\langle y, x \rangle} \quad \text{for all } x, y \in L \\ \langle x, x \rangle &\geq 0 \quad \text{for all } x \in L, \\ \langle x, x \rangle &= 0 \quad \text{if and only if } x = 0\end{aligned}$$

(the over-bar indicates complex conjugation). A vector space endowed with an inner product is called an *inner product space* or *pre-Hilbert space*. If L is an inner product space, $\|x\| = (\langle x, x \rangle)^{1/2}$ defines a norm on L ; this is a consequence of the Cauchy-Schwarz inequality, to be proved in Section 3.2. If, with this norm, L is complete, L is said to be a *Hilbert space*. Thus a Hilbert space is a Banach space whose norm is determined by an inner product.

Finally, a *topological vector space* is a vector space L with a topology such that addition and scalar multiplication are continuous, in other words, the mappings

$$(x, y) \rightarrow x + y \quad \text{of } L \times L \text{ into } L$$

and

$$(a, x) \rightarrow ax \quad \text{of } C \times L \text{ into } L$$

are continuous, with the product topology on $L \times L$ and $C \times L$. In many books, the topology is required to be Hausdorff, but we find it more convenient not to make this assumption.

A Banach space is a topological vector space with the topology induced by the metric $d(x, y) = \|x - y\|$. For if $x_n \rightarrow x$ and $y_n \rightarrow y$, then

$$\|x_n + y_n - (x + y)\| \leq \|x_n - x\| + \|y_n - y\| \rightarrow 0;$$

if $a_n \rightarrow a$ and $x_n \rightarrow x$, then

$$\begin{aligned}\|a_n x_n - ax\| &\leq \|a_n x_n - a_n x\| + \|a_n x - ax\| \\ &= |a_n| \|x_n - x\| + |a_n - a| \|x\| \rightarrow 0.\end{aligned}$$

The above definitions remain unchanged if L is a real vector space, except of course that C is replaced by R . Also, we may drop the complex conjugate in the symmetry requirement for inner product and simply write $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in L$.

3.1.2 Examples. (a) If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $1 \leq p \leq \infty$, $\| \cdot \|_p$ is a seminorm on the vector space $L^p(\Omega, \mathcal{F}, \mu)$. If we pass to equivalence classes by identifying functions that agree a.e. $[\mu]$, we obtain $L^p(\Omega, \mathcal{F}, \mu)$, a Banach space (see 2.4). When $p = 2$, the norm $\| \cdot \|_p$ is determined by an inner product

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

Hence $L^2(\Omega, \mathcal{F}, \mu)$ is a Hilbert space.

If \mathcal{F} consists of all subsets of Ω and μ is counting measure, then $f = g$ a.e. $[\mu]$ implies $f \equiv g$. Thus it is not necessary to pass to equivalence classes; $L^p(\Omega, \mathcal{F}, \mu)$ is a Banach space, denoted for simplicity by $l^p(\Omega)$.

By 2.4.12, if $1 \leq p < \infty$, then $l^p(\Omega)$ consists of all functions $f = (f(\alpha), \alpha \in \Omega)$ from Ω to C such that $f(\alpha) = 0$ for all but countably many α , and $\|f\|_p^p = \sum_{\alpha} |f(\alpha)|^p < \infty$. When $p = 2$, the norm on $l^2(\Omega)$ is induced by the inner product

$$\langle f, g \rangle = \sum_{\alpha} f(\alpha)\overline{g(\alpha)}.$$

When $p = \infty$, the situation is slightly different. The space $l^\infty(\Omega)$ is the collection of all bounded complex-valued functions on Ω , with the *sup norm*

$$\|f\| = \sup\{|f(x)| : x \in \Omega\}.$$

Similarly, if Ω is a topological space and L is the class of all bounded continuous complex-valued functions on Ω , then L is a Banach space under the sup norm, for we may verify directly that the sup norm is actually a norm, or equally well we may use the fact that $L \subset l^\infty(\Omega)$. Thus we need only check completeness, and this follows because a uniform limit of continuous functions is continuous.

(b) Let c be the set of all convergent sequences of complex numbers, and put the sup norm on c ; if $f = \{a_n, n \geq 1\} \in c$, then

$$\|f\| = \sup\{|a_n| : n = 1, 2, \dots\}.$$

Again, to show that c is a Banach space we need only establish completeness. Let $\{f_n\}$ be a Cauchy sequence in c ; if $f_n = \{a_{nk}, k \geq 1\}$, then

$\lim_{k \rightarrow \infty} a_{nk}$ exists from each n since $f_n \in c$, and $b_k = \lim_{n \rightarrow \infty} a_{nk}$ exists,

uniformly in k , since $|a_{nk} - a_{mk}| \leq \|f_n - f_m\| \rightarrow 0$ as $n, m \rightarrow \infty$. By the standard double limit theorem,

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} a_{nk} = \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} a_{nk}.$$

In particular, $\lim_{k \rightarrow \infty} b_k$ exists, so if $f = \{b_k, k \geq 1\}$, then $f \in C$. But $\|f_n - f\| = \sup_k |a_{nk} - b_k| \rightarrow 0$ as $n \rightarrow \infty$ since $a_{nk} \rightarrow b_k$ uniformly in k . This proves completeness.

(c) Let L be the collection of all complex valued functions on S , where S is an arbitrary set. Put the topology of pointwise convergence on L , so that a sequence or net $\{f_n\}$ of functions in L converges to the function $f \in L$ if and only if $f_n(x) \rightarrow f(x)$ for each $x \in S$. (See the appendix on general topology, Section A1, for properties of nets.) With this topology, L is a topological vector space. To show that addition is continuous, observe that if $f_n \rightarrow f$ and $g_n \rightarrow g$ pointwise, then $f_n + g_n \rightarrow f + g$ pointwise. Similarly if $a_n \in C$, $n = 1, 2, \dots$, $a_n \rightarrow a$, and $f_n \rightarrow f$ pointwise, then $a_n f_n \rightarrow a f$ pointwise.

3.2 Basic Properties of Hilbert Spaces

Hilbert spaces are a natural generalization of finite-dimensional Euclidean spaces in the sense that many of the familiar geometric results in R^n carry over. First recall the definition of the inner product (or “dot product”) on R^n : If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, then $\langle x, y \rangle = \sum_{j=1}^n x_j y_j$. (This becomes $\sum_{j=1}^n x_j \bar{y}_j$ in the space C^n of all n -tuples of complex numbers.) The length of a vector in R^n is given by $\|x\| = (\langle x, x \rangle)^{1/2} = (\sum_{j=1}^n x_j^2)^{1/2}$, and the distance between two points of R^n is $d(x, y) = \|x - y\|$. In order to show that d is a metric, the triangle inequality must be established; this in turn follows from the Cauchy–Schwarz inequality $|\langle x, y \rangle| \leq \|x\| \|y\|$. In fact the Cauchy–Schwarz inequality holds in any inner product space, as we now prove:

3.2.1 Cauchy–Schwarz Inequality. If L is an inner product space, and $\|x\| = (\langle x, x \rangle)^{1/2}$, $x \in L$, then

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad \text{for all } x, y \in L.$$

Equality holds iff x and y are linearly dependent.

PROOF. For any $a \in C$,

$$\begin{aligned} 0 &\leq \langle x + ay, x + ay \rangle = \langle x + ay, x \rangle + \langle x + ay, ay \rangle \\ &= \langle x, x \rangle + a\langle y, x \rangle + \bar{a}\langle x, y \rangle + |a|^2\langle y, y \rangle. \end{aligned}$$

Set $a = -\langle x, y \rangle / \langle y, y \rangle$ (if $\langle y, y \rangle = 0$, then $y = 0$ and the result is trivial). Since $\langle y, x \rangle = \langle x, y \rangle$, we have

$$0 \leq \langle x, x \rangle - 2 \frac{|\langle x, y \rangle|^2}{\langle y, y \rangle} + \frac{|\langle x, y \rangle|^2}{\langle y, y \rangle}$$

proving the inequality.

Since $\langle x + ay, x + ay \rangle = 0$ iff $x + ay = 0$, equality holds iff x and y are linearly dependent. ■

3.2.2 Corollary. If L is an inner product space and $\|x\| = (\langle x, x \rangle)^{1/2}$, $x \in L$, then $\| \cdot \|$ is a norm on L .

PROOF. It is immediate that $\|x\| \geq 0$, $\|ax\| = |a| \|x\|$, and $\|x\| = 0$ iff $x = 0$. Now

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle \\ &= \|x\|^2 + \|y\|^2 + 2 \operatorname{Re} \langle x, y \rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\| \|y\| \quad \text{by 3.2.1.} \end{aligned}$$

Therefore $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$. ■

3.2.3 Corollary. An inner product is (jointly) continuous in both variables, that is, $x_n \rightarrow x$, $y_n \rightarrow y$ implies $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$ ($\{(x_n, y_n)\}$ can be a net as well as a sequence).

PROOF.

$$\begin{aligned} |\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n, y_n - y \rangle + \langle x_n - x, y \rangle| \\ &\leq \|x_n\| \|y_n - y\| + \|x_n - x\| \|y\| \quad \text{by 3.2.1.} \end{aligned}$$

But by subadditivity of the norm,

$$|\|x_n\| - \|x\|| \leq \|x_n - x\| \rightarrow 0;$$

hence $\|x_n\| \rightarrow \|x\|$. It follows that $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$. ■

The computation of 3.2.2 establishes the following result, which says geometrically that the sum of the squares of the lengths of the diagonals of a parallelogram is twice the sum of the squares of the lengths of the sides:

3.2.4 Parallelogram Law. In an inner product space,

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

PROOF.

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2 \operatorname{Re}\langle x, y \rangle,$$

and

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2 \operatorname{Re}\langle x, y \rangle. \blacksquare$$

Now suppose that x_1, \dots, x_n are mutually perpendicular unit vectors in R^k , $k \geq n$. If x is an arbitrary vector in R^k , we try to approximate x by a linear combination $\sum_{j=1}^n a_j x_j$. The reader may recall that $\sum_{j=1}^n a_j x_j$ will be closest to x in the sense of Euclidean distance when $a_j = \langle x, x_j \rangle$. This result holds in an arbitrary inner product space.

3.2.5 Definition. Two elements x and y in an inner product space L are said to be *orthogonal* or *perpendicular* iff $\langle x, y \rangle = 0$. If $B \subset L$, B is said to be *orthogonal* iff $\langle x, y \rangle = 0$ for all $x, y \in B$ such that $x \neq y$; B is *orthonormal* iff it is orthogonal and $\|x\| = 1$ for all $x \in B$.

The computation of 3.2.2 shows that if x_1, x_2, \dots, x_n are orthogonal, the *Pythagorean relation* holds: $\|\sum_{i=1}^n x_i\|^2 = \sum_{i=1}^n \|x_i\|^2$.

3.2.6 Theorem. If $\{x_1, \dots, x_n\}$ is an orthonormal set in the inner product space L , and $x \in L$,

$$\left\| x - \sum_{j=1}^n a_j x_j \right\| \text{ is minimized when } a_j = \langle x, x_j \rangle, \quad j = 1, \dots, n.$$

PROOF.

$$\begin{aligned} \left\| x - \sum_{j=1}^n a_j x_j \right\|^2 &= \left\langle x - \sum_{j=1}^n a_j x_j, x - \sum_{k=1}^n a_k x_k \right\rangle \\ &= \|x\|^2 - \sum_{k=1}^n \bar{a}_k \langle x, x_k \rangle - \sum_{j=1}^n a_j \langle x_j, x \rangle \\ &\quad + \left\langle \sum_{j=1}^n a_j x_j, \sum_{k=1}^n a_k x_k \right\rangle. \end{aligned}$$

The last term on the right is $\sum_{j=1}^n |a_j|^2$ since the x_j are orthonormal. Furthermore, $-\bar{a}_j \langle x, x_j \rangle - a_j \langle x_j, x \rangle + |a_j|^2 = -|\langle x, x_j \rangle|^2 + |a_j - \langle x, x_j \rangle|^2$. Thus

$$0 \leq \left\| x - \sum_{j=1}^n a_j x_j \right\|^2 = \|x\|^2 - \sum_{j=1}^n |\langle x, x_j \rangle|^2 + \sum_{j=1}^n |a_j - \langle x, x_j \rangle|^2, \quad (1)$$

so that we can do no better than to take $a_j = \langle x, x_j \rangle$. ■

The above computation establishes the following important inequality.

3.2.7 Bessel's Inequality. If B is an arbitrary orthonormal subset of the inner product space L and x is an element of L , then

$$\|x\|^2 \geq \sum_{y \in B} |\langle x, y \rangle|^2.$$

In other words, $\langle x, y \rangle = 0$ for all but countably many $y \in B$, say $y = x_1, x_2, \dots$, and

$$\|x\|^2 \geq \sum_j |\langle x, x_j \rangle|^2.$$

Equality holds iff $\sum_{j=1}^n \langle x, x_j \rangle x_j \rightarrow x$ as $n \rightarrow \infty$.

PROOF. If $x_1, \dots, x_n \in B$, set $a_j = \langle x, x_j \rangle$ in Eq. (1) of 3.2.6 to obtain $\|x - \sum_{j=1}^n \langle x, x_j \rangle x_j\|^2 = \|x\|^2 - \sum_{j=1}^n |\langle x, x_j \rangle|^2 \geq 0$. ■

We now consider another basic geometric idea, that of *projection*. If M is a subspace of R^n and x is any vector in R^n , x can be resolved into a component in M and a component perpendicular to M . In other words, $x = y + z$ where $y \in M$ and z is orthogonal to every vector in M . Before generalizing to an arbitrary space, we indicate some terminology.

3.2.8 Definitions. A *subspace* or *linear manifold* of a vector space L is a subset M of L that is also a vector space; that is, M is closed under addition and scalar multiplication. If L is a topological vector space, M is said to be a *closed subspace* of L if M is a subspace and is also a closed set in the topology of L .

A subset M of the vector space L is said to be *convex* iff for all $x, y \in M$, we have $ax + (1 - a)y \in M$ for all real $a \in [0, 1]$.

The key fact that we need is that if M is a closed convex subset of the Hilbert space H and x is an arbitrary point of H , there is a unique point of M closest to x .

3.2.9 Theorem. Let M be a nonempty closed convex subset of the Hilbert space H . If $x \in H$, there is a unique element $y_0 \in M$ such that

$$\|x - y_0\| = \inf\{\|x - y\| : y \in M\}.$$

PROOF. Let $d = \inf\{\|x - y\| : y \in M\}$, and pick points $y_1, y_2, \dots \in M$ with $\|x - y_n\| \rightarrow d$ as $n \rightarrow \infty$; we show that $\{y_n\}$ is a Cauchy sequence.

Since $\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$ for all $u, v \in H$ by the parallelogram law 3.2.4, we may set $u = y_n - x$, $v = y_m - x$ to obtain

$$\|y_n + y_m - 2x\|^2 + \|y_n - y_m\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2$$

or

$$\|y_n - y_m\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\frac{1}{2}(y_n + y_m) - x\|^2.$$

Since $\frac{1}{2}(y_n + y_m) \in M$ by convexity, $\|\frac{1}{2}(y_n + y_m) - x\|^2 \geq d^2$, and it follows that $\|y_n - y_m\| \rightarrow 0$ as $n, m \rightarrow \infty$.

By completeness of H , y_n approaches a limit y_0 , hence $\|x - y_n\| \rightarrow \|x - y_0\|$. But then $\|x - y_0\| = d$, and $y_0 \in M$ since M is closed; this finishes the existence part of the proof.

To prove uniqueness, let $y_0, z_0 \in M$, with $\|x - y_0\| = \|x - z_0\| = d$. In the parallelogram law, take $u = y_0 - x$, $v = z_0 - x$, to obtain

$$\|y_0 + z_0 - 2x\|^2 + \|y_0 - z_0\|^2 = 2\|y_0 - x\|^2 + 2\|z_0 - x\|^2 = 4d^2.$$

But $\|y_0 + z_0 - 2x\|^2 = 4\|\frac{1}{2}(y_0 + z_0) - x\|^2 \geq 4d^2$; hence $\|y_0 - z_0\| = 0$, so $y_0 = z_0$. ■

If M is a closed subspace of H , the element y_0 found in 3.2.9 is called the *projection* of x on M . The following result helps to justify this terminology.

3.2.10 Theorem. Let M be a closed subspace of the Hilbert space H , and y_0 an element of M . Then

$$\|x - y_0\| = \inf\{\|x - y\| : y \in M\} \quad \text{iff} \quad x - y_0 \perp M,$$

that is, $\langle x - y_0, y \rangle = 0$ for all $y \in M$.

PROOF. Assume $x - y_0 \perp M$. If $y \in M$, then

$$\begin{aligned} \|x - y\|^2 &= \|x - y_0 - (y - y_0)\|^2 \\ &= \|x - y_0\|^2 + \|y - y_0\|^2 - 2 \operatorname{Re}\langle x - y_0, y - y_0 \rangle \\ &= \|x - y_0\|^2 + \|y - y_0\|^2 \quad \text{since } y - y_0 \in M \\ &\geq \|x - y_0\|^2. \end{aligned}$$

Therefore, $\|x - y_0\| = \inf\{\|x - y\| : y \in M\}$.

Conversely, assume $\|x - y_0\| = \inf\{\|x - y\|: y \in M\}$. Let $y \in M$ and let c be an arbitrary complex number. Then $y_0 + cy \in M$ since M is a subspace, hence $\|x - y_0 - cy\| \geq \|x - y_0\|$. But

$$\|x - y_0 - cy\|^2 = \|x - y_0\|^2 + |c|^2\|y\|^2 - 2 \operatorname{Re}\langle x - y_0, cy \rangle;$$

hence

$$|c|^2\|y\|^2 - 2 \operatorname{Re}\langle x - y_0, cy \rangle \geq 0.$$

Take $c = b\langle x - y_0, y \rangle$, b real. Then $\langle x - y_0, cy \rangle = b|\langle x - y_0, y \rangle|^2$. Thus $|\langle x - y_0, y \rangle|^2[b^2\|y\|^2 - 2b] \geq 0$. But the expression in square brackets is negative if b is positive and sufficiently close to 0; hence $\langle x - y_0, y \rangle = 0$. ■

We may give still another way of characterizing the projection of x on M .

3.2.11 Projection Theorem. Let M be a closed subspace of the Hilbert space H . If $x \in H$, then x has a unique representation $x = y + z$ where $y \in M$ and $z \perp M$. Furthermore, y is the projection of x on M .

PROOF. Let y_0 be the projection of x on M , and take $y = y_0$, $z = x - y_0$. By 3.2.10, $z \perp M$, proving the existence of the desired representation. To prove uniqueness, let $x = y + z = y' + z'$ where $y, y' \in M$, $z, z' \perp M$. Then $y - y' \in M$ since M is a subspace, and $y - y' \perp M$ since $y - y' = z' - z$. Thus $y - y'$ is orthogonal to itself, hence $y = y'$. But then $z = z'$, proving uniqueness. ■

If M is any subset of H , the set $M^\perp = \{x \in H: x \perp M\}$ is a closed subspace by definition of the inner product and 3.2.3. If M is a closed subspace, M^\perp is called the *orthogonal complement* of H , and the projection theorem is expressed by saying that H is the *orthogonal direct sum* of M and M^\perp , written $H = M \oplus M^\perp$.

In R^n , it is possible to construct an orthonormal basis, that is, a set $\{x_1, \dots, x_n\}$ of n mutually perpendicular unit vectors. Any vector x in R^n may then be represented as $x = \sum_{i=1}^n \langle x, x_i \rangle x_i$, so that $\langle x, x_i \rangle$ is the component of x in the direction of x_i . We are now able to generalize this idea to an arbitrary Hilbert space. The following terminology will be used.

3.2.12 Definitions. If B is a subset of the topological vector space L , the *space spanned by B* , denoted by $S(B)$, is the smallest closed subspace of L containing all elements of B . If $L(B)$ is the linear manifold generated by B ,

that is, $L(B)$ consists of all elements $\sum_{i=1}^n a_i x_i$, $a_i \in C$, $x_i \in B$, $i = 1, \dots, n$, $n = 1, 2, \dots$, then $S(B) = \overline{L(B)}$.

If B is a subset of the Hilbert space H , B is said to be an *orthonormal basis* for H iff B is a maximal orthonormal subset of H , in other words, B is not a proper subset of any other orthonormal subset of H . An orthonormal set $B \subset H$ is maximal iff $S(B) = H$, and there are several other conditions equivalent to this, as we now prove.

3.2.13 Theorem. Let $B = \{x_\alpha, \alpha \in I\}$ be an orthonormal subset of the Hilbert space H . The following conditions are equivalent:

- (a) B is an orthonormal basis.
- (b) B is a “complete orthonormal set,” that is, the only $x \in H$ such that $x \perp B$ is $x = 0$.
- (c) B spans H , that is, $S(B) = H$.
- (d) For all $x \in H$, $x = \sum_\alpha \langle x, x_\alpha \rangle x_\alpha$. (Let us explain this notation. By 3.2.7, $\langle x, x_\alpha \rangle = 0$ for all but countably many x_α , say for x_1, x_2, \dots ; the assertion is that $\sum_{j=1}^n \langle x, x_j \rangle x_j \rightarrow x$, and this holds regardless of the order in which the x_j are listed.)
- (e) For all $x, y \in H$, $\langle x, y \rangle = \sum_\alpha \langle x, x_\alpha \rangle \langle x_\alpha, y \rangle$.
- (f) For all $x \in H$, $\|x\|^2 = \sum_\alpha |\langle x, x_\alpha \rangle|^2$.

Condition (f) [and sometimes (e) as well] is referred to as the *Parseval relation*.

PROOF. (a) implies (b): If $x \perp B$, $x \neq 0$, let $y = x/\|x\|$. Then $B \cup \{y\}$ is an orthonormal set, contradicting the maximality of B .

(b) implies (c): If $x \in H$, write $x = y + z$ where $y \in S(B)$ and $z \perp S(B)$ (see 3.2.11). By (b), $z = 0$; hence $x \in S(B)$.

(c) implies (d): Since $S(B) = \overline{L(B)}$, given $x \in H$ and $\varepsilon > 0$ there is a finite set $F \subset I$ and complex numbers a_α , $\alpha \in F$, such that

$$\left\| x - \sum_{\alpha \in F} a_\alpha x_\alpha \right\| \leq \varepsilon.$$

By 3.2.6, if G is any finite subset of I such that $F \subset G$,

$$\begin{aligned} \left\| x - \sum_{\alpha \in G} \langle x, x_\alpha \rangle x_\alpha \right\| &\leq \left\| x - \sum_{\alpha \in G} a_\alpha x_\alpha \right\| \quad \text{where } a_\alpha = 0 \text{ for } \alpha \notin F \\ &= \left\| x - \sum_{\alpha \in F} a_\alpha x_\alpha \right\| \leq \varepsilon. \end{aligned}$$

Thus if x_1, x_2, \dots is any ordering of the points $x_\alpha \in B$ for which $\langle x, x_\alpha \rangle \neq 0$, $\|x - \sum_{j=1}^n \langle x, x_j \rangle x_j\| \leq \varepsilon$ for sufficiently large n , as desired.

(d) implies (e): This is immediate from 3.2.3.

(e) implies (f): Set $x = y$ in (e).

(f) implies (a): Let C be an orthonormal set with $B \subset C$, $B \neq C$. If $x \in C$, $x \notin B$, we have $\|x\|^2 = \sum_\alpha |\langle x, x_\alpha \rangle|^2 = 0$ since by orthonormality of C , x is orthogonal to everything in B . This is a contradiction because $\|x\| = 1$ for all $x \in C$. ■

3.2.14 Corollary. Let $B = \{x_\alpha, \alpha \in I\}$ be an orthonormal subset of H , not necessarily a basis.

(a) B is an orthonormal basis for $S(B)$. [Note that $S(B)$ is a closed subspace of H , hence is itself a Hilbert space with the same inner product.]

(b) If $x \in H$ and y is the projection of x on $S(B)$, then

$$y = \sum_\alpha \langle x, x_\alpha \rangle x_\alpha$$

[see 3.2.13(d) for the interpretation of the series].

PROOF. (a) Let $x \in S(B)$, $x \perp B$; then $x \perp L(B)$. If $y \in S(B)$, let $y_1, y_2, \dots \in L(B)$ with $y_n \rightarrow y$. Since $\langle x, y_n \rangle = 0$ for all n , we have $\langle x, y \rangle = 0$ by 3.2.3. Thus $x \perp S(B)$, so that $\langle x, x \rangle = 0$, hence $x = 0$. The result follows from 3.2.13(b).

(b) By part (a) and 3.2.13(d), $y = \sum_\alpha \langle y, x_\alpha \rangle x_\alpha$. But $x - y \perp S(B)$ by 3.2.11, hence $\langle x, x_\alpha \rangle = \langle y, x_\alpha \rangle$ for all α . ■

A standard application of Zorn's lemma shows that every Hilbert space has an orthonormal basis; an additional argument shows that any two orthonormal bases have the same cardinality (see Problem 5). This fact may be used to classify all possible Hilbert spaces, as follows:

3.2.15 Theorem. Let S be an arbitrary set, and let H be a Hilbert space with an orthonormal basis B having the same cardinality as S . Then there is an isometric isomorphism (a one-to-one-onto, linear, norm-preserving map) between H and $l^2(S)$.

PROOF. We may write $B = \{x_\alpha, \alpha \in S\}$. If $x \in H$, 3.2.13(d) then gives $x = \sum_\alpha \langle x, x_\alpha \rangle x_\alpha$, where $\sum_\alpha |\langle x, x_\alpha \rangle|^2 = \|x\|^2 < \infty$ by 3.2.13(f). The map $x \rightarrow (\langle x, x_\alpha \rangle, \alpha \in S)$ of H into $l^2(S)$ is therefore norm-preserving; since it is also linear, it must be one-to-one. To show that the map is onto, consider any

collection of complex numbers a_α , $\alpha \in S$, with $\sum_\alpha |a_\alpha|^2 < \infty$. Say $a_\alpha = 0$ except for $\alpha = \alpha_1, \alpha_2, \dots$, and let $x = \sum_\alpha a_{\alpha_j} x_{\alpha_j}$. [The series converges to an element of H because of the following fact, which occurs often enough to be stated separately: If $\{y_1, y_2, \dots\}$ is an orthonormal subset of H , the series $\sum_j c_j y_j$ converges to some element of H iff $\sum_j |c_j|^2 < \infty$. To see this, observe that $\|\sum_{j=1}^m c_j y_j\|^2 = \sum_{j=1}^m |c_j|^2 \|y_j\|^2 = \sum_{j=1}^m |c_j|^2$; thus the partial sums form a Cauchy sequence iff $\sum_j |c_j|^2 < \infty$.]

Since the x_α are orthonormal, it follows that $\langle x, x_\alpha \rangle = a_\alpha$ for all α , so that x maps onto $(a_\alpha, \alpha \in S)$. ■

We may also characterize Hilbert spaces that are separable, that is, have a countable dense set.

3.2.16 Theorem. A Hilbert space H is separable iff it has a countable orthonormal basis. If the orthonormal basis has n elements, H is isometrically isomorphic to C^n ; if the orthonormal basis is infinite, H is isometrically isomorphic to l^2 , that is, $l^2(S)$ with $S = \{1, 2, \dots\}$.

PROOF. Let B be an orthonormal basis for H . Now $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2$ for all $x, y \in B$, $x \neq y$, hence the balls $A_x = \{y: \|y - x\| < \frac{1}{2}\}$, $x \in B$, are disjoint. If D is dense in H , D must contain a point in each A_x , so that if B is uncountable, D must be also, and therefore H cannot be separable.

Now assume B is a countable set $\{x_1, x_2, \dots\}$. If U is a nonempty open subset of H [$= S(B) = \overline{L(B)}$], U contains an element of the form $\sum_{j=1}^n a_j x_j$ with the $a_j \in C$; in fact the a_j may be assumed to be rational, in other words, to have rational numbers as real and imaginary parts. Thus

$$D = \left\{ \sum_{j=1}^n a_j x_j : n = 1, 2, \dots, \text{ the } a_j \text{ rational} \right\}$$

is a countable dense set, so that H is separable. The remaining statements of the theorem follow from 3.2.15. ■

A linear norm-preserving map from one Hilbert space to another automatically preserves inner products; this is a consequence of the following proposition:

3.2.17 Polarization Identity. In any inner product space,

$$4\langle x, y \rangle = \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2.$$

PROOF.

$$\begin{aligned}\|x + y\|^2 &= \|x\|^2 + \|y\|^2 + 2 \operatorname{Re}\langle x, y \rangle \\ \|x - y\|^2 &= \|x\|^2 + \|y\|^2 - 2 \operatorname{Re}\langle x, y \rangle \\ \|x + iy\|^2 &= \|x\|^2 + \|y\|^2 + 2 \operatorname{Re}\langle x, iy \rangle \\ \|x - iy\|^2 &= \|x\|^2 + \|y\|^2 - 2 \operatorname{Re}\langle x, iy \rangle\end{aligned}$$

But $\operatorname{Re}\langle x, iy \rangle = \operatorname{Re}[-i\langle x, y \rangle] = \operatorname{Im}\langle x, y \rangle$, and the result follows. ■

Problems

1. In the Hilbert space $l^2(S)$, show that the elements e_α , $\alpha \in S$, form an orthonormal basis, where

$$e_\alpha(s) = \begin{cases} 0, & s \neq \alpha, \\ 1, & s = \alpha. \end{cases}$$

2. (a) If A is an arbitrary subset of the Hilbert space H , show that $A^{\perp\perp} = S(A)$.
(b) If M is a linear manifold of H , show that M is dense in H iff $M^\perp = \{0\}$.
3. Let x_1, \dots, x_n be elements of a Hilbert space. Show that the x_i are linearly dependent iff the Gramian (the determinant of the inner products $\langle x_i, x_j \rangle$, $i, j = 1, \dots, n$) is 0.
4. (Gram-Schmidt process) Let $B = \{x_1, x_2, \dots\}$ be a countable linearly independent subset of the Hilbert space H . Define $e_1 = x_1/\|x_1\|$; having chosen orthonormal elements e_1, \dots, e_n , let y_{n+1} be the projection of x_{n+1} on the space spanned by e_1, \dots, e_n :

$$y_{n+1} = \sum_{i=1}^n \langle x_{n+1}, e_i \rangle e_i.$$

Define

$$e_{n+1} = \frac{x_{n+1} - y_{n+1}}{\|x_{n+1} - y_{n+1}\|}.$$

- (a) Show that $L\{e_1, \dots, e_n\} = L\{x_1, \dots, x_n\}$ for all n , hence $x_{n+1} \neq y_{n+1}$ and the process is well defined.
(b) Show that the e_n form an orthonormal basis for $S(B)$.

Comments. Consider the space $H = L^2(-1, 1)$; if we take $x_n(t) = t^n$, $n = 0, 1, \dots$, the Gram-Schmidt process yields the *Legendre polynomials* $e_n(t) = a_n d^n[(t^2 - 1)^n]/dt^n$, where a_n is chosen so that $\|e_n\| = 1$. Similarly, if in $L^2(-\infty, \infty)$ we take $x_n(t) = t^n e^{-t^2/2}$, $n = 0, 1, \dots$, we obtain the *Hermite polynomials* $e_n(t) = a_n (-1)^n e^{t^2} d^n(e^{-t^2})/dt^n$.

5. (a) Show that every Hilbert space has an orthonormal basis.
 (b) Show that any two orthonormal bases have the same cardinality.
6. Let U be an open subset of the complex plane, and let $H(U)$ be the collection of all functions f analytic on U such that

$$\|f\|^2 = \iint_U |f(x+iy)|^2 dx dy < \infty.$$

If we define

$$\langle f, g \rangle = \iint_U f(x+iy)\bar{g}(x+iy) dx dy, \quad f, g \in H(U),$$

$H(U)$ becomes an inner product space.

- (a) If K is a compact subset of U and $f \in H(U)$, show that

$$\sup\{|f(z)| : z \in K\} \leq \|f\|/\sqrt{\pi d_0}$$

where d_0 is the Euclidean distance from K to the complement of U . Therefore convergence in $H(U)$ implies uniform convergence on compact subsets of U . (If $z \in K$, the Cauchy integral formula yields

$$f(z) = (2\pi)^{-1} \int_0^{2\pi} f(z + re^{i\theta}) d\theta, \quad r < d_0.$$

Integrate this equation with respect to r , $0 \leq r \leq d < d_0$. Note also that if U is the entire plane, we may take $d_0 = \infty$, and it follows that $H(C) = \{0\}$.)

- (b) Show that $H(U)$ is complete, and hence is a Hilbert space.
7. (a) If f is analytic on the unit disk $D = \{z : |z| < 1\}$ with Taylor expansion $f(z) = \sum_{n=0}^{\infty} a_n z^n$, show that

$$\sup_{0 \leq r < 1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta = \sum_{n=0}^{\infty} |a_n|^2.$$

It follows that if H^2 is the collection of all functions f analytic on D such that

$$N^2(f) = \sup_{0 \leq r < 1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta < \infty,$$

- then H^2 , with norm N , is a pre-Hilbert space.
- (b) If $f \in H^2$, show that

$$\iint_D |f(x+iy)|^2 dx dy \leq \pi N^2(f);$$

- hence $H^2 \subset H(D)$ and convergence in H^2 implies convergence in $H(D)$.
- If $f_n(z) = z^n$, $n = 0, 1, \dots$, show that $f_n \rightarrow 0$ in $H(D)$ but not in H^2 .
 - Show that H^2 is complete, and hence is a Hilbert space. [By (a), H^2 is isometrically isomorphic to a subspace of l^2 .]
 - If $e_n(z) = z^n$, $n = 0, 1, \dots$, show that the e_n form an orthonormal basis for H^2 .
 - If $e_n(z) = [(2n+2)/2\pi]^{1/2}z^n$, $n = 0, 1, \dots$, show that the e_n form an orthonormal basis for $H(D)$.
8. Let M be a closed convex subset of the Hilbert space H , and y_0 an element of M . If $x \in H$, show that
- $$\|x - y_0\| = \inf\{\|x - y\| : y \in M\}$$
- iff
- $$\operatorname{Re}\langle x - y_0, y - y_0 \rangle \leq 0 \quad \text{for all } y \in M.$$
9. (a) If g is a continuous complex-valued function on $[0, 2\pi]$ with $g(0) = g(2\pi)$, use the Stone–Weierstrass theorem to show that g can be uniformly approximated by trigonometric polynomials $\sum_{k=-n}^n c_k e^{ikt}$. Conclude that the trigonometric polynomials are dense in $L^2[0, 2\pi]$.
- (b) If $f \in L^2[0, 2\pi]$, show that the Fourier series $\sum_{n=-\infty}^{\infty} a_n e^{int}$, $a_n = (1/2\pi) \int_0^{2\pi} f(t) e^{-int} dt$, converges to f in L^2 , that is,
- $$\int_0^{2\pi} \left| f(t) - \sum_{k=-n}^n a_k e^{ikt} \right|^2 dt \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$
- (c) Show that $\{e^{int}/\sqrt{2\pi}, n = 0, \pm 1, \pm 2, \dots\}$ yields an orthonormal basis for $L^2[0, 2\pi]$.
10. (a) Give an example to show that if M is a nonempty, closed, but not convex subset of a Hilbert space H , there need not be an element of minimum norm in M . Thus the convexity hypothesis cannot be dropped from Theorem 3.2.9, even if we restrict ourselves to existence and forget about uniqueness.
- (b) Show that convexity is not necessary in the existence part of 3.2.9 if H is finite-dimensional.

3.3 Linear Operators on Normed Linear Spaces

The idea of a linear transformation from one Euclidean space to another is familiar. If A is a linear map from R^n to R^m , then A is completely specified by giving its values on a basis e_1, \dots, e_n : $A(\sum_{i=1}^n c_i e_i) = \sum_{i=1}^n c_i A(e_i)$;

furthermore, A is always continuous. If elements of R^n and R^m are represented by column vectors, A is represented by an $m \times n$ matrix. If $n = m$, so that A is a linear transformation on R^n , A is one-to-one iff it is onto, and if A^{-1} exists, it is always continuous (as well as linear).

Linear transformations on infinite-dimensional spaces have many features not found on the finite-dimensional case, as we shall see.

In this section, we study mappings A from one normed linear space L to another such space M . The mapping A will be a *linear operator*, that is, $A(ax + by) = aA(x) + bA(y)$ for all $x, y \in L$, $a, b \in C$. We use the symbol $\| \cdot \|$ for the norm on both spaces; no confusion should result. Linear operators can of course be defined on arbitrary vector spaces, but in this section, it is always understood that the domain and range are normed.

Linearity does not imply continuity; to study this idea, we introduce a new concept.

3.3.1 Definitions and Comments. If A is a linear operator, the *norm* of A is defined by:

- (a) $\|A\| = \sup\{\|Ax\| : x \in L, \|x\| \leq 1\}$. We may express $\|A\|$ in two other ways.
- (b) $\|A\| = \sup\{\|Ax\| : x \in L, \|x\| = 1\}$.
- (c) $\|A\| = \sup\{\|Ax\|/\|x\| : x \in L, x \neq 0\}$.

To see this, note that (b) \leq (a) is clear; if $x \neq 0$, then $\|Ax\|/\|x\| = \|A(x/\|x\|)\|$, and $x/\|x\|$ has norm 1; hence (c) \leq (b). Finally if $\|x\| \leq 1, x \neq 0$, then $\|Ax\| \leq \|Ax\|/\|x\|$, so (a) \leq (c).

It follows from (c) that $\|Ax\| \leq \|A\| \|x\|$, and in fact $\|A\|$ is the smallest number k such that $\|Ax\| \leq k\|x\|$ for all $x \in L$.

The linear operator A is said to be *bounded* iff $\|A\| < \infty$. Boundedness is often easy to check, a very fortunate circumstance because we can show that boundedness is equivalent to continuity.

3.3.2 Theorem. A linear operator A is continuous iff it is bounded.

PROOF. If A is bounded and $\{x_n\}$ is a sequence in L converging to 0, then $\|Ax_n\| \leq \|A\| \|x_n\| \rightarrow 0$. (We use here the fact that the mapping $x \rightarrow \|x\|$ of L into the nonnegative reals is continuous; this follows because $|\|x\| - \|y\|| \leq \|x - y\|$.) Thus A is continuous at 0, and therefore, by linearity, is continuous everywhere. On the other hand, if A is unbounded, we can find elements $x_n \in L$ with $\|x_n\| \leq 1$ and $\|Ax_n\| \rightarrow \infty$. Let $y_n = x_n/\|Ax_n\|$; then $y_n \rightarrow 0$, but $\|Ay_n\| = 1$ for all n , hence Ay_n does not converge to 0. Consequently, A is discontinuous. ■

We are going to show that the set of all bounded linear operators from L to M is itself a normed linear space, but first we consider some examples:

3.3.3 Examples. (a) Let $L = C[a, b]$, the set of all continuous complex-valued functions on the closed bounded interval $[a, b]$ of reals. Put the *sup norm* on L :

$$\|x\| = \sup\{|x(t)| : a \leq t \leq b\}, \quad x \in C[a, b].$$

With this norm, L is a Banach space [see 3.1.2(a)]. Let $K = K(s, t)$ be a continuous complex-valued function on $[a, b] \times [a, b]$, and define a linear operator on L by

$$(Ax)(s) = \int_a^b K(s, t)x(t) dt, \quad a \leq s \leq b.$$

(By the dominated convergence theorem, Ax actually belongs to L .) We show that A is bounded:

$$\begin{aligned} \|Ax\| &= \sup_{a \leq s \leq b} \left| \int_a^b K(s, t)x(t) dt \right| \\ &\leq \sup_{a \leq t \leq b} |x(t)| \sup_{a \leq s \leq b} \int_a^b |K(s, t)| dt \\ &= \|x\| \max_{a \leq s \leq b} \int_a^b |K(s, t)| dt \end{aligned}$$

(note that $\int_a^b |K(s, t)| dt$ is a continuous function of s , by the dominated convergence theorem). Thus

$$\|A\| \leq \max_{a \leq s \leq b} \int_a^b |K(s, t)| dt < \infty.$$

In fact $\|A\| = \max_{a \leq s \leq b} \int_a^b |K(s, t)| dt$ (see Problem 3).

(b) Let $L = l^p(S)$, where S is the set of all integers and $1 \leq p < \infty$. Define a linear operator T on L by $(Tf)_n = f_{n+1}$, $n \in S$; T is called the *two-sided shift* or the *bilateral shift*. It follows from the definition that T is one-to-one onto, and $(T^{-1}f)_n = f_{n-1}$, $n \in S$. Also, $\|Tf\| = \|f\|$ for all $f \in L$; hence $\|T^{-1}f\| = \|f\|$ for all $f \in L$, so that T is an isometric isomorphism of L with itself. (In particular, $\|T\| = \|T^{-1}\| = 1$.)

If we replace S by the positive integers and define T as above, the resulting operator is called the *one-sided* or *unilateral* shift. The one-sided shift is onto with norm 1, but is not one-to-one; $T(f_1, f_2, \dots) = (f_2, f_3, \dots)$.

The shift operators we have defined are shifts to the left. We may also define shifts to the right; in the two-sided case we take $(Af)_n = f_{n-1}$, $n \in S$ (so that $A = T^{-1}$). In the one-sided case, we set $(Af)_n = f_{n-1}$, $n \geq 2$; $(Af)_1 = 0$;

thus $A(f_1, f_2, \dots) = (0, f_1, f_2, \dots)$. The operator A is one-to-one but not onto; $A(L)$ is the closed subspace of L consisting of those sequences whose first coordinate is 0.

(c) Assume that the Banach space L is the direct sum of the two closed subspaces M and N , in other words each $x \in L$ can be represented in a unique way as $y + z$ for some $y \in M$, $z \in N$. (We have already encountered this situation with L a Hilbert space, M a closed subspace, $N = M^\perp$.) We define a linear operator P on L by

$$Px = y.$$

P is called a *projection*; specifically, P is the projection of L on M . P has the following properties:

(1) P is *idempotent*; that is, $P^2 = P$, where P^2 is the composition of P with itself.

(2) P is continuous.

Property (1) follows from the definition of P ; property (2) will be proved later, as a consequence of the closed graph theorem (see after 3.4.16).

Conversely, let P be a continuous idempotent linear operator on L . Define

$$M = \{x \in L : Px = x\}, \quad N = \{x \in L : Px = 0\}.$$

Then we can prove that M and N are closed subspaces, L is the direct sum of M and N , and P is the projection of L on M .

By continuity of P , M and N are closed subspaces. If $x \in L$, then $x = Px + (I - P)x = y + z$ where $y \in M$, $z \in N$. Since $M \cap N = \{0\}$, L is the direct sum of M and N . Furthermore, $Px = y$ by definition of y , so that P is the projection of L on M .

If f is a linear operator from a vector space L to the scalar field, f is called a *linear functional*. (The norm of a scalar b is taken as $|b|$.) Considerable insight is gained about normed linear spaces by studying ways of representing continuous linear functionals on such spaces. We give some examples:

3.3.4 Representations of Continuous Linear Functionals. (a) Let f be a continuous linear functional on the Hilbert space H . We show that there is a unique element $y \in H$ such that

$$f(x) = \langle x, y \rangle \quad \text{for all } x \in H.$$

This is one of several results called the *Riesz representation theorem*.

If the desired y exists, it must be unique, for if $\langle x, y \rangle = \langle x, z \rangle$ for all x , then $y - z$ is orthogonal to everything in H (including itself), so $y = z$.

To prove existence, let N be the *null space* of f , that is, $N = \{x \in H : f(x) = 0\}$. If $N^\perp = \{0\}$, then $N = H$ by the projection theorem; hence $f \equiv 0$, and we may take $y = 0$. Thus assume we have an element $u \in N^\perp$ with $u \neq 0$. Then $u \notin N$, and if we define $z = u/f(u)$, we have $z \in N^\perp$ and $f(z) = 1$.

If $x \in H$ and $f(x) = a$, then

$$x = (x - az) + az, \quad \text{with } x - az \in N, \quad az \perp N.$$

If $y = z/\|z\|^2$, then

$$\begin{aligned} \langle x, y \rangle &= \langle x - az, y \rangle + a\langle z, y \rangle \\ &= a\langle z, y \rangle \quad \text{since } y \perp N \\ &= a = f(x) \end{aligned}$$

as desired.

The above argument shows that if f is not identically 0, then N^\perp is one-dimensional. For if $x \in N^\perp$ and $f(x) = a$, then $x - az \in N \cap N^\perp$, hence $x = az$. Therefore $N^\perp = \{az : a \in C\}$.

Notice also that if $\|f\|$ is the norm of f , considered as a linear operator, then

$$\|f\| = \|y\|.$$

For $|f(x)| = |\langle x, y \rangle| \leq \|x\| \|y\|$ for all x , hence $\|f\| \leq \|y\|$; but $|f(y)| = |\langle y, y \rangle| = \|y\| \|y\|$, so $\|f\| \geq \|y\|$.

Now consider the space H^* of all continuous linear functionals on H ; H^* is a vector space under the usual operations of addition and scalar multiplication. If to each $f \in H^*$ we associate the element of H given by the Riesz representation theorem, we obtain a map $\psi : H^* \rightarrow H$ that is one-to-one onto, norm-preserving, and *conjugate linear*; that is,

$$\psi(af + bg) = \bar{a}\psi(f) + \bar{b}\psi(g), \quad f, g \in H^*, \quad a, b \in C.$$

[Note that if $f(x) = \langle x, y \rangle$ for all x , then $af(x) = \langle x, \bar{a}y \rangle$.] Such a map is called a *conjugate isometry*.

(b) Let f be a continuous linear functional on l^p ($= l^p(S)$, where S is the set of positive integers), $1 < p < \infty$. We show that if q is defined by $(1/p) + (1/q) = 1$, there is a unique element $y = (y_1, y_2, \dots) \in l^q$ such that

$$f(x) = \sum_{k=1}^{\infty} x_k y_k \quad \text{for all } x \in l^p.$$

Furthermore,

$$\|f\| = \|y\| = \left(\sum_{k=1}^{\infty} |y_k|^q \right)^{1/q}.$$

To prove this, let e_n be the sequence in l^p defined by $e_n(j) = 0$, $j \neq n$; $e_n(n) = 1$. If $x \in l^p$, then $\|x - \sum_{k=1}^n x_k e_k\|^p = \sum_{k=n+1}^{\infty} |x_k|^p \rightarrow 0$; hence $x = \sum_{k=1}^{\infty} x_k e_k$, where the series converges in l^p . By continuity of f ,

$$f(x) = \sum_{k=1}^{\infty} x_k y_k, \quad \text{where } y_k = f(e_k). \quad (1)$$

Now write y_k in polar form, that is, $y_k = r_k e^{i\theta_k}$, $r_k \geq 0$. Let

$$z_n = (r_1^{q-1} e^{-i\theta_1}, \dots, r_n^{q-1} e^{-i\theta_n}, 0, 0, \dots);$$

by (1),

$$f(z_n) = \sum_{k=1}^n r_k^{q-1} e^{-i\theta_k} r_k e^{i\theta_k} = \sum_{k=1}^n |y_k|^q. \quad (2)$$

But

$$\begin{aligned} |f(z_n)| &\leq \|f\| \|z_n\| \\ &= \|f\| \left(\sum_{k=1}^n |y_k|^{(q-1)p} \right)^{1/p} \\ &= \|f\| \left(\sum_{k=1}^n |y_k|^q \right)^{1/p} \end{aligned}$$

By Eq. (2),

$$\left(\sum_{k=1}^n |y_k|^q \right)^{1/q} \leq \|f\|;$$

hence $y \in l^q$ and $\|y\| \leq \|f\|$. To prove that $\|f\| \leq \|y\|$, observe that Eq. (1) is of the form $f(x) = \int_{\Omega} xy \, d\mu$ where Ω is the positive integers and μ is counting measure. By Hölder's inequality, $|f(x)| \leq \|x\| \|y\|$, so that $\|f\| \leq \|y\|$.

Finally, we prove uniqueness. If $y, z \in l^q$ and $f(x) = \sum_{k=1}^{\infty} x_k y_k = \sum_{k=1}^{\infty} x_k z_k$ for all $x \in l^p$, then $g(x) = \sum_{k=1}^{\infty} x_k (y_k - z_k) = 0$ for all $x \in l^p$. The above argument with f replaced by g shows that $\|g\| = \|y - z\|$; hence $\|y - z\| = 0$, and thus $y = z$.

(c) Let f be a continuous linear functional on l^1 . We show that there is a unique element $y \in l^{\infty}$ such that

$$f(x) = \sum_{k=1}^{\infty} x_k y_k \quad \text{for all } x \in l^1.$$

The argument of (b) may be repeated up to Eq. (1). In this case, however, if $y_k = r_k e^{i\theta_k}$, $k = 1, 2, \dots$, we take

$$z_n = (0, \dots, 0, e^{-i\theta_n}, 0, \dots), \quad \text{with } e^{-i\theta_n} \text{ in position } n.$$

Thus by (1),

$$f(z_n) = e^{-i\theta_n} r_n e^{i\theta_n} = |y_n|.$$

But

$$|f(z_n)| \leq \|f\| \|z_n\| = \|f\|.$$

Therefore $|y_n| \leq \|f\|$ for all n , so that $y \in l^\infty$ and $\|y\| \leq \|f\|$. But by (1),

$$|f(x)| \leq \left(\sup_k |y_k| \right) \sum_{k=1}^{\infty} |x_k| = \|y\| \|x\|;$$

hence $\|f\| \leq \|y\|$. Uniqueness is proved as in (b).

In 3.3.4(b) and (c), the map $f \rightarrow y$ of $(l^p)^*$ to l^q [$q = \infty$ in 3.3.4(c)] is linear, and is one-to-one by uniqueness of y . To show that it is onto, observe that any linear functional of the form $f(x) = \sum_{k=1}^{\infty} x_k y_k$, $x \in l^p$, $y \in l^q$, satisfies $\|f\| \leq \|y\|$ by the analysis of 3.3.4(b) and (c). Therefore f is continuous, so that every $y \in l^q$ is the image of some $f \in (l^p)^*$. Since $\|f\| = \|y\|$, we have an isometric isomorphism of $(l^p)^*$ and l^q .

If we replace y_k by \bar{y}_k , we obtain the result that there is a unique $y \in l^q$ such that $f(x) = \sum_{k=1}^{\infty} x_k \bar{y}_k$ for all $x \in l^p$. This makes the map $f \rightarrow y$ a conjugate isometry rather than an isometric isomorphism. If $p = 2$, then $q = 2$ also, and thus we have another proof of the Riesz representation theorem for separable Hilbert spaces (see 3.2.16). In fact, essentially the same argument may be used in an arbitrary Hilbert space if the e_n are replaced by an arbitrary orthonormal basis. (Other examples of representation of continuous linear functionals are given in Problems 10 and 11.)

We now show that the set of bounded linear operators from one normed linear space to another can be made into a normed linear space.

3.3.5 Theorem. Let L and M be normed linear spaces, and let $[L, M]$ be the collection of all bounded linear operators from L to M . The operator norm defined by 3.3.1 is a norm on $[L, M]$, and if M is complete, then $[L, M]$ is complete. In particular, the set L^* of all continuous linear functionals on L is a Banach space (whether or not L is complete).

PROOF. It follows from 3.3.1 that $\|A\| \geq 0$ and $\|aA\| = |a| \|A\|$ for all $A \in [L, M]$ and $a \in C$. Also by 3.3.1, if $\|A\| = 0$, then $Ax = 0$ for all $x \in L$, hence $A = 0$. If $A, B \in [L, M]$, then again by 3.3.1,

$$\|A + B\| = \sup\{\|(A + B)x\| : x \in L, \|x\| \leq 1\}.$$

Since $\|(A + B)x\| \leq \|Ax\| + \|Bx\|$ for all x ,

$$\|A + B\| \leq \|A\| + \|B\|$$

and it follows that $[L, M]$ is a vector space and the operator norm is in fact a norm on $[L, M]$.

Now let A_1, A_2, \dots be a Cauchy sequence in $[L, M]$. Then

$$\|(A_n - A_m)x\| \leq \|A_n - A_m\| \|x\| \rightarrow 0 \quad \text{as } n, m \rightarrow \infty. \quad (1)$$

Therefore $\{A_n x\}$ is a Cauchy sequence in M for each $x \in L$, hence A_n converges pointwise on L to an operator A . Since the A_n are linear, so is A (observe that $A_n(ax + by) = aA_n x + bA_n y$, and let $n \rightarrow \infty$). Now given $\varepsilon > 0$, choose N such that $\|A_n - A_m\| \leq \varepsilon$ for $n, m \geq N$. Fix $n \geq N$ and let $m \rightarrow \infty$ in Eq. (1) to conclude that $\|(A_n - A)x\| \leq \varepsilon \|x\|$ for $n \geq N$; therefore $\|A_n - A\| \rightarrow 0$ as $n \rightarrow \infty$. Since $\|A\| \leq \|A - A_n\| + \|A_n\|$, we have $A \in [L, M]$ and $A_n \rightarrow A$ in the operator norm. ■

In the above proof we have talked about two different types of convergence of sequences of operators.

3.3.6 Definitions and Comments. Let $A, A_1, A_2, \dots \in [L, M]$. We say that A_n converges *uniformly* to A iff $\|A_n - A\| \rightarrow 0$ (notation: $A_n \xrightarrow{u} A$). Since $\|(A_n - A)x\| \leq \|A_n - A\| \|x\|$, uniform operator convergence means that $A_n x \rightarrow Ax$, uniformly for $\|x\| \leq 1$ (or equally well for $\|x\| \leq k$, k any positive real number).

We say that A_n converges *strongly* to A (notation: $A_n \xrightarrow{s} A$) iff $A_n x \rightarrow Ax$ for each $x \in L$. Thus strong operator convergence is pointwise convergence on all of L .

Uniform convergence implies strong convergence, but not conversely. For example, let $\{e_1, e_2, \dots\}$ be an orthonormal set in a Hilbert space, and let $A_n x = \langle x, e_n \rangle e_n$, $n = 1, 2, \dots$. Then $A_n \xrightarrow{s} 0$ by Bessel's inequality, but A_n does not converge uniformly to 0. In fact $\|A_n e_n\| = 1$ for all n , hence $\|A_n\| \equiv 1$.

There is an important property of finite-dimensional spaces that we are now in a position to discuss. In the previous section, we regarded C^n as a Hilbert space, so that if $x \in C^n$, the norm of x was taken as the Euclidean norm $\|x\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$. The metric associated with this norm yields the standard topology on C^n . However, we may put various other norms on C^n , for example, the L^p norm $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, $1 \leq p < \infty$, or the sup norm $\|x\|_\infty = \max(|x_1|, \dots, |x_n|)$. Since the space is finite-dimensional, there are no convergence difficulties and all elements have finite norm. Fortunately, the proliferation of norms causes no confusion because *all norms on a given finite-dimensional space induce the same topology*. The proof of this result is outlined in Problems 6 and 7.

Problems

1. Let f be a linear functional on the normed linear space L . If f is not identically 0, show that the following are equivalent:
 - (a) f is continuous.
 - (b) The null space $N = f^{-1}\{0\}$ is closed.
 - (c) N is not dense in L .
 - (d) f is bounded on some neighborhood of 0.

[To prove that (c) implies (d), show that if $B(x, \varepsilon) \cap N = \emptyset$, and f is unbounded on $B(0, \varepsilon)$, then $f(B(0, \varepsilon)) = C$. In particular, there is a point $z \in B(0, \varepsilon)$ such that $f(z) = -f(x)$, and this leads to a contradiction.]
2. Show that any infinite-dimensional normed linear space had a discontinuous linear functional. (Let e_1, e_2, \dots be an infinite sequence of linearly independent elements such that $\|e_n\| = 1$ for all n . Define f appropriately on the e_n and extend f to the whole space using linearity.)
3. In Example 3.3.3(a), show that $\|A\| = \max_{a \leq s \leq b} \int_a^b |K(s, t)| dt$. [If $\int_a^b |K(s, t)| dt$ assumes a maximum at the point u , and $K(u, t) = r(t)e^{i\theta(t)}$, let $z(t) = r(t)e^{-i\theta(t)}$. Let x_1, x_2, \dots be continuous functions such that $|x_n(t)| \leq 1$ for all n and t , and $\int_a^b |x_n(t) - z(t)| dt \rightarrow 0$ as $n \rightarrow \infty$. Since $\|Ax_n\| \leq \|A\|$ and $(Ax_n)(s) \rightarrow \int_a^b K(s, t)z(t) dt$ as $n \rightarrow \infty$, it follows that $\int_a^b |K(u, t)| dt \leq \|A\|$.]

The same argument, with integrals replaced by sums, shows that if A is a matrix operator on C^n , with sup norm, $\|A\| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$.

4. Let M be a linear manifold in the normed linear space L . Denote by $[x]$ the coset of x modulo M , that is, $\{x + y : y \in M\}$. Define

$$\|[x]\| = \inf\{\|y\| : y \in [x]\};$$

note that $\|[x]\| = \inf\{\|x - z\| : z \in M\} = \text{dist}(x, M)$. Show that the above formula defines a seminorm on the quotient space L/M , a norm if M is closed.

5. If L is a Banach space and M is a closed subspace, show that L/M is a Banach space.
6. (a) Let A be a linear operator from L to M . Show that A is one-to-one with A^{-1} continuous on its domain $A(L)$, iff there is a finite number $m > 0$ such that $\|Ax\| \geq m\|x\|$ for all $x \in L$.
- (b) Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on the linear space L . Show that the norms induce the same topology iff there are finite numbers $m, M > 0$ such that

$$m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1 \quad \text{for all } x \in L.$$

(This may be done using part (a), or it may be shown directly that, for example, if $\|x\|_1 \leq (1/m)\|x\|_2$ for all x , then the topology induced by $\|\cdot\|_1$ is weaker than (that is, included in) the topology induced by $\|\cdot\|_2$.)

7. Let L be a finite-dimensional normed linear space, with basis e_1, \dots, e_n . Let $\|\cdot\|_1$ be any norm on L , and define

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}, \quad \text{where } x = \sum_{i=1}^n x_i e_i.$$

- (a) Show that for some positive real number k , $\|x\|_1 \leq k \|x\|_2$ for all $x \in L$.
 - (b) Show that for some positive real number m , $\|x\|_1 \geq m$ on $\{x : \|x\|_2 = 1\}$. [By (a), the map $x \rightarrow \|x\|_1$ is continuous in the topology induced by $\|\cdot\|_2$.]
 - (c) Show that $\|x\|_1 \geq m \|x_2\|$ for all $x \in L$, and conclude that all norms on a finite-dimensional space induce the same topology.
 - (d) Let M be an arbitrary normed linear space, and let L be a finite-dimensional subspace of M . Show that L is closed in M .
8. (*Riesz lemma*) Let M be a closed proper subspace of the normed linear space L . Show that if $0 < \delta < 1$, there is an element $x_\delta \in L$ such that $\|x_\delta\| = 1$ and $\|x - x_\delta\| \geq \delta$ for all $x \in M$. [Choose $x_1 \notin M$, and let $d = \text{dist}(x_1, M) > 0$ since M is closed. Choose $x_0 \in M$ such that $\|x_1 - x_0\| \leq d/\delta$, and set $x_\delta = (x_1 - x_0)/\|x_1 - x_0\|$.]
9. Let L be a normed linear space. Show that the following are equivalent:
- (a) L is finite-dimensional.
 - (b) L is topologically isomorphic to a Euclidean space C^n (or R^n if L is a real vector space), that is, there is a one-to-one, onto, linear, bicontinuous map between the two spaces.
 - (c) L is locally compact (every point of L has a neighborhood whose closure is compact).
 - (d) Every closed bounded subset of L is compact.
 - (e) The set $\{x \in L : \|x\| = 1\}$ is compact.
 - (f) The set $\{x \in L : \|x\| = 1\}$ is totally bounded, that is, can be covered by a finite number of open balls of any preassigned radius.
- [Problem 7 shows that (a) implies (b); (b) implies (c), (d) implies (e), and (e) implies (f) are obvious, and (c) implies (d) is easy. To prove that (f) implies (a), use Problem 8.]
10. Let c be the space of convergent sequences of complex numbers [see 3.1.2(b)]. If f is a continuous linear functional on c , show that there is a unique element $y = (y_0, y_1, \dots) \in l^1$ such that for all $x \in c$,

$$f(x) = \left(\lim_{n \rightarrow \infty} x_n \right) \left(y_0 - \sum_{k=1}^{\infty} y_k \right) + \sum_{k=1}^{\infty} x_k y_k.$$

Furthermore, $\|f\| = |y_0 - \sum_{k=1}^{\infty} y_k| + \sum_{k=1}^{\infty} |y_k|$. If c_0 is the closed subspace of c consisting of those sequences converging to 0, the representation of a continuous linear functional on c_0 is simpler: $f(x) = \sum_{k=1}^{\infty} x_k y_k$, $\|f\| = \sum_{k=1}^{\infty} |y_k|$. Thus c_0^* is isometrically isomorphic to L^1 .

11. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

- (a) Assume μ finite, $1 < p < \infty$, $(1/p) + (1/q) = 1$. If f is a continuous linear functional on $L^p = L^p(\Omega, \mathcal{F}, \mu)$ show that there is an element $y \in L^q$ such that

$$f(x) = \int_{\Omega} xy \, d\mu \quad \text{for all } x \in L^p.$$

Furthermore, $\|f\| = \|y\|_q$. If y_1 is another such function, show that $y = y_1$ a.e. $[\mu]$. [Define $\lambda(A) = f(I_A)$, $A \in \mathcal{F}$, and apply the Radon–Nikodym theorem.]

- (b) Drop the finiteness assumption on μ . If $A \in \mathcal{F}$ and $\mu(A) < \infty$, part (a) applied to $(A, \mathcal{F} \cap A, \mu)$ provides an essentially unique $y_A \in L^q$ such that $y_A = 0$ on A^c and

$$f(xI_A) = \int_{\Omega} xy_A \, d\mu \quad \text{for all } x \in L^p;$$

also, $\|y_A\|_q$ is the norm of the restriction of f to $L^p(A)$, so

$$\|y_A\|_q \leq \|f\|.$$

- (i) If $\mu(A) < \infty$ and $\mu(B) < \infty$, show that $y_A = y_B$ a.e. $[\mu]$ on $A \cap B$. Thus $y_{A \cup B}$ may be obtained by piecing together y_A and y_B .
- (ii) Let A_n , $n = 1, 2, \dots$, be sets with $\|y_{A_n}\|_q \rightarrow k = \sup\{\|y_A\|_q : A \in \mathcal{F}, \mu(A) < \infty\} \leq \|f\|$. Show that y_{A_n} converges in L_q to a limit function y , and y is essentially independent of the particular sequence $\{A_n\}$. Furthermore, $\|y\|_q \leq \|f\|$.
- (iii) Show that $f(x) = \int_{\Omega} xy \, d\mu$ for all $x \in L^p$. Since $\|y\|_q \leq \|f\|$ by (ii) and $\|f\| \leq \|y\|_q$ by Hölder's inequality, we have $\|f\| = \|y\|_q$. Thus the result of (a) holds for arbitrary μ .
- (c) Prove (a) with μ finite, $p = 1, q = \infty$.
- (d) Prove (a) with μ σ -finite, $p = 1, q = \infty$. [For an extension of this result, see Kelley and Namioka (1963, Problem 14M).]

It follows that there is an isometric isomorphism of $(L^p)^*$ and L^q if $1 < p < \infty$, $1 < q < \infty$, $(1/p) + (1/q) = 1$; if μ is σ -finite, this is true also for $p = 1, q = \infty$.

3.4 Basic Theorems of Functional Analysis

Almost every area of functional analysis leans heavily on at least one of the three basic results of this section: the Hahn–Banach theorem, the uniform boundedness principle, and the open mapping theorem. We are going to establish these results and discuss applications.

We first consider an extension problem. If f is a linear functional defined on a subspace M of a vector space L , there is no difficulty in extending f to a linear functional on all of L ; simply extend a Hamel basis of M to a Hamel basis for L , define f arbitrarily on the basis vectors not belonging to M , and extend by linearity. However, if L is normed and we require that the extension of f have the same norm as the original functional, the problem becomes more difficult. We first prove a preliminary result.

3.4.1 Lemma. Let L be a real vector space, not necessarily normed, and let p be a map from L to R satisfying

$$\begin{aligned} p(x+y) &\leq p(x) + p(y) && \text{for all } x, y \in L \\ p(ax) &= ap(x) && \text{for all } x \in L \text{ and all } a > 0. \end{aligned}$$

[The first property is called *subadditivity*, the second *positive-homogeneity*. Note that positive-homogeneity implies that $p(0) = 0$ [set $x = 0$ to obtain $p(0) = ap(0)$ for all $a > 0$]. A subadditive, positive-homogeneous map is sometimes called a *sublinear functional*.]

Let M be a subspace of L and g a linear functional defined on M such that $g(x) \leq p(x)$ for all $x \in M$. Let x_0 be a fixed element of L . For any real number c , the following are equivalent:

- (1) $g(x) + \lambda c \leq p(x + \lambda x_0)$ for all $x \in M$ and all $\lambda \in R$.
- (2) $-p(-x - x_0) - g(x) \leq c \leq p(x + x_0) - g(x)$ for all $x \in M$.

Furthermore, there is a real number c satisfying (2), and hence (1).

PROOF. To prove that (1) implies (2), first set $\lambda = 1$, and then set $\lambda = -1$ and replace x by $-x$ [note $g(-x) = -g(x)$ by linearity]. Conversely, if (2) holds and $\lambda > 0$, replace x by x/λ in the right-hand inequality of (2); if $\lambda < 0$, replace x by x/λ in the left-hand inequality. In either case, the positive homogeneity of p yields (1). If $\lambda = 0$, (1) is true by hypothesis.

To produce the desired c , let x and y be arbitrary elements of M . Then

$$\begin{aligned} g(x) - g(y) &= g(x - y) \leq p(x - y) \\ &\leq p(x + x_0) + p(-y - x_0) && \text{by subadditivity.} \end{aligned}$$

It follows that

$$\sup_{y \in M} [-p(-y - x_0) - g(y)] \leq \inf_{x \in M} [p(x + x_0) - g(x)].$$

Any c between the sup and the inf will work. ■

We may now prove the main extension theorem.

3.4.2 Hahn–Banach Theorem. Let p be a subadditive, positive-homogeneous functional on the real linear space L , and g a linear functional on the subspace M , with $g \leq p$ on M . There is a linear functional f on L such that $f = g$ on M and $f \leq p$ on all of L .

PROOF. If $x_0 \notin M$, consider the subspace $M_1 = L(M \cup \{x_0\})$, consisting of all elements $x + \lambda x_0$, $x \in M$, $\lambda \in R$. We may extend g to a linear functional on M_1 by defining $g_1(x + \lambda x_0) = g(x) + \lambda c$, where c is any real number. If we choose c to satisfy (1) of 3.4.1, then $g_1 \leq p$ on M_1 .

Now let \mathcal{C} be the collection of all pairs (h, H) where h is an extension of g to the subspace $H \supset M$, and $h \leq p$ on H . Partially order \mathcal{C} by $(h_1, H_1) \leq (h_2, H_2)$ iff $H_1 \subset H_2$ and $h_1 = h_2$ on H_1 ; then every chain in \mathcal{C} has an upper bound (consider the union of all subspaces in the chain). By Zorn's lemma, \mathcal{C} has a maximal element (f, F) . If $F \neq L$, the first part of the proof yields an extension of f to a larger subspace, contradicting maximality. ■

There is a version of the Hahn–Banach theorem for complex spaces. First, we observe that if L is a vector space over C , L is automatically a vector space over R , since we may restrict scalar multiplication to real scalars. For example, C^n is an n -dimensional space over C , with basis vectors $(1, 0, \dots, 0)$, $(0, \dots, 0, 1)$. If C^n is regarded as a vector space over R , it becomes $2n$ -dimensional, with basis vectors

$$(1, 0, \dots, 0), \dots, (0, \dots, 0, 1), \quad (i, 0, \dots, 0), \dots, (0, \dots, 0, i).$$

Now if f is a linear functional on L , with $f_1 = \operatorname{Re} f$, $f_2 = \operatorname{Im} f$, then f_1 and f_2 are linear functionals on L' , where L' is L regarded as a vector space over R . Also, for all $x \in L$,

$$f(ix) = f_1(ix) + if_2(ix).$$

But

$$f(ix) = if(x) = -f_2(x) + if_1(x).$$

Thus $f_1(ix) = -f_2(x)$, $f_2(ix) = f_1(x)$; consequently

$$f(x) = f_1(x) - if_1(ix) = f_2(ix) + if_2(x).$$

Therefore f is determined by f_1 (or by f_2). Conversely, let f_1 be a linear functional on L' . Then f_1 is a map from L to R such that $f_1(ax + by) = af_1(x) + bf_1(y)$ for all $x, y \in L$ and all $a, b \in R$. Define $f(x) = f_1(x) - if_1(ix)$, $x \in L$. It follows that f is a linear functional on L (and $f_1 = \operatorname{Re} f$). For f_1 is additive, hence so is f , and if $a, b \in R$, we have

$$\begin{aligned} f((a+ib)x) &= f_1(ax+ibx) - if_1(-bx+iax) \\ &= af_1(x) + bf_1(ix) + ibf_1(x) - iaf_1(ix) \\ &= (a+ib)[f_1(x) - if_1(ix)] \\ &= (a+ib)f(x). \end{aligned}$$

Note that homogeneity of f_1 does not immediately imply homogeneity of f . For $f_1(\lambda x) = \lambda f_1(x)$ for real λ but not in general for complex λ . [For example, let $L = C$, $f(x) = x$, $f_1(x) = \operatorname{Re} x$.]

We now prove the complex version of the Hahn–Banach theorem.

3.4.3 Theorem. Let L be a vector space over C , and p a seminorm on L . If g is a linear functional on the subspace M , and $|g| \leq p$ on M , there is a linear functional f on L such that $f = g$ on M and $|f| \leq p$ on L .

PROOF. Since p is a seminorm, it is subadditive and absolutely homogeneous, and hence positive-homogeneous. If $g_1 = \operatorname{Re} g$, then $g_1 \leq |g| \leq p$; so by 3.4.2, there is an extension of g_1 to a linear map f_1 of L to R such that $f_1 = g_1$ on M and $f_1 \leq p$ on L . Define $f(x) = f_1(x) - if_1(ix)$, $x \in L$. Then f is a linear functional on L and $f = g$ on M . Fix $x \in L$, and let $f(x) = re^{i\theta}$, $r \geq 0$. Then

$$\begin{aligned} |f(x)| &= r = f(e^{-i\theta}x) \\ &= f_1(e^{-i\theta}x) \quad \text{since } r \text{ is real} \\ &\leq p(e^{-i\theta}x) \quad \text{since } f_1 \leq p \text{ on } L \\ &= p(x) \quad \text{by absolute homogeneity.} \blacksquare \end{aligned}$$

3.4.4 Corollary. Let g be a continuous linear functional on the subspace M of the normed linear space L . There is an extension of g to a continuous linear functional f on L such that $\|f\| = \|g\|$.

PROOF. Let $p(x) = \|g\| \|x\|$; then p is a seminorm on L and $|g| \leq p$ on M by definition of $\|g\|$. The result follows from 3.4.3. \blacksquare

A direct application of the Hahn–Banach theorem is the result that in a normed linear space, there are enough continuous linear functionals to distinguish points; in other words, if $x \neq y$, there is a continuous linear functional f such that $f(x) \neq f(y)$. We now prove this, along with other related results.

3.4.5 Theorem. Let M be a subspace of the normed linear space L , and let L^* be the collection of all continuous linear functionals on L .

- (a) If $x_0 \notin \bar{M}$, there is an $f \in L^*$ such that $f = 0$ on M , $f(x_0) = 1$, and $\|f\| = 1/d$, where d is the distance from x_0 to M .
- (b) $x_0 \in \bar{M}$ iff every $f \in L^*$ that vanishes on M also vanishes at x_0 .
- (c) If $x_0 \neq 0$, there is an $f \in L^*$ such that $\|f\| = 1$ and $f(x_0) = \|x_0\|$; thus the maximum value of $|f(x)|/\|x\|$, $x \neq 0$, is achieved at x_0 . In particular, if $x \neq y$, there is an $f \in L^*$ such that $f(x) \neq f(y)$.

PROOF. (a) First note that $L(M \cup \{x_0\})$ is the set of all elements $y = x + ax_0$, $x \in M$, $a \in C$, and since $x_0 \notin M$, a is uniquely determined by y . Define f on $N = L(M \cup \{x_0\})$ by $f(x + ax_0) = a$; f is linear, and furthermore, $\|f\| = 1/d$, as we now prove. By 3.3.1 we have

$$\begin{aligned} \|f\| &= \sup \left\{ \frac{|f(y)|}{\|y\|} : y \in N, \quad y \neq 0 \right\} \\ &= \sup \left\{ \frac{|a|}{\|x + ax_0\|} : x \in M, \quad a \in C, \quad x \neq 0 \quad \text{or} \quad a \neq 0 \right\} \\ &= \sup \left\{ \frac{|a|}{\|x + ax_0\|} : x \in M, \quad a \in C, \quad a \neq 0 \right\} \end{aligned}$$

since $f(y) = 0$ when $a = 0$. Now

$$\frac{|a|}{\|x + ax_0\|} = \frac{1}{\left\| x_0 + \frac{x}{a} \right\|} = \frac{1}{\|x_0 - z\|} \quad \text{for some } z \in M;$$

hence $\|f\| = (\inf\{\|x_0 - z\| : z \in M\})^{-1} = 1/d < \infty$. The result now follows from 3.4.4.

- (b) This is immediate from (a).
- (c) Apply (a) with $M = \{0\}$, to obtain $g \in L^*$ with $g(x_0) = 1$ and $\|g\| = 1/\|x_0\|$; set $f = \|x_0\|g$. ■

The Hahn–Banach theorem is basic in the study of the concept of reflexivity, which we now discuss. Let L be a normed linear space, and L^* the

set of continuous linear functionals on L ; L^* is sometimes called the *conjugate space* of L . By 3.3.5, L^* is a Banach space, so that we may talk about L^{**} , the conjugate space of L^* , or the *second conjugate space* of L . We may identify L with a subspace of L^{**} as follows: If $x \in L$, we define $x^{**} \in L^{**}$ by

$$x^{**}(f) = f(x), \quad f \in L^*.$$

If $\|f_n - f\| \rightarrow 0$, then $f_n(x) \rightarrow f(x)$; hence x^{**} is in fact a continuous linear functional on L^* . Let us examine the map $x \rightarrow x^{**}$ of L into L^{**} .

3.4.6 Theorem. Define $h: L \rightarrow L^{**}$ by $h(x) = x^{**}$. Then h is an isometric isomorphism of L and the subspace $h(L)$ of L^{**} ; therefore, if $x \in L$, we have, by 3.3.1, $\|x\| = \|x^{**}\| = \sup\{|f(x)| : f \in L^*, \|f\| \leq 1\}$.

PROOF: To show that h is linear, we write

$$[h(ax + by)](f) = f(ax + by) = af(x) + bf(y) = [ah(x)](f) + [bh(y)](f).$$

We now prove that h is norm-preserving ($\|h(x)\| = \|x\|$ for all $x \in L$); consequently, h is one-to-one. If $x \in L$, $|[h(x)](f)| = |f(x)| \leq \|x\| \|f\|$, and hence $\|h(x)\| \leq \|x\|$. On the other hand, by 3.4.5(c), there is an $f \in L^*$ such that $\|f\| = 1$ and $|f(x)| = \|x\|$. Thus

$$\sup\{|[h(x)](f)| : f \in L^*, \|f\| = 1\} \geq \|x\|$$

so that $\|h(x)\| \geq \|x\|$, and consequently $\|h(x)\| = \|x\|$. ■

If $h(L) = L^{**}$, L is said to be *reflexive*. Note that L^{**} is complete by 3.3.5 so by 3.4.6, a reflexive normed linear space is necessarily complete. We shall now consider some examples.

3.4.7 Examples. (a) Every Hilbert space is reflexive. For if ψ is the conjugate isometry of 3.3.4(a), H^* becomes a Hilbert space if we take $\langle f, g \rangle = \langle \psi(g), \psi(f) \rangle$. Thus if $q \in H^{**}$, we have, for some $g \in H^*$, $q(f) = \langle f, g \rangle = \langle \psi(g), \psi(f) \rangle = f(x)$, where $x = \psi(g)$. Therefore $q = h(x)$.

(b) If $1 < p < \infty$, l^p is reflexive. For by 3.3.4(b), $(l^p)^*$ is isometrically isomorphic to l^q , where $(1/p) + (1/q) = 1$. Thus if $t \in (l^p)^{**}$ we have $t(y) = \sum_{k=1}^{\infty} y_k z_k$, $y \in l^q$, where z is an element of $(l^q)^* = l^p$. But then $t = h(z)$.

Essentially the same argument, with the aid of Problem 11 of Section 3.3, shows that if $(\Omega, \mathcal{F}, \mu)$ is an arbitrary measure space $L^p(\Omega, \mathcal{F}, \mu)$ is reflexive for $1 < p < \infty$.

(c) The space l^1 is not reflexive. This depends on the following result:

3.4.8 Theorem. If L is a normed linear space and L^* is separable, so is L . Thus if L is reflexive and separable (so that L^{**} is separable), then so is L^* .

PROOF. Let f_1, f_2, \dots form a countable dense subset of $\{f \in L^* : \|f\| = 1\}$ (note that any subset of a separable metric space is separable). Since $\|f_n\| = 1$, we can find points $x_n \in L$ with $\|x_n\| = 1$ and $|f_n(x_n)| \geq \frac{1}{2}$ for all n . Let M be the space spanned by the x_n ; we claim that $M = L$. If not, 3.4.5(a) yields an $f \in L^*$ with $f = 0$ on M and $\|f\| = 1$. But then $\frac{1}{2} \leq |f_n(x_n)| = |f_n(x_n) - f(x_n)| \leq \|f_n - f\|$ for all n , contradicting the assumption that $\{f_1, f_2, \dots\}$ is dense. ■

To return to 3.4.7(c), we note that l^1 is separable since $\{x \in l^1 : x_k = 0 \text{ for all but finitely many } k\}$ is dense. [If $x \in l^1$ and $x^{(n)} = (x_1, \dots, x_n, 0, 0, \dots)$ then $x^{(n)} \rightarrow x$ in l^1 .] But $(l^1)^* = l^\infty$ by 3.3.4(c), and this space is not separable. For if $S = \{x \in l^\infty : x_k = 0 \text{ or } 1 \text{ for all } k\}$, then S is uncountable and $\|x - y\| = 1$ for all $x, y \in S, x \neq y$. Thus the sets $B_x = \{y \in l^\infty : \|y - x\| < \frac{1}{2}\}, x \in S$, form an uncountable family of disjoint open sets. If there were a countable dense set D , there would be at least one point of D in each B_x , a contradiction. Thus l^1 is not reflexive.

We now consider the second basic result of this section. Suppose that the A_i, i belonging to the arbitrary index set I , are bounded linear operators from L to M , where L and M are normed linear spaces. The uniform boundedness principle asserts that if L is complete and the A_i are pointwise bounded, that is, $\sup\{\|A_i x\| : i \in I\} < \infty$ for each $x \in L$, then the A_i are uniformly bounded, that is, $\sup\{\|A_i\| : i \in I\} < \infty$. Completeness of L is essential; to see this, let L be the set of all sequences $x = (x_1, x_2, \dots)$ of complex numbers such that $x_k = 0$ for all but finitely many k , with the l^p norm, $1 \leq p \leq \infty$. Take $M = C$, and $A_n x = n x_n, n = 1, 2, \dots$. For any x , $A_n x = 0$ for sufficiently large n , so the A_n are pointwise bounded, although $\|A_n\| = n \rightarrow \infty$.

The proof that we shall give uses the Baire category theorem, one version of which states if a complete metric space is a countable union of closed sets, one of the sets must have a nonempty interior. (See the appendix on general topology, Theorem A9.2.)

3.4.9 Principle of Uniform Boundedness. Let $A_i, i \in I$, be bounded linear operators from the Banach space L to the normed linear space M . If the A_i are pointwise bounded, they are uniformly bounded.

PROOF. Let $C_n = \{x \in L : \sup_i \|A_i x\| \leq n\}, n = 1, 2, \dots$. Since the A_i are pointwise bounded, $\bigcup_{n=1}^{\infty} C_n = L$, and since the A_i are continuous, each C_n is

closed. By the Baire category theorem, for some n there is a closed ball $\bar{B} = \{x \in L: \|x - x_0\| \leq r\} \subset C_n$. Now if $\|y\| \leq 1$ and $i \in I$, we have

$$\begin{aligned}\|A_i y\| &= \frac{1}{r} \|A_i z\| \quad \text{where } z = ry \\ &\leq \frac{1}{r} \|A_i(x_0 + z)\| + \frac{1}{r} \|A_i x_0\| \\ &\leq \frac{2n}{r} \quad \text{since } x_0 + z \text{ and } x_0 \text{ belong to } \bar{B}.\end{aligned}$$

Thus $\|A_i\| \leq 2n/r$. ■

The uniform boundedness principle is used in an important way in the study of weak convergence, a concept that we now describe.

3.4.10 Definitions and Comments. The sequence (or net) $\{x_n\}$ in the normed linear space L is said to converge *weakly* to $x \in L$ iff $f(x_n) \rightarrow f(x)$ for every $f \in L^*$ (notation: $x_n \xrightarrow{w} x$). Convergence in the metric of L ($\|x_n - x\| \rightarrow 0$) will be called *strong convergence* and will be written simply as $x_n \rightarrow x$. [In this terminology, strong convergence of the sequence of linear operators A_n to the linear operator A means strong convergence of $A_n x$ to Ax for each x (see 3.3.6).]

For each $x_0 \in L$, consider the collection of sets

$$U(x_0) = \{x \in L: |f_i(x) - f_i(x_0)| < \varepsilon_i, \quad i = 1, \dots, n\},$$

where $\varepsilon_1, \dots, \varepsilon_n > 0, f_1, \dots, f_n \in L^*, n = 1, 2, \dots$

We form a topology by taking the $U(x_0)$ as a base for the neighborhood system at x_0 . (There is no difficulty checking that the required axioms for a neighborhood system are satisfied.) The resulting (Hausdorff) topology is called the *weak topology* on L ; it follows from the definition that convergence relative to the weak topology coincides with weak convergence. (The topology induced by the norm of L will be called the *strong topology*.)

It follows from the definitions that strong convergence implies weak convergence, hence every set that is open in the weak topology is open in the strong topology (and similarly for closed sets). To see that the converse does not hold, let $\{e_1, e_2, \dots\}$ be an infinite orthonormal sequence in a Hilbert space H . If $x \in H$, then $\langle e_n, x \rangle \rightarrow 0$ as $n \rightarrow \infty$ by Bessel's inequality, hence $e_n \xrightarrow{w} 0$. But $\|e_n\| \equiv 1$, so e_n does not converge strongly to 0.

In a Euclidean space, strong and weak convergence coincide. For if $x_n = a_{1n}e_1 + \dots + a_{kn}e_k$ converges weakly to $x = a_1e_1 + \dots + a_ke_k$ (where the e_i

are basis vectors for C^k), let f be a continuous linear functional that is 1 at e_i and 0 at $e_j, j \neq i$. Then $f(x_n) \rightarrow f(x)$, so that $a_{in} \rightarrow a_i$ as $n \rightarrow \infty$ ($i = 1, \dots, k$). Thus $x_n \rightarrow x$ strongly.

The weak topology is an important example of a product topology (see the appendix on general topology, Section A3). For L may be identified (see 3.4.6) with a subset $h(L)$ of the set C^{L^*} of all complex-valued functions on L^* , and the weak topology is the product topology, also called the “topology of pointwise convergence,” on C^{L^*} , relativized to $h(L)$.

We give a few properties of weak convergence.

3.4.11 Theorem. (a) A weakly convergent sequence $\{x_n\}$ is bounded, that is, $\sup_n \|x_n\| < \infty$. In fact if $x_n \xrightarrow{w} x_0$, then $\|x_0\| \leq \liminf_{n \rightarrow \infty} \|x_n\|$.

(b) If A is a bounded linear operator from L to M and the net $\{x_i\}$ converges weakly to x_0 in L , then Ax_i converges weakly to Ax_0 in M .

(c) If the net $\{x_i\}$ converges weakly to x_0 , then x_0 belongs to the subspace spanned by the x_i .

(d) If M is a linear manifold of L , the closures of M relative to the weak and strong topologies coincide. Consequently if M is closed in the strong topology, it is closed in the weak topology.

PROOF. (a) The x_n may be regarded as continuous linear functionals on L^* with $x_n(f) = f(x_n)$ (see 3.4.6). The x_n are pointwise bounded on L^* since $f(x_n) \rightarrow f(x_0)$; hence by the uniform boundedness principle, $\sup_n \|x_n\| < \infty$. Also, if $f \in L^*$,

$$|f(x_0)| = \lim_{n \rightarrow \infty} |f(x_n)| = \liminf_{n \rightarrow \infty} |f(x_n)| \leq \|f\| \liminf_{n \rightarrow \infty} \|x_n\|.$$

Since $\|x_0\| = \sup\{|f(x_0)| : f \in L^*, \|f\| \leq 1\}$, we may conclude that $\|x_0\| \leq \liminf_{n \rightarrow \infty} \|x_n\|$.

(b) If $g \in M^*$, define $f = g \circ A$; since A is continuous, we have $f \in L^*$, so that $f(x_i) \rightarrow f(x_0)$, that is, $g(Ax_i) \rightarrow g(Ax_0)$. But g is arbitrary; hence $Ax_i \xrightarrow{w} Ax_0$.

(c) If this is false, 3.4.5(a) yields an $f \in L^*$ with $f(x_0) = 1$ and $f(x_i) \equiv 0$, contradicting $x_i \xrightarrow{w} x_0$.

(d) The strong closure is always a subset of the weak closure, so assume x belongs to the weak closure of M . Let $\{x_i\}$ be a net in M with $x_i \xrightarrow{w} x$. By (c), x is a strong limit of a net of finite linear combinations of the x_i . But these finite linear combinations belong to M since M is a subspace; hence x belongs to the strong closure. ■

In order to characterize weak convergence in specific spaces, the following result is useful.

3.4.12 Theorem. Let E be a subset of L^* such that $S(E) = L^*$, and let $\{x_n\}$ be a sequence in L . If $x_0 \in L$, then $x_n \xrightarrow{w} x_0$ iff $\sup_n \|x_n\| < \infty$ and $f(x_n) \rightarrow f(x_0)$ for all $f \in E$.

PROOF. The “only if” part follows from 3.4.11(a) and the definition of weak convergence. For the “if” part, let $f \in L^*$, and choose elements $f_k \in L(E)$ with $\|f - f_k\| \rightarrow 0$. Then

$$\begin{aligned}|f(x_n) - f(x_0)| &\leq |f(x_n) - f_k(x_n)| + |f_k(x_n) - f_k(x_0)| + |f_k(x_0) - f(x_0)| \\&\leq \|f - f_k\| \|x_n\| + |f_k(x_n) - f_k(x_0)| + \|f_k - f\| \|x_0\|.\end{aligned}$$

Since the x_n are bounded in norm, given $\varepsilon > 0$, we may choose k such that the right-hand side is at most $|f_k(x_n) - f_k(x_0)| + \varepsilon$; but $f_k(x_n) \rightarrow f_k(x_0)$ as $n \rightarrow \infty$ since $f_k \in L(E)$, and since ε is arbitrary, the result follows. ■

We now describe weak convergence in l^p and \mathbf{L}^p .

3.4.13 Theorem. Assume $1 < p < \infty$.

- (a) Let $x_n = (x_{n1}, x_{n2}, \dots) \in l^p$, $n = 1, 2, \dots$. If $z \in l^p$, then $x_n \xrightarrow{w} z$ iff $\sup_n \|x_n\| < \infty$ and $x_{nk} \xrightarrow{w} z_k$ as $n \rightarrow \infty$ for each k .
- (b) Let $x_n \in \mathbf{L}^p(\Omega, \mathcal{F}, \mu)$, $n = 1, 2, \dots$, where μ is assumed finite. If $z \in \mathbf{L}^p(\Omega, \mathcal{F}, \mu)$, then $x_n \rightarrow z$ iff $\sup_n \|x_n\| < \infty$ and $\int_A x_n d\mu \rightarrow \int_A z d\mu$ for each $A \in \mathcal{F}$. (It will often be convenient to blur the distinction between l^p and \mathbf{L}^p , and treat the elements of \mathbf{L}^p as functions rather than equivalence classes.)

PROOF. (a) Define $f_k \in (l^p)^*$ by $f_k(x) = x_k$; then f_k corresponds to the sequence in l^q with a one in position k and zeros elsewhere [see 3.3.4(b)]. Take $E = \{f_1, f_2, \dots\}$ and apply 3.4.12.

(b) For each $A \in \mathcal{F}$, define $f_A \in (\mathbf{L}^p)^*$ by $f_A(x) = \int_A x d\mu$; f_A corresponds to the indicator function $I_A \in \mathbf{L}^q$ (see Problem 11, Section 3.3). Take E as the set of all f_A , $A \in \mathcal{F}$; E spans $(\mathbf{L}^p)^*$ because the simple functions are dense in \mathbf{L}^q , and the result follows from 3.4.12. ■

We now consider the third basic result, the open mapping theorem. This will allow us to conclude that under certain conditions, the inverse of a one-to-one continuous linear operator is continuous. We cannot make this assertion in general, as the following example shows: Let L be the set of all continuous complex-valued functions x on $[0, 1]$ such that $x(0) = 0$, $M = \{x \in L: x \text{ has a continuous derivative on } [0, 1]\}$; put the sup norm on L and M . If A is defined by $(Ax)(t) = \int_0^t x(s) ds$, $0 \leq t \leq 1$, then A is a one-to-one, bounded, linear

operator from L onto M . But A^{-1} is discontinuous; for example, if $x_n(t) = \sin nt$, then $y_n(t) = (Ax_n)(t) = (1 - \cos nt)/n$, so that $y_n \rightarrow 0$ in M , but x_n has no limit in L . A hypothesis under which continuity of the inverse holds is the completeness of both spaces L and M . In the above example, M is not complete; for example, a sequence of polynomials may converge uniformly to a continuous function without a continuous derivative (in fact to a continuous nowhere differentiable function).

We now state the third basic theorem.

3.4.14 Open Mapping Theorem. Let A be a bounded linear operator from the Banach space L onto the Banach space M . Then A is an open map, that is, if D is an open subset of L , then $A(D)$ is an open subset of M . Consequently if A is also one-to-one, then A^{-1} is bounded.

PROOF. We defer this until the next section (see 3.5.11) since it will not be much more difficult to prove a more general statement about mappings of topological vector spaces. ■

The open mapping theorem allows us to prove the closed graph theorem, which is often useful in proving that a particular linear operator is bounded.

First we observe that if L and M are normed linear spaces, we may define a norm on the product space $L \times M$ by $\|(x, y)\| = (\|x\|^p + \|y\|^p)^{1/p}$, $x \in L$, $y \in M$, where p is any fixed real number in $[1, \infty)$. For if $x, x' \in L$, $y, y' \in M$,

$$\begin{aligned} (\|x + x'\|^p + \|y + y'\|^p)^{1/p} &\leq [(\|x\| + \|x'\|)^p + (\|y\| + \|y'\|)^p]^{1/p} \\ &\leq (\|x\|^p + \|y\|^p)^{1/p} + (\|x'\|^p + \|y'\|^p)^{1/p} \end{aligned}$$

by Minkowski's inequality applied to R^2 .

Therefore the triangle inequality is satisfied and we have defined a norm on $L \times M$ (the other requirements for a norm are immediate). Furthermore, $(x_n, y_n) \rightarrow (x, y)$ iff $x_n \rightarrow x$ and $y_n \rightarrow y$; thus regardless of the value of p , the norm on $L \times M$ induces the product topology; also, if L and M are complete, so is $L \times M$. [The same result is obtained using the analog of the L^∞ norm, that is, $\|(x, y)\| = \max(\|x\|, \|y\|)$.]

3.4.15 Definition. Let A be a linear operator from L to M , where L and M are normed linear spaces. We say that A is *closed* iff the graph $G(A) = \{(x, Ax) : x \in L\}$ is a closed subset of $L \times M$. Equivalently, A is closed iff the following condition holds:

If $x_n \in L$, $x_n \rightarrow x$, and $Ax_n \rightarrow y$, then $(x, y) \in G(A)$; in other words, $y = Ax$. This formulation shows that every bounded linear operator is closed. The converse holds if L and M are Banach spaces.

3.4.16 Closed Graph Theorem. If A is a closed linear operator from the Banach space L to the Banach space M , then A is bounded.

PROOF. Since $G(A)$ is a closed subspace of $L \times M$, it is a Banach space. Define $P: G(A) \rightarrow L$ by $P(x, Ax) = x$. Then P is linear and maps onto L , and $\|P(x, Ax)\| = \|x\| \leq \|(x, Ax)\|$; hence $\|P\| \leq 1$ so that P is bounded. [Alternatively, if $(x_n, Ax_n) \rightarrow (x, y)$, then $x_n \rightarrow x$, proving continuity of P .] Similarly, the linear operator $Q: G(A) \rightarrow M$ given by $Q(x, Ax) = Ax$ is bounded. If $P(x, Ax) = 0$, then $x = Ax = 0$, so P is one-to-one. By 3.4.14, P^{-1} is bounded, and since $A = Q \circ P^{-1}$, A is bounded. ■

As an application of the closed graph theorem, we show that if P is the projection of a Banach space L on a closed subspace M , then P is continuous [see 3.3.3(c)]. Let $\{x_n\}$ be a sequence of points in L with $x_n \rightarrow x$, and assume Px_n converges to the element $y \in M$. Recall that in defining a projection operator it is assumed that L is the direct sum of closed subspaces M and N ; thus

$$x_n = y_n + z_n \quad \text{where} \quad y_n = Px_n \in M, \quad z_n \in N.$$

Since $x_n \rightarrow x$ and $y_n \rightarrow y$, it follows that $z_n \rightarrow z = x - y$, necessarily in N . Therefore $x = y + z$, $y \in M$, $z \in N$, so that $y = Px$, proving P closed. By 3.4.16, P is continuous.

Problems

1. Show that a subadditive, absolutely homogeneous functional on a vector space must be nonnegative, and hence a seminorm. Give an example of a subadditive, positive-homogeneous functional that fails to be nonnegative.
2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and assume \mathcal{F} is countably generated, that is, there is a countable set $\mathcal{C} \subset \mathcal{F}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$. (Note that the minimal field \mathcal{F}_0 over \mathcal{C} is also countable; see Problem 9, Section 1.2.) If μ is σ -finite on \mathcal{F}_0 and $1 \leq p < \infty$, show that $\mathbf{L}^p(\Omega, \mathcal{F}, \mu)$ is separable. If in addition there is an infinite collection of disjoint sets $A \in \mathcal{F}$ with $\mu(A) > 0$, show that $\mathbf{L}^1(\Omega, \mathcal{F}, \mu)$ is not reflexive.

3. If L and M are normed linear spaces and $[L, M]$ is complete, show that M must be complete.
4. Let $A \in [L, M]$, where L and M are normed linear spaces. The *adjoint* of A is an operator $A^* : M^* \rightarrow L^*$, defined as follows: If $f \in M^*$ we take $(A^*f)(x) = f(Ax)$, $x \in L$. Establish the following results:
 - (a) $\|A^*\| = \|A\|$.
 - (b) $(aA + bB)^* = aA^* + bB^*$ for all $a, b \in C$, $A, B \in [L, M]$.
 - (c) If $A \in [L, M]$, $B \in [M, N]$, then $(BA)^* = A^*B^*$, where BA is the composition of A and B .
 - (d) If $A \in [L, M]$, A maps onto M , and A^{-1} exists and belongs to $[M, L]$, then $(A^{-1})^* = (A^*)^{-1}$.
5. Define the *annihilator* of the subset K of the normed linear space L as $K^\perp = \{f \in L^* : f(x) = 0 \text{ for all } x \in K\}$. Similarly, if $J \subset L^*$, define $J^\perp = \{x \in L : f(x) = 0 \text{ for all } f \in J\}$. If $A \in [L, M]$, we denote by $N(A)$ the null space $\{x \in L : Ax = 0\}$, and by $R(A)$ the closure of the range of A , that is, the closure of $\{Ax : x \in L\}$. Establish the following:
 - (a) For any $K \subset L$, $K^{\perp\perp} = S(K)$, the space spanned by K .
 - (b) $R(A)^\perp = N(A^*)$ and $R(A) = N(A^*)^\perp$.
 - (c) $R(A) = M$ iff A^* is one-to-one.
 - (d) $R(A^*)^\perp = N(A)$.
 - (e) For any $J \subset L^*$, $S(J) \subset J^{\perp\perp}$; $S(J) = J^{\perp\perp}$ if L is reflexive.
 - (f) $R(A^*) \subset N(A)^\perp$; $R(A^*) = N(A)^\perp$ if L is reflexive.
 - (g) If $R(A^*) = L^*$, then A is one-to-one; the converse holds if L is reflexive.
6. Consider the Hahn–Banach theorem 3.4.2, with the additional assumption that L is a normed linear space (or more generally, a topological vector space) and p is continuous at 0; hence continuous on all of L since $|p(x) - p(y)| \leq p(x - y)$. Show that if L is separable, the theorem may be proved without Zorn's lemma. It follows that 3.4.3 and 3.4.4 do not require Zorn's lemma under the above hypothesis.
7. If μ_0 is a finitely additive, nonnegative real-valued set function on a field \mathcal{F}_0 of subsets of a set Ω , use the Hahn–Banach theorem to show that μ_0 has an extension to a finitely additive, nonnegative real-valued set function on the class of *all* subsets of Ω . Thus in one respect, at least, finite additivity is superior to countable additivity.
8. If the sequence (or net) of bounded linear operators A_n on a Banach space converges strongly to the (necessarily linear) operator A , show that A is bounded; in fact

$$\|A\| \leq \liminf_{n \rightarrow \infty} \|A_n\| \leq \sup_n \|A_n\| < \infty.$$

9. Let $\{A_i, i \in I\}$ be a family of continuous linear operators from the Banach space L to the normed linear space M . Assume the A_i are weakly bounded, that is, $\sup_i |f(A_i x)| < \infty$ for all $x \in L$ and all $f \in M^*$. Show that the A_i are uniformly bounded, that is, $\sup_i \|A_i\| < \infty$.
10. (a) If the elements $x_i, i \in I$, belong to the normed linear space L , and $\sup_i |f(x_i)| < \infty$ for each $f \in L^*$, show that $\sup_i \|x_i\| < \infty$.
(b) If A is a linear operator from the normed linear space L to the normed linear space M , and $f \circ A$ is continuous for each $f \in M^*$, show that A is continuous.
11. Let L , M , and N be normed linear spaces, with L or M complete, and let $B: L \times M \rightarrow N$ be a bilinear form, that is, $B(x, y)$ is linear in x for each fixed y , and linear in y for each fixed x . If for each $f \in N^*$, $f(B(x, y))$ is continuous in x for each fixed y , and continuous in y for each fixed x , show that B is bounded, that is,

$$\sup\{|B(x, y)| : \|x\|, \|y\| \leq 1\} < \infty.$$

- Equivalently, for some positive constant k we have $|B(x, y)| \leq k\|x\|\|y\|$ for all x, y .
12. Give an example of a closed unbounded operator from one normed linear space to another.
 13. Let A be a bounded linear operator from the Banach space L onto the Banach space M . Show that there is a positive number k such that for each $y \in M$ there is an $x \in L$ with $y = Ax$ and $\|x\| \leq k\|y\|$. This result is sometimes called the *solvability theorem*.

3.5 Some Properties of Topological Vector Spaces

Recall that a topological vector space is a vector space L with a topology that makes addition and scalar multiplication continuous; that is, the maps $(x, y) \rightarrow x + y$ of $L \times L \rightarrow L$ and $(a, x) \rightarrow ax$ of $C \times L \rightarrow L$ are continuous, with the product topology on $L \times L$ and $C \times L$. In this section we look at some of the important properties of such spaces.

One cannot put an arbitrary topology on a vector space L and obtain a topological vector space. For example, consider the discrete topology (all subsets of L are open); with this topology, L is not a topological vector space (exclude the trivial case $L = \{0\}$). To see this, pick a nonzero $x \in L$, and let $a_n = 1/n$, $n = 1, 2, \dots$. If scalar multiplication were continuous, we would have $a_n x \rightarrow 0 \cdot x = 0$; since $\{0\}$ is open, this implies that $a_n x = 0$ for sufficiently large n , a contradiction.

The following result is helpful in checking whether a given topology makes L into a topological vector space. [Throughout this section, we use the following notation: If A and B are subsets of L ,

$$\begin{aligned} A + B &= \{x + y: x \in A, y \in B\}, \\ A - B &= \{x - y: x \in A, y \in B\}. \end{aligned}$$

If $x \in L$,

$$\begin{aligned} x + B &= \{x + y: y \in B\} \\ x - B &= \{x - y: y \in B\}. \end{aligned}$$

3.5.1 Theorem. Let L be a topological vector space, and let \mathcal{U} be a base of overneighborhoods at 0 (the sets in \mathcal{U} are overneighborhoods of 0, and for each neighborhood V of 0 there is a set $U \in \mathcal{U}$ with $U \subset V$). Then

- (a) if $U, V \in \mathcal{U}$, there is a set $W \in \mathcal{U}$ such that $W \subset U \cap V$;
- (b) if $U \in \mathcal{U}$ then for each $x \in L$ there is a $\delta > 0$ such that $ax \in U$ for all complex numbers a with $|a| \leq \delta$;
- (c) for each $U \in \mathcal{U}$, there is a set $V \in \mathcal{U}$ with $V + V \subset U$.

Conversely, if \mathcal{U} is a collection of subsets of L satisfying (a)–(c), and in addition,

- (d) each $U \in \mathcal{U}$ is *circled*, in other words, if $x \in U$, then $ax \in U$ whenever $|a| \leq 1$,

there is a unique topology that makes L a topological vector space with \mathcal{U} a base of overneighborhoods at 0.

PROOF. (a) This follows since the intersection of two neighborhoods of 0 is a neighborhood of 0.

(b) By continuity of scalar multiplication, $ax \rightarrow 0$ as $a \rightarrow 0$, with x fixed; hence $ax \in U$ for sufficiently small $|a|$.

(c) By continuity of the map $(x, y) \rightarrow x + y$ at $x = y = 0$, there are sets $V_1, V_2 \in \mathcal{U}$ such that if $x \in V_1$ and $y \in V_2$, then $x + y \in U$; in other words, $V_1 + V_2 \subset U$. By (a), there is a set $V \in \mathcal{U}$ with $V \subset V_1 \cap V_2$; then $V + V \subset V_1 + V_2 \subset U$.

Conversely, let \mathcal{U} satisfy (a)–(d). Take \mathcal{U} as an overneighborhood base at 0, and, for each $x \in L$, $x + \mathcal{U} = \{x + U: U \in \mathcal{U}\}$ as an overneighborhood base at x . Thus $\mathcal{V}(x) = \{V \subset L: V \supset x + U \text{ for some } U \in \mathcal{U}\}$ is the proposed class of overneighborhoods of x . Now:

- (i) If $V \in \mathcal{V}(x)$, then $x \in V$. This follows because 0 belongs to each $U \in \mathcal{U}$, by (b).

- (ii) If $V_1, V_2 \in \mathcal{V}(x)$, then $V_1 \cap V_2 \in \mathcal{V}(x)$. For if $x + U_i \subset V_i$, $i = 1, 2$, by (a) there is a set $W \in \mathcal{U}$ with $W \subset U_1 \cap U_2$. Then $x + W \subset (x + U_1) \cap (x + U_2) \subset V_1 \cap V_2$.
- (iii) If $V \in \mathcal{V}(x)$ and $V \subset W$, then $W \in \mathcal{V}(x)$. This is immediate from the definition of $\mathcal{V}(x)$.
- (iv) If $V \in \mathcal{V}(x)$, there is a set $W \in \mathcal{V}(x)$ such that $W \subset V$ and $V \in \mathcal{V}(y)$ for each $y \in W$. For if $x + U \subset V$, by (c) there is a set $T \in \mathcal{U}$ with $T + T \subset U$. If $W = x + T$ and $y = x + t \in W$, then $y + T = x + t + T \subset x + T + T \subset x + U \subset V$, hence $V \in \mathcal{V}(y)$.

Thus the axioms of a system of overneighborhoods are satisfied, so there is a topology on L with \mathcal{U} an overneighborhood base at 0. By definition of the $\mathcal{V}(x)$, we have $x_n \rightarrow x$ iff $x_n - x \rightarrow 0$; hence if the addition operation $(x, y) \rightarrow x + y$ is continuous at $(0, 0)$, it is continuous everywhere. But continuity of addition at $(0, 0)$ is immediate from (c).

Finally, we look at continuity of the multiplication operation $(a, x) \rightarrow ax$. By (b), the map is continuous at $a = 0$ for each fixed x , hence, as above, continuous at any a for each fixed x . If $U \in \mathcal{U}$ and $a \in C$, choose a positive integer $n \geq |a|$ and apply (c) inductively to obtain $V \in \mathcal{U}$ such that $V + \cdots + V$ (n times) is a subset of U . (Note, for example, that $V + V + V \subset V + V + V + V$ since $0 \in V$.) If $nV = \{ny: y \in V\}$, then $nV \subset V + \cdots + V \subset U$; hence by (d), $aV \subset U$, so that multiplication is continuous at $x = 0$ for each fixed a , and thus continuous at any x for each fixed a . Now if $U \in \mathcal{U}$, then by (d), if $|a| \leq 1$ and $x \in U$, then $ax \in U$, so that multiplication is jointly continuous at $(0, 0)$. Thus if $x_n \rightarrow x$ and $a_n \rightarrow a$, then $(a_n - a)(x_n - x) \rightarrow 0$, and by the above analysis, $a_n x \rightarrow ax$, $ax_n \rightarrow ax$. It follows that $a_n x_n \rightarrow ax$.

Finally, to prove uniqueness of the topology, observe that if \mathcal{U} is a base of overneighborhoods at 0, continuity of addition and subtraction makes the map $y \rightarrow x + y$, $y \in L$, a homeomorphism, and therefore $\mathcal{V}(x)$ is a system of overneighborhoods of x . ■

We have seen that a seminorm on a vector space L induces a topology under which L becomes a topological vector space. We now show that a similar result holds if instead of a single seminorm we use an entire family.

3.5.2 Theorem. Let $\{p_i, i \in I\}$ be a family of seminorms on the vector space L . Let \mathcal{U} be the class of all finite intersections of sets $\{x \in L: p_i(x) < \delta_i\}$ where $i \in I$, $\delta_i > 0$. Then \mathcal{U} is a neighborhood base for a topology that makes L a topological vector space. This topology is the weakest, making all the p_i continuous, and for a net $\{x_n\}$ in L , $x_n \rightarrow x$ iff $p_i(x_n - x) \rightarrow 0$ for each $i \in I$.

PROOF. We show that the conditions of 3.5.1 are satisfied. Condition (a) follows because an intersection of two finite intersections of sets $\{x: p_i(x) < \delta_i\}$ is a finite intersection of such sets. If $x \in L$, then $p_i(ax) = |a|p_i(x)$, which is less than δ_i if $|a|$ is sufficiently small; this proves (b). To prove (c), suppose that $U = \bigcap_{i=1}^n \{x: p_i(x) < \delta_i\}$, and let $V = \bigcap_{i=1}^n \{x: p_i(x) < \delta\}$ where $0 < \delta \leq \frac{1}{2} \min_i \delta_i$. If $y, z \in V$, then $p_i(y+z) \leq p_i(y) + p_i(z) < \delta_i$, hence $y+z \in U$. Finally, each $U \in \mathcal{U}$ is circled since $p_i(ax) = |a|p_i(x)$, proving (d). If $x \in U = \bigcap_{i=1}^n \{z: p_i(z) < \delta_i\}$ and $y \in V = \bigcap_{i=1}^n \{z: p_i(z) < \min_j [\delta_j - p_j(x)]\}$, then $p_i(x+y) \leq p_i(x) + p_i(y) < p_i(x) + \delta_i - p_i(x) = \delta_i$; hence $x+V \subset U$. Thus U is an overneighborhood of each of its points, so the sets in \mathcal{U} are open and \mathcal{U} is a neighborhood base, not merely an overneighborhood base.

Now continuity of p_i at 0 is equivalent to continuity everywhere [$|p_i(x) - p_i(y)| \leq p_i(x-y)$] and the sets $U \in \mathcal{U}$ are open in any topology for which the p_i are continuous at 0. This proves that the given topology is the weakest, making the p_i continuous. Finally, $x_n \rightarrow x$ iff $x_n - x \rightarrow 0$ (see the proof of 3.5.1), and $x_n - x \rightarrow 0$ iff for each i , $p_i(x_n - x) \rightarrow 0$, by definition of \mathcal{U} . ■

The topology induced by a family of seminorms is *locally convex*; that is, for each $x \in L$, the family of convex overneighborhoods forms a base of overneighborhoods at x . This follows because the sets $\{x: p_i(x) < \delta_i\}$ are convex. We shall prove later (see 3.5.7) that every locally convex topology is generated by a family of seminorms.

If for each $x \neq 0$ there is an $i \in I$ such that $p_i(x) \neq 0$, the topology is Hausdorff. For if $x \neq y$ and $p_i(x-y) = \delta > 0$, then $\{z: p_i(z-x) < \delta/2\}$ and $\{z: p_i(z-y) < \delta/2\}$ are disjoint neighborhoods of x and y .

3.5.3 Examples. (a) Let L be a vector space of complex-valued functions on a topological space Ω . For each compact subset K of Ω , define $p_K(x) = \sup\{|x(t)| : t \in K\}$. In the (Hausdorff) topology induced by the seminorms p_K , convergence means uniform convergence on all compact subsets of Ω . If we restrict the K to finite subsets of Ω , we obtain the topology of pointwise convergence. In general, if the K are restricted to a class \mathcal{C} of subsets of Ω , we obtain the topology of uniform convergence on sets in \mathcal{C} .

(b) Let $L = C^\infty[a, b]$, the collection of all infinitely differentiable complex-valued functions on the closed bounded interval $[a, b] \subset R$. For each n , let $p_n(x) = \sup\{|x^{(n)}(t)| : a \leq t \leq b\}$ where $x^{(n)}$ is the n th derivative of x . In the topology induced by the p_n , convergence means uniform convergence of all derivatives.

We now examine convex sets in more detail.

3.5.4 Definitions. Let K be a subset of the vector space L . Then K is said to be *radial at x* iff K contains a line segment through x in each direction, in other words, if $y \in L$, there is a $\delta > 0$ such that $x + \lambda y \in K$ for $0 \leq \lambda < \delta$. (If K is radial at x , x is sometimes called an *internal point* of K .) If K is convex and radial at 0, the *Minkowski functional* of K is defined as

$$p(x) = \inf \left\{ r > 0 : \frac{x}{r} \in K \right\}.$$

Intuitively, $p(x)$ is the factor by which x has to be shrunk in order to reach the boundary of K . The Minkowski functional has the following properties.

3.5.5 Lemma. Let K be a convex subset of L , radial at 0, and let p be the Minkowski functional of K .

- (a) The functional p is sublinear, that is, subadditive and positive-homogeneous.
- (b) $\{x \in L : p(x) < 1\}$ is the radial kernel of K , defined by $\text{rad ker } K = \{x \in K : K \text{ is radial at } x\}$; also, $K \subset \{x \in L : p(x) \leq 1\}$.
- (c) If K is circled, then p is a seminorm.
- (d) If L is a topological vector space and 0 belongs to the interior K^0 of K , then p is continuous, $\bar{K} = \{x \in L : p(x) \leq 1\}$, and $K^0 = \{x \in L : p(x) < 1\}$; hence $\{x \in L : p(x) = 1\}$ is the boundary of K .

PROOF. (a) If $x/r \in K$ and $y/s \in K$, then

$$\frac{x+y}{r+s} = \frac{r}{r+s} \frac{x}{r} + \frac{s}{r+s} \frac{y}{s} \in K \quad \text{by convexity.}$$

Thus $p(x+y) \leq r+s$; take the inf over r , then over s , to obtain $p(x+y) \leq p(x) + p(y)$, proving subadditivity. Positive homogeneity follows from the definition of p .

(b) If $p(x) < 1$, then $sx \in K$ for some $s > 1$; hence $x \in K$ by convexity. [Write $x = (1/s)(sx) + [1 - (1/s)]0$.] Now if $p(x) < 1$ and $y \in L$, then $p(x + \lambda y) \leq p(x) + \lambda p(y) < 1$ for λ sufficiently small and nonnegative; hence K is radial at x . Conversely, if K is radial at x , then $x + \lambda x \in K$ for some $\lambda > 0$; hence $p(x + \lambda x) \leq 1$ by definition of p . Thus $p(x) \leq (1 + \lambda)^{-1} < 1$. The last statement follows from the definition of p .

(c) If $x/r \in K$ and $a \in C$, $a \neq 0$, then $ax/|a|r \in K$ since K is circled; consequently $p(ax) \leq |a|r$. Take the inf over r to obtain $p(ax) \leq |a|p(x)$. Replace x by x/a to obtain $p(x) \leq |a|p(x/a)$ or, with $b = 1/a$, $p(bx) \geq |b|p(x)$. Now $p(0) = 0$ since $0 \in K$, and the result follows.

(d) Since $0 \in K^0$ there is a neighborhood U of 0 such that $U \subset K$. If $\lambda > 0$ and $y \in \lambda U$, that is, $y = \lambda x$ for some $x \in U$, then $p(y) = \lambda p(x) \leq \lambda$.

[Note $x \in K$ implies $p(x) \leq 1$, by (b).] Thus p is continuous at the origin, and therefore continuous everywhere by subadditivity.

Since p is continuous, $\{x: p(x) \leq 1\}$ is closed, so by (b), $\bar{K} \subset \{x: p(x) \leq 1\}$. But if $0 < \lambda < 1$ and $p(x) \leq 1$, then $p(\lambda x) < 1$; hence $\lambda x \in K$. If $\lambda \rightarrow 1$, then $\lambda x \rightarrow x$; therefore $p(x) \leq 1$ implies $x \in \bar{K}$.

Again, by continuity of p , $\{x: p(x) < 1\}$ is open, and hence is a subset of K^0 . But if $p(x) \geq 1$, then by considering λx with $\lambda > 1$ we see that x is a limit of a sequence of points not in K , hence $x \notin K^0$. ■

Before characterizing locally convex spaces, we need the following result.

3.5.6 Lemma. If U is a neighborhood of 0 in the topological vector space L , there is a circled neighborhood V of 0 with $V \subset U$, and a closed circled overneighborhood W of 0 with $W \subset U$. If L is locally convex, V and W can be taken as convex.

PROOF. Choose $T \in \mathcal{U}$ and $\delta > 0$ such that $aT \subset U$ for $|a| \leq \delta$, and take $V = \bigcup\{aT: |a| \leq \delta\}$. Now if $A \subset L$, we claim that

$$\bar{A} = \bigcap\{A + N: N \in \mathcal{N}\}, \quad (1)$$

where \mathcal{N} is the family of all neighborhoods of 0. [This may be written as $\bar{A} = \bigcap\{A - N: N \in \mathcal{N}\}$; note that $N \in \mathcal{N}$ iff $-N \in \mathcal{N}$ since the map $y \mapsto -y$ is a homeomorphism.] For if $x \in \bar{A}$ and $N \in \mathcal{N}$, then $x + N$ is a neighborhood of x ; hence $(x + N) \cap A \neq \emptyset$. If $y \in (x + N) \cap A$, then $x \in y - N$. If $x \notin \bar{A}$, then $(x + N) \cap A = \emptyset$ for some $N \in \mathcal{N}$; hence $x \notin A - N$.

Now if U is a neighborhood of 0, let V_1 be a circled neighborhood of 0 with $V_1 + V_1 \subset U$. By (1), $\bar{V}_1 \subset V_1 + V_1$, and since V_1 is circled, so is \bar{V}_1 . Thus we may take $W = \bar{V}_1$.

In the locally convex case, we may as well assume U convex [the interior of a convex set is convex, by 3.5.5(d) and the fact that a translation of a convex set is convex]. If V_2 is a circled neighborhood included in U , the convex hull

$$\hat{V}_2 = \left\{ \sum_{i=1}^n \lambda_i x_i: x_i \in V_2, \quad \lambda_i \geq 0, \quad \sum_{i=1}^n \lambda_i = 1, \quad n = 1, 2, \dots \right\},$$

the smallest convex overset of V_2 , is also included in U . Since V_2 is circled so is \hat{V}_2 , and therefore so is $(\hat{V}_2)^0$. (If $x + N \subset \hat{V}_2$, $N \in \mathcal{N}$, and $|a| \leq 1$, then $ax + aN \subset \hat{V}_2$.) Thus we may take $V = (\hat{V}_2)^0$.

Finally, if V_3 is a circled convex neighborhood of 0 with $V_3 + V_3 \subset U$, then $W = \bar{V}_3$ is closed, circled, and convex, and by (1), $W \subset V_3 + V_3 \subset U$. ■

3.5.7 Theorem. If L is a locally convex topological vector space, the topology of L is generated, in the sense of 3.5.2, by a family of seminorms. Specifically, if \mathcal{U} is the collection of all circled convex neighborhoods of 0, the Minkowski functionals p_U of the sets $U \in \mathcal{U}$ are the desired seminorms.

PROOF. By 3.5.5(c), the p_U are seminorms, and 3.5.6, \mathcal{U} is a base at 0 for the topology of L . By 3.5.5(d), for each $U \in \mathcal{U}$ we have $U = \{x: p_U(x) < 1\}$, and it follows that the topology of L is the same as the topology induced by the p_U . ■

The fact that the Minkowski functional is sublinear suggests that the Hahn–Banach theorem may provide useful information. The next result illustrates this idea.

3.5.8 Theorem. Let K be a convex subset of the real vector space L ; assume K is radial at 0 and has Minkowski functional p .

- (a) If f is a linear functional on L , then $f \leq 1$ on K iff $f \leq p$ on L .
- (b) If g is a linear functional on the subspace M , and $g \leq 1$ on $K \cap M$, then g may be extended to a linear functional f on L such that $f \leq 1$ on K .
- (c) If in addition K is circled and L is a complex vector space, and g is a linear functional on the subspace M with $|g| \leq 1$ on $K \cap M$, then g may be extended to a linear functional f on L such that $|f| \leq 1$ on K .
- (d) A continuous linear functional g on a subspace M of a locally convex topological vector space L may be extended to a continuous linear functional on L .

PROOF. (a) If $f \leq p$ on L , then $f \leq 1$ on K by 3.5.5(b). Conversely, assume $f \leq 1$ on K . If $x/r \in K$, then $f(x/r) \leq 1$, so $f(x) \leq r$. Take the inf over r to obtain $f(x) \leq p(x)$.

(b) By (a), $g \leq p$ on M , so by the Hahn–Banach theorem, g extends to a linear functional f with $f \leq p$ on L . By (a), $f \leq 1$ on K .

(c) Let L' be L regarded as a real vector space; by (b), $g_1 = \operatorname{Re} g$ may be extended to a linear function f_1 on L' with $f_1 \leq 1$ on K . If $f(x) = f_1(x) - i f_1(ix)$, then (as in 3.4.3) f is a linear functional on L , and on M we have $f(x) = g_1(x) - i g_1(ix) = g(x)$. Finally, if $f(x) = r e^{i\theta}$, $r \geq 0$, then $|f(x)| = r = f(e^{-i\theta} x) = f_1(e^{-i\theta} x)$ since r is real. If $x \in K$, then $e^{-i\theta} x \in K$ since K is circled, so $|f(x)| \leq 1$.

(d) By continuity, g must be bounded on some neighborhood of 0 in M , so that by 3.5.6 there is a convex circled neighborhood K of 0 in L such that $|g|$ is bounded on $K \cap M$. By (c), g has an extension to a linear functional f on L with $|f|$ bounded on K . But if $|f| \leq r$ on K , then, given $\delta > 0$, we have $|f| \leq \delta$ on the neighborhood $(\delta/r)K$, proving continuity. ■

In 3.4.14 we promised a general version of the open mapping theorem, and we now consider this question. The following technical lemma is the key step.

3.5.9 Lemma. Let A be a continuous linear operator from L to M , where L is a complete, metrizable topological vector space and M is a Hausdorff topological vector space. Assume that for every neighborhood W of 0 in L , $\overline{A(W)}$ is an overneighborhood of 0 in M . Then in fact $A(W)$ is an overneighborhood of 0 in M .

PROOF. Denote by \mathcal{U} the collection of overneighborhoods of 0 in L , and by \mathcal{V} the corresponding collection in M . It suffices to show that if U and V are neighborhoods of 0 in L , then $\overline{A(U)} \subset A(\overline{U + V})$. For then given the neighborhood W we may find a closed circled neighborhood U with $U + U + U \subset W$; then [see 3.5.6, Eq. (1)] $\overline{U + U} \subset U + U + U \subset W$. Now $\overline{A(U)}$, which belongs to \mathcal{V} by hypothesis, is a subset of $A(\overline{U + U})$ by the result to be established, and this in turn is a subset of $A(W)$, as desired.

Thus let U and V be neighborhoods of 0 in L . Since L is metrizable it has a countable base of neighborhoods at 0, say U_1, U_2, \dots . It may be assumed that $U_{n+1} + U_{n+1} \subset U_n$ for all n [3.5.1(c)], and that $U_1 = U$, $U_2 + U_2 \subset U \cap V$.

Let $y \in \overline{A(U)}$; we try to find elements $x_n \in U_n$ such that if $y_n = Ax_n$, then, for all $n = 1, 2, \dots$,

$$y_1 + \cdots + y_n - y \in \overline{A(U_{n+1})}. \quad (1)$$

Since $y \in \overline{A(U_1)}$ ($= \overline{A(U)}$) and $y + \overline{A(U_2)}$ is an overneighborhood of y , there is an element $y_1 \in A(U_1) \cap (y + \overline{A(U_2)})$. Thus $y_1 = Ax_1$ for some $x_1 \in U_1$, and $y_1 - y \in \overline{A(U_2)}$. If x_1, \dots, x_n have been found such that $y_1 + \cdots + y_n - y \in \overline{A(U_{n+1})}$, then since $y - (y_1 + \cdots + y_n) + \overline{A(U_{n+2})}$ is an overneighborhood of $y - (y_1 + \cdots + y_n)$, there is an element $y_{n+1} \in A(U_{n+1}) \cap (y - (y_1 + \cdots + y_n) + \overline{A(U_{n+2})})$, so that we have $x_{n+1} \in U_{n+1}$, $y_{n+1} = Ax_{n+1}$, and $y_1 + \cdots + y_{n+1} - y \in \overline{A(U_{n+2})}$, completing the induction.

Let $z_n = \sum_{i=1}^n x_i$; then $z_{n+k} - z_n = x_{n+1} + \cdots + x_{n+k} \in U_{n+1} + \cdots + U_{n+k} \subset U_n$. (Note $U_{n+k-1} + U_{n+k} \subset U_{n+k-1} + U_{n+k-1} \subset U_{n+k-2}$, and proceed backward.) Since the U_n form a base at 0, $\{z_n\}$ is a Cauchy sequence in L ; hence z_n converges to some $x \in L$ by completeness. [By the appendix on general topology, Theorem A10.9, the metric d of L may be assumed invariant, that is, $d(x, y) = d(x + z, y + z)$ for all z . We know that $z_m - z_n \rightarrow 0$ as $n, m \rightarrow \infty$;

hence by invariance, $d(z_m, z_n) = d(z_m - z_n, 0) \rightarrow 0$, so that $\{z_n\}$ is Cauchy.] Since

$$x_1 + \cdots + x_n \in U_1 + \cdots + U_n \subset U_1 + (U_2 + U_2) \subset U + (U \cap V) \subset U + V$$

it follows that $x \in \overline{U + V}$. If we can show that $y = Ax$, then $y \in A(\overline{U + V})$ and we have finished.

If $y \neq Ax$, the Hausdorff hypothesis for M yields a closed circled $W' \in \mathcal{V}$ such that $Ax + W'$ and $y + W'$ are disjoint; let W be a closed set in \mathcal{V} such that $W + W \subset W'$. By continuity of A there is a set $W_1 \in \mathcal{U}$ such that $A(W_1) \subset W$. Since the U_n form a base at 0, $U_n \subset W_1$ for large n ; hence $A(U_n) \subset A(W_1) \subset W$; since W is closed, we have

$$\overline{A(U_n)} \subset W \quad \text{for sufficiently large } n. \quad (2)$$

If $n < m$,

$$\begin{aligned} A\left(\sum_{j=1}^m x_j\right) - y &= \sum_{j=1}^n y_j - y + \sum_{j=n+1}^m y_j \\ &\in \overline{A(U_{n+1})} + A(U_{n+1} + \cdots + U_m) \quad \text{by (1).} \end{aligned}$$

By (2), $A(\sum_{j=1}^m x_j) - y \in W + A(U_{n+1} + \cdots + U_m) \subset \overline{W + A(U_n)} \subset \overline{W + W}$. Let $m \rightarrow \infty$; by continuity of A we find $Ax - y \in \overline{W + W} \subset \overline{W'} = W' \subset W' + W'$, contradicting the disjointness of $Ax + W'$ and $y + W'$. ■

Under the hypothesis of 3.5.9, if $W \in \mathcal{V}$, then by continuity of A there is a set U_n in the countable base at 0 in L such that $A(U_n) \subset W$. By 3.5.9, $A(U_n) \in \mathcal{V}$ for all n ; hence M also has a countable base at 0. It is shown in the appendix on general topology (Theorem A10.8) that the existence of a countable base at 0 in a topological vector space is equivalent to pseudo-metrizability. Thus under the hypothesis of 3.5.9, M , being Hausdorff, is metrizable.

One more preliminary result is needed.

3.5.10 Lemma. Let A be a linear operator from the topological vector space L to the topological vector space M . Assume $A(L)$ is of the second category in M ; that is, $A(L)$ cannot be expressed as a countable union of nowhere dense sets. If U is an overneighborhood of 0 in L , then $\overline{A(U)}$ is an overneighborhood of 0 in M .

PROOF. Let \mathcal{U} and \mathcal{V} be the overneighborhoods of 0 in L and M , respectively. If $U \in \mathcal{U}$, choose a circled $V \in \mathcal{V}$ with $V + V \subset U$. By 3.5.1(b) applied to V

we have $L = \bigcup_{n=1}^{\infty} nV$, hence $A(L) = \bigcup_{n=1}^{\infty} nA(V)$. Since $A(L)$ is of the second category, $nA(V)$ [$= n\overline{A(V)}$] has a nonempty interior for some n , hence $\overline{A(V)}$ has a nonempty interior. If B is a nonempty open subset of $\overline{A(V)}$, then

$$\begin{aligned} 0 \in B - B &\subset \overline{A(V)} - \overline{A(V)} \subset \overline{A(V)} - A(V) \\ &= \overline{A(V) + A(V)} \quad \text{since } V, \text{ hence } A(V), \text{ is circled} \\ &\subset \overline{A(U)}. \end{aligned}$$

But $B - B = \bigcup\{x - B: x \in B\}$, an open set; hence $\overline{A(U)} \in \mathcal{V}$. ■

3.5.11 Open Mapping Theorem. Let L and M be metrizable topological vector spaces, with L complete. If A is a continuous linear operator from L to M and $A(L)$ is of the second category in M , then $A(L) = M$ and A is an open map.

PROOF. Take \mathcal{U} and \mathcal{V} as in 3.5.9 and 3.5.10. Let U be any set in \mathcal{U} ; by 3.5.9 and 3.5.10, $A(U) \in \mathcal{V}$. If $y \in M$, then $y \in nA(U)$ for some n by 3.5.1(b); hence $y = A(nx)$ for some $x \in U$, proving that $A(L) = M$. If G is open in L and $x_0 \in G$, choose $V \in \mathcal{U}$ such that $x_0 + V \subset G$. Then $A(V) \in \mathcal{V}$ and $Ax_0 + A(V) \subset A(G)$, proving that $A(G)$ is open. ■

3.5.12 Corollary. If A is a continuous linear operator from L onto M , where L and M are complete metrizable topological vector spaces, then A is an open map.

PROOF. By completeness of M , $A(L)$ ($= M$) is of category 2 in M , and the result follows from 3.5.11. ■

We conclude this section with a discussion of separation theorems and their applications. If K_1 and K_2 are disjoint convex subsets of R^3 , there is a plane P such that K_1 is on one side of P and K_2 is on the other side. Now P can be described as $\{x \in R^3: f(x) = c\}$ for some linear functional f and real number c ; hence we have $f(x) \leq c$ for all x in one of the two convex sets, and $f(x) \geq c$ for all x in the other set. We are going to consider generalizations of this idea.

The following theorem is the fundamental result of this type.

3.5.13 Basic Separation Theorem. Let K_1 and K_2 be disjoint, nonempty convex subsets of the real vector space L , and assume that K_1 has at least one

internal point. There is a linear functional f on L separating K_1 and K_2 , that is, $f \not\equiv 0$ and $f(x) \leq f(y)$ for all $x \in K_1$ and $y \in K_2$ [for short, $f(K_1) \leq f(K_2)$].

PROOF. First assume that 0 is an internal point of K_1 . Pick an element $z \in K_2$; then $-z$ is an internal point of $-z + K_1 \subset K_1 - K_2$; hence 0 is an internal point of the convex set $K = z + K_1 - K_2$. If $z \in K$, then $K_1 \cap K_2 \neq \emptyset$, contradicting the hypothesis; therefore $z \notin K$; so if p is the Minkowski functional of K , we have $p(z) \geq 1$ by 3.5.5(b). Define a linear functional g on the subspace $M = \{\lambda z : \lambda \in R\}$ by $g(\lambda z) = \lambda$. If $a > 0$, then $g(az) = a = a(1) \leq ap(z) = p(az)$, and if $a \leq 0$, then $g(az) = a \leq 0 \leq p(az)$. By the Hahn–Banach theorem, g extends to a (nontrivial) linear functional f with $f \leq p$ on L . By 3.5.8(a), $f \leq 1$ on K . Since $f(z) = g(z) = 1$, $f(K_1) \leq f(K_2)$.

If x is an internal point of K_1 , then 0 is an internal point of $-x + K_1$, and $-x + K_1$ is disjoint from $-x + K_2$. If $f(-x + K_1) \leq f(-x + K_2)$, then $f(K_1) \leq f(K_2)$. ■

If L is a complex vector space, the discussion after 3.4.2 shows that there is a nontrivial linear functional f on L such that $\operatorname{Re} f(x) \leq \operatorname{Re} f(y)$ for all $x \in K_1$, $y \in K_2$; this is what we shall mean by “ f separates K_1 and K_2 ” in the complex case.

3.5.14 Corollaries.

Assume the hypothesis of 3.5.13.

- (a) If L is a topological vector space and K_1 has nonempty interior, the linear functional f constructed in 3.5.13 is continuous.
- (b) If L is locally convex, and K_1 and K_2 disjoint convex subsets of L , with K_1 closed and K_2 compact, there is a continuous linear functional f on L strongly separating K_1 and K_2 , that is, $f(x) \leq c_1 < c_2 \leq f(y)$ for some real numbers c_1 and c_2 and all $x \in K_1$, $y \in K_2$. In particular if L is Hausdorff and $x, y \in L$, $x \neq y$, there is a continuous linear functional f on L with $f(x) \neq f(y)$.
- (c) If L is locally convex, M a closed subspace of L , and $x_0 \notin M$, there is a continuous linear functional f on L such that $f = 0$ on M and $f(x_0) \neq 0$.

PROOF. (a) Since $K_1^0 \neq \emptyset$, the interior and internal points of K_1 coincide, by 3.5.5(b) and (d). Thus in the proof of 3.5.13, 0 is an interior point of K , so that $K^0 = \{x : p(x) < 1\}$ by 3.5.5(d). Since $f \leq p$ on L it follows that $f \leq 1$ on some neighborhood U of 0 (for example, $U = K^0$); since U may be assumed circled, $f(-x) \leq 1$ for all $x \in U$; hence $|f| \leq 1$ on U . But then $|f| \leq \delta$ on the neighborhood δU , proving continuity.

(b) Let $K = K_2 - K_1$; K is convex, and is also closed since the sum of a compact set A and a closed set B in a topological vector space is closed. (If

$x_n \in A$, $y_n \in B$, $x_n + y_n \rightarrow z$, find a subnet $x_{n_i} \rightarrow x \in A$; then $y_{n_i} \rightarrow z - x$, hence $z - x \in B$ since B is closed.) Now $0 \notin K$ by disjointness of K_1 and K_2 ; so by local convexity, there is a convex circled neighborhood V of 0 such that $V \cap K = \emptyset$. By 3.5.13 and part (a) above there is an $f \in L^*$ separating V and K , say $f(V) \leq c \leq f(K)$. Now f cannot be identically 0 on V , for if so $f \equiv 0$ on L by 3.5.1(b). Thus there is an $x \in V$ with $f(x) \neq 0$, and since V is circled, we may assume $f(x) > 0$. Thus c must be greater than 0 ; hence $f(K_2) \geq f(K_1) + c$, $c > 0$, as desired.

(c) Let $K_1 = M$, $K_2 = \{x_0\}$, and apply (b) to obtain $f \in L^*$ with $f(M) \leq c_1 < c_2 \leq f(x_0)$. Since f is bounded above on the subspace M , f must be identically 0 on M . [If $f(x) \neq 0$, consider $f(\lambda x)$ for arbitrarily large λ .] But then $f(x_0) \geq c_2 > 0$. ■

If we adopt the above definition of separation in complex vector spaces, all parts of 3.5.14 extend immediately to the complex case.

Separation theorems may be applied effectively in the study of weak topologies. In 3.4.10 we defined the weak topology on a normed linear space; the definition is identical for an arbitrary topological vector space L . Specifically, for each $f \in L^*$, $p_f(x) = |f(x)|$ defines a seminorm on L . The locally convex topology induced by the seminorms p_f , $f \in L^*$, is called the *weak topology* on L . By 3.5.1 and 3.5.2, a base at x_0 for the weak topology consists of finite intersections of sets of the form $\{x : p_f(x - x_0) < \epsilon\}$, so in the case of a normed linear space we obtain the topology defined in 3.4.10.

There is a dual topology defined on L^* ; if $x \in L$, then $p_x(f) = |f(x)|$ defines a seminorm on L^* . The locally convex topology induced by the seminorms p_x is called the *weak* topology* on L^* .

By 3.5.2, the weak topology is the weakest topology on L making each $f \in L^*$ continuous, so the weak topology is weaker than the original topology of L . Convergence of x_n to x in the weak topology means $f(x_n) \rightarrow f(x)$ for each $f \in L^*$. The weak* topology on L^* is the weakest topology making all evaluation maps $f \rightarrow f(x)$ continuous. Convergence of f_n to f in the weak* topology means $f_n(x) \rightarrow f(x)$ for all $x \in L$; thus weak* convergence is simply pointwise convergence, so if L is a normed linear space, the weak* topology is weaker than the norm topology on L^* .

We have observed in 3.4.10 that the weak topology is an example of a product topology. Since weak* convergence is pointwise convergence, the weak* topology is the product topology on the set C^L of all complex-valued functions on L , relativized to L^* .

In distinguishing between the weak topology and the original topology on L , it will be convenient to call the original topology the *strong topology*.

By the above discussion, a weakly closed subset of L is closed in the original topology. Under certain conditions there is a converse statement:

3.5.15 Theorem. Let L be a locally convex topological vector space. If K is a convex subset of L , then K is strongly closed in L iff it is weakly closed.

PROOF. Assume K strongly closed. If $y \notin K$, then by 3.5.14(b) there are real numbers c_1 and c_2 and an $f \in L^*$ with $\operatorname{Re} f(x) \leq c_1 < c_2 \leq \operatorname{Re} f(y)$ for all $x \in K$. But then $W = \{x \in L : |f(x) - f(y)| < c_2 - c_1\}$ is a weak neighborhood of y , and if $x \in K$, we have $|f(x - y)| \geq |\operatorname{Re} f(y) - \operatorname{Re} f(x)| \geq c_2 - c_1$; therefore $W \cap K = \emptyset$, proving K^c weakly open. ■

For the remainder of this section we consider normed linear spaces. The closed unit ball $\{f : \|f\| \leq 1\}$ of L^* is never compact (in the strong topology) unless L (hence L^*) is finite-dimensional; see Problem 9, Section 3.3. However, it is always compact in the weak* topology, as we now prove:

3.5.16 Banach-Alaoglu Theorem. If L is a normed linear space and $B = \{f \in L^* : \|f\| \leq 1\}$, then B is compact in the weak* topology.

PROOF. If $f \in B$, then $|f(x)| \leq \|f\| \|x\| \leq \|x\|$ for all $x \in L$. If $I(x)$ is the set $\{z \in C : |z| \leq \|x\|\}$, then $B \subset \prod\{I(x) : x \in L\}$, the set of all functions defined on L such that $f(x) \in I(x)$ for each $x \in L$. By the Tychonoff theorem (see the appendix on general topology, Theorem A5.4), $\prod\{I(x) : x \in L\}$ is compact in the product topology (the topology of pointwise convergence).

Let $\{f_n\}$ be a net in B ; by the above discussion there is a subnet converging pointwise to a complex-valued function f on L , and since the f_n are linear so is f . Now $|f_n(x)| \leq \|x\|$ for each x , and it follows that $|f(x)| \leq \|x\|$, hence $f \in B$, proving compactness. ■

If L is reflexive, then L and L^{**} may be identified, and the weak topology on L corresponds to the weak* topology on L^{**} . By 3.5.16, the closed unit ball of L is weakly compact. Conversely, weak compactness of the closed unit ball implies reflexivity. To prove this we need the following auxiliary results:

3.5.17 Lemma. Let L be a topological vector space.

(a) If f is a linear functional on L , f is continuous relative to the weak topology iff f is continuous relative to the strong topology, that is, iff $f \in L^*$.

(b) If g is a linear functional on L^* , g is continuous relative to the weak* topology iff there is an $x \in L$ such that $g(f) = f(x)$ for all $f \in L^*$.

PROOF. (a) If $f \in L^*$ and x_n converges weakly to x , then $f(x_n) \rightarrow f(x)$ by definition of the weak topology, so f is weakly continuous. Conversely, if f is continuous relative to the weak topology and x_n converges strongly to x , then x_n converges weakly to x , hence $f(x_n) \rightarrow f(x)$. Thus $f \in L^*$.

(b) The “if” part follows from the definition of the weak* topology, so assume g weak* continuous. A basic neighborhood of 0 in the weak* topology is of the form $U = \bigcap_{i=1}^n \{f \in L^*: |f(x_i)| < \delta_i, i = 1, \dots, n\}$ for some $x_1, \dots, x_n \in L$, $\delta_1, \dots, \delta_n > 0$, $n = 1, 2, \dots$; thus there is a U of this type such that $|g(f)| < 1$ for all $f \in U$. By linearity, $|f(x_i)| < \delta_i/M$ for all i implies $|g(f)| < 1/M$. In particular, if $f(x_i) = 0$ for all $i = 1, \dots, n$, then $|g(f)| < 1/M$, and since M is arbitrary, $g(f) = 0$. But this implies, by a standard algebraic result (see Problem 15) that $g(f) = \sum_{i=1}^n c_i f(x_i)$ for some complex numbers c_1, \dots, c_n . Set $x = \sum_{i=1}^n c_i x_i$. ■

3.5.18 Lemma. Let L be a normed linear space, and let h be the isometric isomorphism of L into L^{**} (see 3.4.6). Let B be the closed unit ball in L , B^{**} the closed unit ball in L^{**} . Then $h(B)$ is dense in B^{**} relative to the weak* topology. Furthermore, $h(L)$ is dense in L^{**} .

PROOF. Let M be the weak* closure of $h(B)$, and let $x^{**} \in B^{**}$, $x^{**} \notin M$. By 3.5.16, B^{**} is weak* compact (and convex); so by 3.5.14(b), there is a linear functional g on L^{**} such that g is continuous relative to the weak* topology and $\operatorname{Re} g(M) \leq c_1 < c_2 \leq \operatorname{Re} g(x^{**})$. Since M is a subspace (in particular is closed), $|g(y^{**})| \leq c_1$ for all $y^{**} \in M$. [If $y^{**} \in M$ and $g(y^{**}) = re^{i\theta}$, then $|g(y^{**})| = g(e^{-i\theta}y^{**}) = r = \operatorname{Re} g(e^{-i\theta}y^{**}) \leq c_1$.] By 3.5.17(b), g is of the form $g(y^{**}) = y^{**}(f)$ for some $f \in L^*$.

If $y \in B$, and $y^{**} = h(y)$ is the corresponding element of $h(B) \subset M$, then $|f(y)| = |y^{**}(f)| \leq c_1$; hence

$$\|f\| \leq c_1 < c_2 \leq \operatorname{Re} g(x^{**}) \leq |x^{**}(f)| \leq \|x^{**}\| \|f\| \leq \|f\|,$$

a contradiction. Consequently $M = B^{**}$.

Finally, if $y^{**} \in L^{**}$, $y^{**} \neq 0$, then $y^{**}/\|y^{**}\| \in B^{**}$, which we have just seen is the weak* closure of $h(B)$. Thus $y^{**}/\|y^{**}\|$ belongs to the weak* closure of $h(L)$, which is a subspace. The result follows. ■

We now characterize reflexive normed linear spaces.

3.5.19 Theorem. A normed linear space L is reflexive iff its closed unit ball B is weakly compact.

PROOF. The “only if” part has been pointed out after 3.5.16, so assume B weakly compact. The isometric isomorphism h of L into L^{**} is a homeomorphism of the weak topology of L and the weak* topology of $h(L)$ [note that $x_n \xrightarrow{w} x$ iff $f(x_n) \rightarrow f(x)$ for all $f \in L^*$; in other words, iff $h(x_n) \xrightarrow{w^*} h(x)$]. Thus $h(B)$ is a weak* compact (hence closed) subset of B^{**} , and is dense in B^{**} by 3.5.18. It follows that $h(B) = B^{**}$, so as in 3.5.18, $h(L) = L^{**}$, proving reflexivity. ■

3.5.20 Corollaries. (a) If M is a closed subspace of the reflexive normed linear space L , then M is reflexive.

(b) If L is a Banach space, L is reflexive iff L^* is reflexive.

PROOF. (a) If B is the closed unit ball in L , then $B \cap M$ is the closed unit ball in M . By hypothesis, $B \cap M$ is strongly closed, hence weakly closed by 3.5.15. But B is weakly compact by 3.5.19, so $B \cap M$ is weakly compact. Again by 3.5.19, M is reflexive.

(b) Assume L reflexive. The weak* topology on L^* then coincides with its weak topology, hence by 3.5.16, the closed unit ball of L^* is weakly compact, and consequently L^* is reflexive by 3.5.19. (Completeness of L is not used in this part.)

If L^* is reflexive, so is L^{**} by the first part of the proof. Since L is complete, so is $h(L)$; thus $h(L)$ is a closed subspace of L^{**} , and the result follows from (a). ■

Problems

- Let \mathcal{U} be the family of overneighborhoods of 0 in a topological vector space L . Show that L is Hausdorff iff $\bigcap\{U: U \in \mathcal{U}\} = \{0\}$.
- If f is a linear functional on the topological vector space L , with f not identically 0, show that the four conditions of Problem 1, Section 3.3, are still equivalent.
- If L is a finite-dimensional Hausdorff topological vector space, show that L is topologically isomorphic to C^n for some n . Show also that a finite-dimensional subspace of an arbitrary Hausdorff topological vector space is always closed. [It can be shown that for a Hausdorff topological vector space, finite-dimensionality is equivalent to local compactness; see Kelley and Namioka (1963, Theorem 7.8).]
- (a) Let K be a convex set, radial at 0, in the vector space L . If g is any nonnegative sublinear functional on L such that $\{x \in L: g(x) < 1\} \subset K \subset \{x \in L: g(x) \leq 1\}$, show that g is the Minkowski functional of K . [Thus by 3.5.5(d), if $0 \in K^0$, then the Minkowski functionals of K^0 , K , and \bar{K} coincide.]

- (b) Use part (a) to find the Minkowski functionals of each of the following subsets of R^2 :
- (i) $\{(x, y): |x| + |y| \leq 1\}$.
 - (ii) $\{(x, y): -1 \leq x < 1, -1 \leq y < 1\}$
 - (iii) $\{(x, y): x^2 + y^2 < 1\}$.
- (c) If K_1 and K_2 are convex and radial at 0, with Minkowski functionals p_1 and p_2 , show that the Minkowski functional of $K_1 \cap K_2$ is $\max(p_1, p_2)$.
5. Let L be the set of all complex-valued functions on $[0, 1]$, with the topology of pointwise convergence. Then L is a locally convex topological vector space, with topology induced by the seminorms $p_x(f) = |f(x)|$, $f \in L$, where x ranges over $[0, 1]$; see 3.5.3(a). Show that L is not metrizable. (The same result holds even if we restrict L to the continuous complex-valued functions on $[0, 1]$.)
6. Let L be a locally convex topological vector space. Show (without using the theory of uniform spaces given in the appendix on general topology, Section A10) that the following conditions are equivalent:
- (a) L is pseudometrizable, that is, there is a pseudometric inducing the topology of L .
 - (b) L has a countable base at 0 (hence at every point of L).
 - (c) The topology of L is generated by a countable family of seminorms.
7. Let L be the collection of analytic functions on the unit disk $D = \{z: |z| < 1\}$, with the topology of uniform convergence on compact subsets [see 3.5.3(a)]. Show that L is metrizable but not normable.
8. Let $L = L^p[0, 1]$ where $0 < p < 1$; with the pseudometric $d(x, y) = \int_0^1 |x(t) - y(t)|^p dt$ (see 2.4.12), L is a topological vector space. [Apply 3.5.1, using the sets $U = \{x: \int_0^1 |x(t)|^p dt < \varepsilon\}$, $\varepsilon > 0$.]
- (a) Let f be a continuous linear functional on L . If $f(x) = 1$, show that x can be written as $y + z$ where $d(y, 0) = d(z, 0) = \frac{1}{2}d(x, 0)$.
 - (b) With f as in part (a), show that there are functions $x_1, x_2, \dots \in L$ such that $d(x_n, 0) \rightarrow 0$ as $n \rightarrow \infty$, but $|f(x_n)| \geq 1$ for all n . Conclude that the only continuous linear functional on L is the zero functional. [In part (a), either $|f(y)| \geq \frac{1}{2}$ or $|f(z)| \geq \frac{1}{2}$; if, say, $|f(y)| \geq \frac{1}{2}$, take $x_1 = 2y$ and proceed inductively.]
9. Let L be a topological vector space. Show that there is a nontrivial continuous linear functional on L iff there is a proper convex subset of L with nonempty interior.
10. Let B_1 and B_2 be (not necessarily convex) nonempty subsets of the topological vector space L , with $B_1^0 \neq \emptyset$. If f is a linear functional on

- L that separates B_1 and B_2 , show that f is continuous. [This gives another proof of 3.5.14(a).]
11. Let K be a convex subset of a locally convex topological vector space L . Show that the closure of K in the weak topology coincides with the closure in the strong topology.
 12. Let K be a convex set, radial at 0, in the real vector space L . Let x be a point of L and M a subspace of L such that $x + M$ does not contain any internal points of K . Show that there is a hyperplane H (that is, H is the null space of a nontrivial linear functional on L) such that $M \subset H$ and $x + H$ contains no internal points of K . [Define g on $L(M \cup \{x\})$ by $g(y + ax) = a$, and extend using the Hahn–Banach theorem.]
 13. (a) If L is a topological vector space and $\mathcal{T}_1, \mathcal{T}_2$ are topologies on L such that $\mathcal{T}_1 \subset \mathcal{T}_2$ and L is complete and metrizable under both \mathcal{T}_1 and \mathcal{T}_2 , show that $\mathcal{T}_1 = \mathcal{T}_2$.
 (b) If L is a Banach space under two norms $\|\cdot\|_1$ and $\|\cdot\|_2$, and, for some $k > 0$, $\|x\|_1 \leq k\|x\|_2$ for all $x \in L$, show that for some $r > 0$, $\|x\|_2 \leq r\|x\|_1$ for all $x \in L$; in other words (Problem 6, Section 3.3) the norms induce the same topology.
 14. (a) (*Closed graph theorem*) Let A be a linear map from L to M , where L and M are complete metrizable topological vector spaces. Assume A is closed; in other words, the graph of A is a closed subset of $L \times M$, with the product topology. Show that A is continuous.
 (b) If A is a continuous linear operator from L to M where L and M are topological vector spaces and M is Hausdorff, show that A is closed.
 15. Let g, f_1, \dots, f_n be linear functionals on the vector space L . If N denotes null space and $\bigcap_{i=1}^n N(f_i) \subset N(g)$, show that g is a linear combination of the f_i .
 16. Let L be a normed linear space. If the weak and strong topologies coincide on L , show that L is finite-dimensional, as follows:
 - (a) The unit ball $B = \{x: \|x\| < 1\}$ is strongly open, hence weakly open by hypothesis. Thus we can find $f_1, \dots, f_n \in L^*$ and $\delta_1, \dots, \delta_n > 0$ such that $\{x: |f_i(x)| < \delta_i, i = 1, \dots, n\} \subset B$. Show that [with $N(f)$ denoting the null space of f] $\bigcap_{i=1}^n N(f_i) = \{0\}$.
 - (b) Define $T(x) = (f_1(x), \dots, f_n(x))$; by (a), T is a one-to-one map of L onto a subspace M of C^n . If $\{y_1, \dots, y_k\}$ is a basis for M and $x_j = T^{-1}(y_j)$, $j = 1, \dots, k$, show that $\{x_1, \dots, x_k\}$ is a basis for L , hence L is finite-dimensional.
 17. Let L be a separable normed linear space. If B is the closed unit ball in L^* , show that B is metrizable in the weak* topology. Thus since B is compact by 3.5.16, every sequence in B (hence every norm-bounded

sequence in L^*) has a weak* convergent subsequence. Similarly, if L is reflexive, the closed unit ball of L is metrizable in the weak topology. (Let $\{x_1, x_2, \dots\}$ be a countable dense subset of L , and set

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|f(x_n) - g(x_n)|}{1 + |f(x_n) - g(x_n)|}, \quad f, g \in L^*.$$

Show that weak* convergence implies d -convergence.)

18. Show that a reflexive Banach space is weakly complete; in other words every weak Cauchy sequence $\{f(x_n)\}$ is Cauchy for each $f \in L^*$ converges weakly.
19. A subset B of a topological vector space L is said to be *absorbed* by a subset A iff $B \subset aA$ for sufficiently large $|a|$; B is said to be *bounded* iff it is absorbed by every neighborhood of 0.
 - (a) Show that B is bounded iff for each sequence of points $x_n \in B$ and each sequence of complex numbers $\lambda_n \rightarrow 0$, we have $\lambda_n x_n \rightarrow 0$.
 - (b) A *bornivore* in a topological vector space L is a convex circled set that absorbs every bounded set; L is *bornological* iff L is locally convex and every bornivore is an overneighborhood of 0. Show that every metrizable locally convex space is bornological.
 - (c) Let T be a linear operator from L to M , where L is bornological and M is locally convex. If T maps bounded sets into bounded sets, show that T is continuous. (The converse is true for arbitrary L and M .)
 - (d) If L is a Hausdorff topological vector space, show that there is no bounded subspace of L except $\{0\}$.

3.6 References

There is a vast literature on functional analysis, and we only give a few representative titles. Readable introductory treatments are given in Liusternik and Sobolev (1961), Taylor (1958), Bachman and Narici (1966), and Halmos (1951); the last deals exclusively with Hilbert spaces. Among the more advanced treatments, Dunford and Schwartz (1958, 1963, 1970) emphasize normed spaces, Kelley and Namioka (1963) and Schaefer (1966) emphasize topological vector spaces. Yosida (1968) gives a broad survey of applications to differential equations, semigroup theory, and other areas of analysis.

4

The Interplay between Measure Theory and Topology

4.1 Introduction

A connection between measure theory and topology is established when a σ -field \mathcal{F} is defined in terms of topological properties. In the most common situation, we have a topological space Ω , and \mathcal{F} is taken as the smallest σ -field containing all open sets of Ω . If this is done, there is a natural connection between measure-theoretic and topological questions. For example, if μ is a measure on \mathcal{F} and $A \in \mathcal{F}$, we may ask whether A can be approximated by a compact subset. In other words, we wish to know if $\mu(A) = \sup\{\mu(K) : K \text{ a compact subset of } A\}$. As another example, we may ask whether a function in $L^p(\Omega, \mathcal{F}, \mu)$ can be approximated by a continuous function. One formulation of this is to ask whether the continuous functions are dense in L^p .

In this chapter we investigate questions of this type. The results in the first two sections are not topological, but they serve as basic tools in the later development. We first consider a result that is a companion to the monotone class theorem:

4.1.1 Definition. Let \mathcal{D} be a class of subsets of a set Ω . Then \mathcal{D} is said to be a *Dynkin system* (D-system for short) iff the following conditions hold.

- (a) $\Omega \in \mathcal{D}$.
- (b) If $A, B \in \mathcal{D}$, $B \subset A$, then $A - B \in \mathcal{D}$. Thus \mathcal{D} is closed under proper differences.
- (c) If $A_1, A_2, \dots \in \mathcal{D}$ and $A_n \uparrow A$, then $A \in \mathcal{D}$.

Note that by (a) and (b), \mathcal{D} is closed under complementation; hence by (c), \mathcal{D} is a monotone class. If \mathcal{D} is closed under finite union (or closed under finite intersection), then \mathcal{D} is a field, and hence a σ -field (see 1.2.1).

4.1.2 Dynkin System Theorem. Let \mathcal{S} be a class of subsets of Ω , and assume \mathcal{S} closed under finite intersection. If \mathcal{D} is a Dynkin system and $\mathcal{D} \supset \mathcal{S}$, then \mathcal{D} includes the minimal σ -field $\mathcal{F} = \sigma(\mathcal{S})$.

PROOF. Let \mathcal{D}_0 be the smallest D-system including \mathcal{S} . We show that $\mathcal{D}_0 = \mathcal{F}$, in other words *the smallest D-system and the smallest σ -field over a class closed under finite intersection coincide*. Since $\mathcal{D}_0 \subset \mathcal{D}$, the result will follow.

Now $\mathcal{D}_0 \subset \mathcal{F}$ since \mathcal{F} is a D-system. To show that $\mathcal{F} \subset \mathcal{D}_0$, let $\mathcal{C} = \{A \in \mathcal{D}_0 : A \cap B \in \mathcal{D}_0 \text{ for all } B \in \mathcal{S}\}$. Then $\mathcal{S} \subset \mathcal{C}$ since \mathcal{S} is closed under finite intersection, and since \mathcal{D}_0 is a D-system, so is \mathcal{C} . Thus $\mathcal{D}_0 \subset \mathcal{C}$, hence $\mathcal{D}_0 = \mathcal{C}$.

Now let $\mathcal{C}' = \{C \in \mathcal{D}_0 : C \cap D \in \mathcal{D}_0 \text{ for all } D \in \mathcal{D}_0\}$. The result $\mathcal{D}_0 = \mathcal{C}$ implies that $\mathcal{S} \subset \mathcal{C}'$, and since \mathcal{C}' is a D-system we have $\mathcal{D}_0 \subset \mathcal{C}'$, hence $\mathcal{D}_0 = \mathcal{C}'$. It follows that \mathcal{D}_0 is closed under finite intersection; by 4.1.1, \mathcal{D}_0 is a σ -field, so that $\mathcal{F} \subset \mathcal{D}_0$. ■

If \mathcal{S} is a field and \mathcal{M} is the smallest monotone class including \mathcal{S} , then $\mathcal{M} = \sigma(\mathcal{S})$ (see 1.3.9). In the Dynkin system theorem, we have a weaker hypothesis on \mathcal{S} (it is closed under finite intersection but need not be a field) but a stronger hypothesis on the class of sets including \mathcal{S} (\mathcal{D} is a Dynkin system, not merely a monotone class).

4.1.3 Corollary. Let \mathcal{S} be a class of subsets of Ω , and let μ_1 and μ_2 be finite measures on $\sigma(\mathcal{S})$. Assume $\Omega \in \mathcal{S}$ and \mathcal{S} is closed under finite intersection. If $\mu_1 = \mu_2$ on \mathcal{S} , then $\mu_1 = \mu_2$ on $\sigma(\mathcal{S})$.

PROOF. Let \mathcal{D} be the collection of sets $A \in \sigma(\mathcal{S})$ such that $\mu_1(A) = \mu_2(A)$. Then \mathcal{D} is a D-system and $\mathcal{S} \subset \mathcal{D}$, hence $\mathcal{D} = \sigma(\mathcal{S})$ by 4.1.2. ■

4.1.4 Corollary. Let \mathcal{S} be a class of subsets of Ω ; assume that $\Omega \in \mathcal{S}$ and \mathcal{S} is closed under finite intersection. Let H be a vector space of real-valued functions on Ω , such that $I_A \in H$ for each $A \in \mathcal{S}$.

Suppose that whenever f_1, f_2, \dots are nonnegative functions in H , $|f_n| \leq M < \infty$ for all n , and $f_n \uparrow f$, the limit function f belongs to H . Then $I_A \in H$ for all $A \in \sigma(\mathcal{S})$.

PROOF. Let $\mathcal{D} = \{A \subset \Omega : I_A \in H\}$; then $\mathcal{S} \subset \mathcal{D}$ and \mathcal{D} is a D-system. For if $A, B \in \mathcal{D}$, $B \subset A$, then $I_{A-B} = I_A - I_B \in H$, so that $A - B \in \mathcal{D}$. If $A_n \in \mathcal{D}$ and $A_n \uparrow A$, then $I_{A_n} \uparrow I_A$, hence $I_A \in H$ by hypothesis; thus $A \in \mathcal{D}$. The result now follows from 4.1.2. ■

4.2 The Daniell Integral

One of the basic properties of the integral is linearity; if f and g are μ -integrable functions and a and b are real or complex numbers, then

$$\int_{\Omega} (af + bg) d\mu = a \int_{\Omega} f d\mu + b \int_{\Omega} g d\mu.$$

Thus the integral may be regarded as a linear functional on the vector space of integrable functions. This idea may be used as the basis for a different approach to integration theory. Instead of beginning with a measure and constructing the corresponding integral, we start with a given linear functional E on a vector space. Under appropriate hypotheses, we extend E to a larger space, and finally we show that there is a measure μ such that E is in fact the integral with respect to μ .

We first fix the notation to be used.

4.2.1 Notation. In this section, L will denote a vector space of real-valued functions on a set Ω ; L is assumed *closed under the lattice operations*, in other words if $f, g \in L$ and $f \vee g = \max(f, g), f \wedge g = \min(f, g)$, then $f \vee g, f \wedge g \in L$. There are several familiar examples of such spaces; L can be the class of continuous real-valued functions on a given topological space, or equally well, the bounded continuous real-valued functions. Another possibility is to take L as the collection of all real-valued functions on a given set.

The letter E will denote a *positive linear functional* on L , that is, a linear map E from L to R such that $f \geq 0$ implies $E(f) \geq 0$. This implies that E is monotone, that is, $f \leq g$ implies $E(f) \leq E(g)$.

If H is any class of functions from Ω to R (or \bar{R}), H^+ will denote $\{f \in H : f \geq 0\}$. The collection of functions $f : \Omega \rightarrow \bar{R}$ of the form $\lim_n f_n$ where the f_n form an increasing sequence of functions in L^+ , will be denoted by L' ; if the f_n are allowed to form an increasing net in L^+ , the resulting class is denoted by L'' . (See the appendix on general topology, Section A1, for a discussion of nets.)

If H is as above, $\sigma(H)$ is defined as the smallest σ -field of subsets of Ω making every function in H Borel measurable, that is, the minimal σ -field containing all sets $f^{-1}(B)$, where f ranges over H and B over the Borel sets. If \mathcal{G} is a class of subsets of Ω , $\sigma(\mathcal{G})$ as usual denotes the minimal σ -field over \mathcal{G} .

The following will be assumed throughout:

Hypothesis A: If the functions f_n form a sequence in L decreasing to 0, then $E(f_n)$ decreases to 0. Equivalently, if the f_n belong to L and increase to $f \in L$, then $E(f_n)$ increases to $E(f)$.

We are also going to carry through a parallel development under the following assumption:

Hypothesis B: If the functions f_n form a net in L decreasing to 0, then $E(f_n)$ decreases to 0 (with the equivalent statement just as in hypothesis A).

Hypothesis A is always assumed in the statements of theorems. Corresponding results under hypothesis B will be added in brackets.

4.2.2 Lemma. Let $\{f_m\}$ and $\{f'_n\}$ be sequences in L increasing to f and f' , respectively, with $f \leq f'$ (f and f' need not belong to L). Then

$$\lim_m E(f_m) \leq \lim_n E(f'_n).$$

Hence E may be extended to L' by defining $E(\lim_n f_n) = \lim_n E(f_n)$. [Under hypothesis B (and with “sequence” replaced by “net” in the above statement), E may be extended to L' in the same fashion.]

PROOF. As $n \rightarrow \infty$, $f_m \wedge f'_n \uparrow f_m \wedge f' = f_m$; hence $\lim_n E(f'_n) \geq \lim_n E(f_m \wedge f'_n) = E(f_m)$. Let $m \rightarrow \infty$ to finish the proof. [Under hypothesis B, the proof is the same, with sequences replaced by nets.] ■

We now study the extension of E .

4.2.3 Lemma. The extension of E to L' has the following properties:

- (a) $0 \leq E(f) \leq \infty$ for all $f \in L'$.
- (b) If $f, g \in L'$, $f \leq g$, then $E(f) \leq E(g)$.
- (c) If $f \in L'$ and $0 \leq c < \infty$, then $cf \in L'$ and $E(cf) = cE(f)$.
- (d) If $f, g \in L'$, then $f + g, f \vee g, f \wedge g \in L'$ and

$$E(f + g) = E(f \vee g) + E(f \wedge g) = E(f) + E(g).$$

- (e) If $\{f_n\}$ is a sequence in L' increasing to f , then $f \in L'$ and $E(f_n)$ increases to $E(f)$.

[Under hypothesis B, the extension of E to L'' has exactly the same properties; “sequence” is replaced by “net” in (e).]

PROOF. Properties (a)–(c) follow from 4.2.2. To prove (d), let $f_n, g_n \in L^+$, with $f_n \uparrow f, g_n \uparrow g$. Then

$$\begin{aligned} E(f_n \vee g_n) + E(f_n \wedge g_n) &= E[(f_n \vee g_n) + (f_n \wedge g_n)] \\ &\quad \text{by linearity} \\ &= E(f_n + g_n) \\ &\quad \text{(the sum of two numbers is} \\ &\quad \text{the larger plus the smaller)} \\ &= E(f_n) + E(g_n) \\ &\quad \text{by linearity.} \end{aligned}$$

Let $n \rightarrow \infty$ to obtain (d). To prove (e), let $f_{nm} \in L^+, f_{nm} \uparrow f_n$ as $m \rightarrow \infty$. If $g_m = f_{1m} \vee \dots \vee f_{mm}$, the g_m form an increasing sequence in L^+ , and

$$f_{nm} \leq g_m \leq f_m \quad \text{for } n \leq m; \tag{1}$$

hence

$$E(f_{nm}) \leq E(g_m) \leq E(f_m), \quad n \leq m. \tag{2}$$

In (1) let $m \rightarrow \infty$ to obtain $f_n \leq \lim_m g_m \leq \lim_m f_m = f$; then let $n \rightarrow \infty$ to obtain $g_m \uparrow f$, hence $E(g_m) \uparrow E(f)$. Now let $m \rightarrow \infty$ in (2) to find that $E(f_n) \leq E(f) \leq \lim_m E(f_m)$; then let $n \rightarrow \infty$ to establish that $\lim_n E(f_n) = E(f)$.

[Under hypothesis B, (a)–(d) are proved as above, but (e) is more complicated. Let $\{f_n\}$ be a net in L'' with $f_n \uparrow f$; for each n there is a net $\{f_{nm}\}$ in L^+ with $f_{nm} \uparrow f_n$, so that $f = \sup_{n,m} f_{nm}$. Let D consist of all finite sets of pairs (n, m) . Direct D by inclusion, and define $g_M = \max\{f_{nm} : (n, m) \in M\}$, $M \in D$. The g_M form a monotone net in L^+ and $g_M \uparrow f$; therefore $f \in L''$ and, by 4.2.2, $E(g_M) \uparrow E(f)$. Now for each M , $E(g_M) \leq \sup_{n,m} E(f_{n,m})$, and since $f_{nm} \leq f_n$ for all m , it follows that $E(g_M) \leq \sup_n E(f_n) = \lim_n E(f_n)$. But M is arbitrary, hence $E(f) \leq \lim_n E(f_n)$. The reverse inclusion holds because $f_n \uparrow f$, and the proof is complete.] ■

We now begin the construction of a measure derived from the linear functional E . From now on we assume that all constant functions belong to L and, as a normalization, $E(1) = 1$ (hence $E(c) = c$ for all c).

4.2.4 Lemma. Let $\mathcal{G} = \{G \subset \Omega : I_G \in L'\}$ and define $\mu(G) = E(I_G)$, $G \in \mathcal{G}$. Then \mathcal{G} satisfies the four conditions of 1.3.2, namely:

- (a) $\emptyset, \Omega \in \mathcal{G}$, $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$, $0 \leq \mu(A) \leq 1$ for all $A \in \mathcal{G}$.
- (b) If $G_1, G_2 \in \mathcal{G}$, then $G_1 \cup G_2, G_1 \cap G_2 \in \mathcal{G}$, and

$$\mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) = \mu(G_1) + \mu(G_2).$$

- (c) If $G_1, G_2 \in \mathcal{G}$ and $G_1 \subset G_2$, then $\mu(G_1) \leq \mu(G_2)$.
- (d) If $G_n \in \mathcal{G}$, $n = 1, 2, \dots$ and $G_n \uparrow G$, then $G \in \mathcal{G}$ and $\mu(G_n) \uparrow \mu(G)$.

Thus by 1.3.3 and 1.3.5, $\mu^*(A) = \inf\{\mu(G) : G \in \mathcal{G}, G \supset A\}$ is a probability measure on the σ -field $\mathcal{H} = \{H \subset \Omega : \mu^*(H) + \mu^*(H^c) = 1\}$, and $\mu^* = \mu$ on \mathcal{G} .

[Under hypothesis B, we take $\mathcal{G} = \{G \subset \Omega : I_G \in L''\}$ and replace sequences by nets in 4.2.4(d). The class \mathcal{G} then has exactly the same properties.]

PROOF. Part (a) follows since L contains the constant functions and $E(c) = c$. Part (b) follows from 4.2.3(d) with $f = I_{G_1}$, $g = I_{G_2}$. Part (c) follows from 4.2.3(b), and part (d) from 4.2.3(e). [The proof under hypothesis B is identical.] ■

We now investigate Borel measurability relative to the σ -field $\sigma(\mathcal{G})$.

4.2.5 Lemma. If $f \in L'$ and $a \in R$, then $\{f(\omega) > a\} \in \mathcal{G}$. Hence

$$f : (\Omega, \sigma(\mathcal{G})) \rightarrow (\bar{R}, \mathcal{B}(\bar{R})).$$

[The same result holds for $f \in L''$ under hypothesis B.]

PROOF. Let $f_n \in L^+, f_n \uparrow f$. Then $(f_n - a)^+ = (f_n - a) \vee 0 \in L^+ \subset L'$, and $(f_n - a)^+ \uparrow (f - a)^+$. By 4.2.3(e) (or the definition of L'), $(f - a)^+ \in L'$; hence $k(f - a)^+ \in L'$ for each positive integer k , by 4.2.3(c). But as $k \rightarrow \infty$,

$$1 \wedge k(f - a)^+ \uparrow I_{\{f > a\}};$$

so by 4.2.3(e), $I_{\{f > a\}} \in L'$, and therefore $\{f > a\} \in \mathcal{G}$. [Under hypothesis B, the proof is the same, with $\{f_n\}$ a net instead of a sequence.] ■

4.2.6 Lemma. The σ -fields $\sigma(L)$, $\sigma(L')$, and $\sigma(\mathcal{G})$ are identical. [Under hypothesis B we only have $\sigma(L'') = \sigma(\mathcal{G})$ and $\sigma(L) \subset \sigma(L'')$.]

PROOF. By 4.2.5, $\sigma(\mathcal{G})$ makes every function in L' measurable; hence $\sigma(L') \subset \sigma(\mathcal{G})$. If $G \in \mathcal{G}$, then $I_G \in L'$; hence $G = \{I_G = 1\} \in \sigma(L')$; therefore $\sigma(\mathcal{G}) \subset \sigma(L')$. [Under hypothesis B, $\sigma(L'') = \sigma(\mathcal{G})$ by the same argument.]

If $f \in L$, then $f = f^+ - f^-$, where $f^+, f^- \in L^+ \subset L'$. Since f^+ and f^- are $\sigma(L')$ -measurable, so is f . In other words, $\sigma(L')$ makes every function in L Borel measurable, so that $\sigma(L) \subset \sigma(L')$. [Under hypothesis B, $\sigma(L) \subset \sigma(L'')$ by the same argument.] Now if $f \in L'$, then f is the limit of a sequence $f_n \in L$, and since the f_n are $\sigma(L)$ -measurable, so is f . [This fails under hypothesis B because the limit of a net of measurable functions need not be measurable; see Problem 1.] Thus $\sigma(L') \subset \sigma(L)$. ■

Now by definition of the set function μ^* (see 4.2.4) we have, for all $A \subset \Omega$,

$$\begin{aligned}\mu^*(A) &= \inf\{E(I_G): G \in \mathcal{G}, G \supset A\} \\ &= \inf\{E(f): f = I_G \in L', f \geq I_A\} \\ &\geq \inf\{E(f): f \in L', f \geq I_A\}.\end{aligned}$$

In fact equality holds.

4.2.7 Lemma. For any $A \subset \Omega$, $\mu^*(A) = \inf\{E(f): f \in L', f \geq I_A\}$. [The result is the same under hypothesis B, with L' replaced by L'' .]

PROOF. Let $f \in L', f \geq I_A$. If $0 < a < 1$, then $A \subset \{f > a\}$, which belongs to \mathcal{G} by 4.2.5. Thus $\mu^*(A) \leq \mu(\{f > a\}) = E(I_{\{f > a\}})$. But since $f \geq 0$ we have $f \geq aI_{\{f > a\}}$; hence $E(I_{\{f > a\}}) \leq E(f)/a$. Let $a \rightarrow 1$ to conclude that $\mu^*(A) \leq E(f)$. [The proof is the same under hypothesis B.] ■

We may now prove that μ^* is a measure on $\sigma(\mathcal{G})$:

4.2.8 Lemma. If $\mathcal{H} = \{H \subset \Omega: \mu^*(H) + \mu^*(H^c) = 1\}$, then $\mathcal{G} \subset \mathcal{H}$, hence $\sigma(\mathcal{G}) \subset \mathcal{H}$. [The result is the same under hypothesis B.]

PROOF. Let $G \in \mathcal{G}$; since $I_G \in L'$, there are functions $f_n \in L^+$ with $f_n \uparrow I_G$. Then

$$\mu^*(G) = \mu(G) = E(I_G) = \lim_n E(f_n) \quad \text{by 4.2.3(e).}$$

By 4.2.7, $\mu^*(G^c) = \inf\{E(f): f \in L', f \geq I_{G^c}\}$. But if $f_n \leq I_G$, then $1 - f_n \geq I_{G^c}$; since $1 - f_n \geq 0$, we have $1 - f_n \in L^+ \subset L'$; hence

$$\mu^*(G^c) \leq \inf_n E(1 - f_n) = 1 - \lim_n E(f_n) = 1 - E(I_G).$$

Thus $\mu^*(G) + \mu^*(G^c) \leq 1$; since $\mu^*(G) + \mu^*(G^c)$ is always at least 1 by 1.3.3(b), we have $G \in \mathcal{H}$. [Under hypothesis B, the proof is the same, with $\{f_n\}$ a net instead of a sequence.] ■

We now prove the main theorem; for clarity we state the results under hypotheses A and B separately (in 4.2.9 and 4.2.10).

4.2.9 Daniell Representation Theorem. Let L be a vector space of real-valued functions on the set Ω ; assume that L contains the constant functions and is closed under the lattice operations. Let E be a *Daniell integral* on L , that is, a positive linear functional on L such that $E(f_n) \downarrow 0$ for each sequence of functions $f_n \in L$ with $f_n \downarrow 0$; assume that $E(1) = 1$.

Then there is a unique probability measure P on $\sigma(L)$ ($=\sigma(L') = \sigma(\mathcal{G})$ by 4.2.6) such that each $f \in L$ is P -integrable and $E(f) = \int_{\Omega} f dP$.

PROOF. Let P be the restriction of μ^* to $\sigma(L)$. By 4.2.6 and 4.2.8, P is a probability measure. If $G \in \mathcal{G}$, then

$$E(I_G) = \mu(G) = \mu^*(G) = P(G) = \int_{\Omega} I_G dP.$$

Now if $f \in L'$, we define

$$h_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\{(k-1)/2^n < f \leq k/2^n\}} + nI_{\{f > n\}}.$$

Since $I_{\{a < f \leq b\}} = I_{\{f > a\}} - I_{\{f > b\}}$ for $a < b$, and $\{f > a\}, \{f > b\} \in \mathcal{G}$ by 4.2.5, it follows that $E(h_n) = \int_{\Omega} h_n dP$. But the h_n form a sequence of nonnegative simple functions increasing to f [see 1.5.5(a)], so by the monotone convergence theorem, $E(f) = \int_{\Omega} f dP$.

Now let $f \in L$; $f = f^+ - f^-$, where $f^+, f^- \in L^+ \subset L'$. Then $E(f) = E(f^+) - E(f^-) = \int_{\Omega} f^+ dP - \int_{\Omega} f^- dP = \int_{\Omega} f dP$. (Since $f^+, f^- \in L$, the integrals are finite.)

This establishes the existence of the desired probability measure P . If P' is another such measure, then $\int_{\Omega} f dP = \int_{\Omega} f dP'$ for all $f \in L$, and hence for all $f \in L'$, by the monotone convergence theorem. Set $f = I_G$, $G \in \mathcal{G}$, to show that $P = P'$ on \mathcal{G} . Now \mathcal{G} is closed under finite intersection by 4.2.4(b); hence by 4.1.3, $P = P'$ on $\sigma(\mathcal{G})$, proving uniqueness. ■

4.2.10 Theorem. Let L be a vector space of real-valued functions on the set Ω ; assume that L contains the constant functions and is closed under the lattice operations. Let E be a positive linear functional on L such that $E(f_n) \downarrow 0$ for each net of functions $f_n \in L$ with $f_n \downarrow 0$; assume that $E(1) = 1$.

Then there is a unique probability measure P on $\sigma(L'')$ ($=\sigma(\mathcal{G})$ by 4.2.6) such that:

- (a) Each $f \in L$ is P -integrable and $E(f) = \int_{\Omega} f dP$.
- (b) If $\{G_n\}$ is a net of sets in \mathcal{G} and $G_n \uparrow G$, then $G \in \mathcal{G}$ and $P(G_n) \uparrow P(G)$.

PROOF. Let P be the restriction of μ^* to $\sigma(L'')$. The proof that P satisfies (a) is done exactly as in 4.2.9, with L' replaced by L'' ; P satisfies (b) by 4.2.4(d).

The uniqueness part cannot be done exactly as in 4.2.9 because the monotone convergence theorem fails in general for nets (see Problem 2). Let P' be a probability measure satisfying (a) and (b). If $f \in L''$, there is a net of functions $f_\alpha \in L^+$ with $f_\alpha \uparrow f$. We define

$$h_{n\alpha} = \frac{1}{2^n} \sum_{j=1}^{n2^n} I_{\{f_\alpha > j2^{-n}\}}, \quad n = 1, 2, \dots$$

If $(k-1)/2^n < f_\alpha(\omega) \leq k/2^n$, $k = 1, 2, \dots, n2^n$, then $h_{n\alpha}(\omega) = (k-1)/2^n$, and if $f_\alpha(\omega) > n$, then $h_{n\alpha}(\omega) = n$. Thus the $h_{n\alpha}$, $n = 1, 2, \dots$, are in fact the standard sequence of nonnegative simple functions increasing to f_α [see 1.5.5(a)]. Similarly, if

$$h_n = \frac{1}{2^n} \sum_{j=1}^{n2^n} I_{\{f > j2^{-n}\}},$$

the h_n are nonnegative simple functions increasing to f . Now

$$\begin{aligned} \int_{\Omega} h_n dP' &= \frac{1}{2^n} \sum_{j=1}^{n2^n} P'\{f > j2^{-n}\} \\ &= \frac{1}{2^n} \sum_{j=1}^{n2^n} \lim_{\alpha} P'\{f_\alpha > j2^{-n}\} \quad \text{by (b)} \\ &= \lim_{\alpha} \frac{1}{2^n} \sum_{j=1}^{n2^n} P'\{f_\alpha > j2^{-n}\} \\ &\quad \text{since the sum on } j \text{ is finite} \\ &= \lim_{\alpha} \int_{\Omega} h_{n\alpha} dP'; \end{aligned}$$

but

$$\begin{aligned} \int_{\Omega} f dP' &= \lim_n \int_{\Omega} h_n dP' \\ &\quad \text{by the monotone convergence theorem} \\ &= \lim_n \lim_{\alpha} \int_{\Omega} h_{n\alpha} dP' \\ &= \lim_{\alpha} \lim_n \int_{\Omega} h_{n\alpha} dP' \\ &\quad \text{since } h_{n\alpha} \text{ is monotone in each variable,} \\ &\quad \text{so that "lim" may be replaced by "sup"} \\ &= \lim_{\alpha} \int_{\Omega} f_{\alpha} dP' \\ &\quad \text{by the monotone convergence theorem} \end{aligned}$$

Equation continues

$$\begin{aligned}
 &= \lim_{\alpha} \int_{\Omega} f_{\alpha} dP \quad \text{by (a)} \\
 &= \int_{\Omega} f dP \\
 &\quad \text{by the above argument with } P' \text{ replaced by } P.
 \end{aligned}$$

Set $f = I_G$, $G \in \mathcal{G}$, to show that $P = P'$ on \mathcal{G} ; hence, as in 4.2.9, on $\sigma(\mathcal{G})$. ■

The following approximation theorem will be helpful in the next section.

4.2.11 Theorem. Assume the hypothesis of 4.2.9, and in addition assume that L is closed under limits of uniformly convergent sequences. Let

$$\mathcal{G}' = \{G \subset \Omega : G = \{f > 0\} \text{ for some } f \in L^+\}.$$

Then:

- (a) $\mathcal{G}' = \mathcal{G}$.
- (b) If $A \in \sigma(L)$, then $P(A) = \inf\{P(G) : G \in \mathcal{G}', G \supset A\}$.
- (c) If $G \in \mathcal{G}$, then $P(G) = \sup\{E(f) : f \in L^+, f \leq I_G\}$.

PROOF. (a) We have $\mathcal{G}' \subset \mathcal{G}$ by 4.2.5. Conversely, suppose $G \in \mathcal{G}$, and let $f_n \in L^+$ with $f_n \uparrow I_G$ ($\in L'$). Set $f = \sum_{n=1}^{\infty} 2^{-n} f_n$. Since $0 \leq f_n \leq 1$, the series is uniformly convergent, hence $f \in L^+$. But

$$\{f > 0\} = \bigcup_{n=1}^{\infty} \{f_n > 0\} = \{I_G = 1\} = G.$$

Consequently, $G \in \mathcal{G}'$.

- (b) This is immediate from (a) and the fact that $P = \mu^*$ on $\sigma(L)$.
- (c) If $f \in L^+, f \leq I_G$, then $E(f) \leq E(I_G) = P(G)$. Conversely, let $G \in \mathcal{G}$, with $f_n \in L^+, f_n \uparrow I_G$. Then $P(G) = E(I_G) = \lim_n E(f_n) = \sup_n E(f_n)$, hence $P(G) \leq \sup\{E(f) : f \in L^+, f \leq I_G\}$. ■

Problems

1. Give an example to show that the limit of a net of measurable functions need not be measurable.
2. Give an example of a net of nonnegative Borel measurable functions f_{α} increasing to a Borel measurable function f , with $\lim_{\alpha} \int f_{\alpha} d\mu \neq \int f d\mu$.
3. Let L be the class of real-valued continuous functions on $[0, 1]$, and let $E(f)$ be the Riemann integral of f . Show that E is a Daniell integral on L , and show that $\sigma(L) = \mathcal{B}[0, 1]$ and P is Lebesgue measure.

4.3 Measures on Topological Spaces

We are now in a position to obtain precise results on the interplay between measure theory and topology.

4.3.1 Definitions and Comments. Let Ω be a normal topological space (Ω is Hausdorff, and if A and B are disjoint closed subsets of Ω , there are disjoint open sets U and V with $A \subset U$ and $B \subset V$). The basic property of normal spaces that we need is *Urysohn's lemma*: If A and B are disjoint closed subsets of Ω , there is a continuous function $f: \Omega \rightarrow [0, 1]$ such that $f = 0$ on A and $f = 1$ on B . Other standard results are that every compact Hausdorff space is normal, and every metric space is normal.

The class of *Borel sets* of Ω , denoted by $\mathcal{B}(\Omega)$ or simply by \mathcal{B} , is defined as the smallest σ -field of subsets of Ω containing the open (or equally well the closed) sets. The class of *Baire sets* of Ω , denoted by $\mathcal{A}(\Omega)$ or simply by \mathcal{A} , is defined as the smallest σ -field of subsets of Ω making all continuous real-valued functions (Borel) measurable, that is, \mathcal{A} is the minimal σ -field containing all sets $f^{-1}(B)$ where B ranges over $\mathcal{B}(R)$ and f ranges over the class $C(\Omega)$ of continuous maps from Ω to R . Note that \mathcal{A} is the smallest σ -field making all bounded continuous functions measurable. For let \mathcal{F} be a σ -field that makes all bounded continuous functions measurable. If $f \in C(\Omega)$, then $f^+ \wedge n$ is a bounded continuous function and $f^+ \wedge n \uparrow f^+$ as $n \rightarrow \infty$. Thus f^+ (and similarly f^-) is \mathcal{F} -measurable, hence $f = f^+ - f^-$ is \mathcal{F} -measurable. Thus $\mathcal{A} \subset \mathcal{F}$, as desired.

The class of bounded continuous real-valued functions on Ω will be denoted by $C_b(\Omega)$.

If V is an open subset of R and $f \in C(\Omega)$, then $f^{-1}(V)$ is open in Ω , hence $f^{-1}(V) \in \mathcal{B}(\Omega)$. But the sets $f^{-1}(V)$ generate $\mathcal{A}(\Omega)$, since any σ -field containing the sets $f^{-1}(V)$ for all open sets V must contain the sets $f^{-1}(B)$ for all Borel sets B . (Problem 6 of Section 1.2 may be used to give a formal proof, with \mathcal{C} taken as the class of open sets.) It follows that $\mathcal{A}(\Omega) \subset \mathcal{B}(\Omega)$.

An F_σ set in Ω is a countable union of closed sets, and a G_σ set is a countable intersection of open sets.

4.3.2 Theorem. Let Ω be a normal topological space. Then $\mathcal{A}(\Omega)$ is the minimal σ -field containing the open F_σ sets (or equally well, the minimal σ -field containing the closed G_δ sets).

PROOF. Let $f \in C(\Omega)$; then $\{f > a\} = \bigcup_{n=1}^{\infty} \{f \geq a + (1/n)\}$ is an open F_σ set. As above, the sets $\{f > a\}$, $a \in R$, $f \in C(\Omega)$, generate \mathcal{A} ; hence \mathcal{A} is included

in the minimal σ -field \mathcal{H} over the F_σ sets. Conversely, let $H = \bigcup_{n=1}^{\infty} F_n$, F_n closed, be an open F_σ set. By Urysohn's lemma, there are functions $f_n \in C(\Omega)$ with $0 \leq f_n \leq 1$, $f_n = 0$ on H^c , $f_n = 1$ on F_n . If $f = \sum_{n=1}^{\infty} 2^{-n} f_n$, then $f \in C(\Omega)$, $0 \leq f \leq 1$, and $\{f > 0\} = \bigcup_{n=1}^{\infty} \{f_n > 0\} = H$. Thus $H \in \mathcal{A}$, so that $\mathcal{H} \subset \mathcal{A}$. ■

4.3.3 Corollary. If Ω is a normal topological space, the open F_σ sets are precisely the sets $\{f > 0\}$ where $f \in C_b(\Omega)$, $f \geq 0$.

PROOF. By the argument of 4.3.2. ■

4.3.4 Corollary. If Ω is a metric space, then $\mathcal{A}(\Omega) = \mathcal{B}(\Omega)$.

PROOF. If F is a closed subset of Ω , then F is a G_δ ,

$$\left(F = \bigcap_{n=1}^{\infty} \{ \omega : \text{dist}(\omega, F) < \frac{1}{n} \} \right);$$

hence every open subset of Ω is an F_σ . The result now follows from 4.3.2. ■

Corollary 4.3.4 has a direct proof that avoids use of 4.3.2; see Problem 1.

We may obtain some additional information about the open F_σ sets of a normal space.

4.3.5 Lemma. Let A be an open F_σ set in the normal space Ω , so that by 4.3.3, $A = \{f > 0\}$ where $f \in C_b(\Omega)$, $f \geq 0$. Then I_A is the limit of an increasing sequence of continuous functions.

PROOF. We have $\{f > 0\} = \bigcup_{n=1}^{\infty} \{f \geq 1/n\}$, and by Urysohn's lemma there are functions $f_n \in C(\Omega)$ with $0 \leq f_n \leq 1$, $f_n = 0$ on $\{f = 0\}$, $f_n = 1$ on $\{f \geq 1/n\}$. If $g_n = \max(f_1, \dots, f_n)$, then $g_n \uparrow I_{\{f > 0\}}$. ■

The Daniell theory now gives us a basic approximation theorem.

4.3.6 Theorem. Let P be any probability measure on $\mathcal{A}(\Omega)$, where Ω is a normal topological space. If $A \in \mathcal{A}$, then

- (a) $P(A) = \inf\{P(V) : V \supset A, V \text{ an open } F_\sigma \text{ set}\}$,
- (b) $P(A) = \sup\{P(C) : C \subset A, C \text{ a closed } G_\delta \text{ set}\}$.

PROOF. Let $L = C_b(\Omega)$ and define $E(f) = \int_{\Omega} f dP$, $f \in L$. [Note that $\sigma(L) = \mathcal{A}$, so each $f \in L$ is \mathcal{A} -measurable; furthermore, since f is bounded, $\int_{\Omega} f dP$ is finite. Thus E is well-defined.] Now E is a positive linear functional on L , and by the dominated convergence theorem, E is a Daniell integral. By 4.2.11(b),

$$P(A) = \inf\{P(G): G \in \mathcal{G}', G \supset A\},$$

where

$$\mathcal{G}' = \{G \subset \Omega: G = \{f > 0\} \text{ for some } f \in L^+\}.$$

By 4.3.3, \mathcal{G}' is the class of open F_σ sets, proving (a). Part (b) follows upon applying (a) to the complement of A . ■

4.3.7 Corollary. If Ω is a metric space, and P is a probability measure on $\mathcal{B}(\Omega)$, then for each $A \in \mathcal{B}(\Omega)$,

- (a) $P(A) = \inf\{P(V): V \supset A, V \text{ open}\},$
- (b) $P(A) = \sup\{P(C): C \subset A, C \text{ closed}\}.$

PROOF. In a metric space, every closed set is a G_δ and every open set is an F_σ (see 4.3.4); the result follows from 4.3.6. ■

Under additional hypotheses on Ω , we obtain approximations by compact subsets.

4.3.8 Theorem. Let Ω be a complete separable metric space (sometimes called a “Polish space”). If P is a probability measure on $\mathcal{B}(\Omega)$, then for each $A \in \mathcal{B}(\Omega)$,

$$P(A) = \sup\{P(K): K \text{ compact subset of } A\}.$$

PROOF. By 4.3.7, the approximation property holds with “compact” replaced by “closed.” We are going to show that if $\varepsilon > 0$, there is a compact set K_ε such that $P(K_\varepsilon) \geq 1 - \varepsilon$. This implies the theorem, for if C is closed, then $C \cap K_\varepsilon$ is compact, and $P(C) - P(C \cap K_\varepsilon) = P(C - K_\varepsilon) \leq P(\Omega - K_\varepsilon) \leq \varepsilon$.

Since Ω is separable, there is a countable dense set $\{\omega_1, \omega_2, \dots\}$. Let $B(\omega_n, r)$ (respectively, $\bar{B}(\omega_n, r)$) be the open (respectively, closed) ball with center at ω_n and radius r . Then for every $r > 0$, $\Omega = \bigcup_{n=1}^{\infty} \bar{B}(\omega_n, r)$ so that $\bigcup_{k=1}^m \bar{B}(\omega_k, 1/n) \uparrow \Omega$ as $m \rightarrow \infty$ (n fixed). Thus given $\varepsilon > 0$ and a positive integer n , there is a positive integer $m(n)$ such that $P\left(\bigcup_{k=1}^m \bar{B}(\omega_k, 1/n)\right) \geq 1 - \varepsilon 2^{-n}$ for all $m \geq m(n)$.

Let $K_\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{m(n)} \bar{B}(\omega_k, 1/n)$. Then K_ε is closed, and

$$P(K_\varepsilon^c) \leq \sum_{n=1}^{\infty} P\left(\bigcup_{k=1}^{m(n)} \bar{B}\left[\omega_k, \frac{1}{n}\right]\right)^c \leq \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon.$$

Therefore $P(K_\varepsilon) \geq 1 - \varepsilon$.

It remains to show that K_ε is compact. Let $\{x_1, x_2, \dots\}$ be a sequence in K_ε . Then $x_p \in \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{m(n)} \bar{B}(\omega_k, 1/n)$ for all p ; hence $x_p \in \bigcup_{k=1}^{m(1)} \bar{B}(\omega_k, 1)$ for all p . We conclude that for some k_1 , $x_p \in \bar{B}(\omega_{k_1}, 1)$ for infinitely many p , say, for $p \in T_1$, an infinite subset of the positive integers. But $x_p \in \bigcup_{k=1}^{m(2)} \bar{B}(\omega_k, \frac{1}{2})$ for all p , in particular for all $p \in T_1$; hence for some k_2 , $x_p \in \bar{B}(\omega_{k_1}, 1) \cap \bar{B}(\omega_{k_2}, \frac{1}{2})$ for infinitely many $p \in T_1$, say, for $p \in T_2 \subset T_1$. Continue inductively to obtain integers k_1, k_2, \dots and infinite sets $T_1 \supset T_2 \supset \dots$ such that

$$x_p \in \bigcap_{j=1}^i \bar{B}\left[\omega_{k_j}, \frac{1}{j}\right] \quad \text{for all } p \in T_i.$$

Pick $p_i \in T_i$, $i = 1, 2, \dots$, with $p_1 < p_2 < \dots$. Then if $j < i$, we have $x_{p_i}, x_{p_j} \in \bar{B}(\omega_{k_j}, 1/j)$, so $d(x_{p_i}, x_{p_j}) \leq 2/j \rightarrow 0$ as $j \rightarrow \infty$. Thus $\{x_{p_i}\}$ is a Cauchy sequence, hence converges (to a point of K_ε since K_ε is closed). Therefore $\{x_p\}$ has a subsequence converging to a point of K_ε , so K_ε is compact. ■

We now apply the Daniell theory to obtain theorems on representation of positive linear functionals in a topological context.

4.3.9 Theorem. Let Ω be a compact Hausdorff space, and let E be a positive linear functional on $C(\Omega)$, with $E(1) = 1$. There is a unique probability measure P on $\mathcal{A}(\Omega)$ such that $E(f) = \int_{\Omega} f dP$ for all $f \in C(\Omega)$.

PROOF. Let $L = C(\Omega)$. If $f_n \in L$, $f_n \downarrow 0$, then $f_n \rightarrow 0$ uniformly (this is *Dini's theorem*). For given $\delta > 0$, we have $\Omega = \bigcup_{n=1}^{\infty} \{f_n < \delta\}$; hence by compactness,

$$\begin{aligned} \Omega &= \bigcup_{n=1}^N \{f_n < \delta\} \quad \text{for some } N \\ &= \{f_N < \delta\} \quad \text{by monotonicity of } \{f_n\}. \end{aligned}$$

Thus $n \geq N$ implies $0 \leq f_n(\omega) \leq f_N(\omega) < \delta$ for all ω , proving uniform convergence.

Thus if $\delta > 0$ is given, eventually $0 \leq f_n \leq \delta$, so $0 \leq E(f_n) \leq E(\delta) = \delta$. Therefore $E(f_n) \downarrow 0$, hence E is a Daniell integral. The result follows from 4.2.9. ■

A somewhat different result is obtained if we use the Daniell theory with hypothesis B.

4.3.10 Theorem. Let Ω be a compact Hausdorff space, and let E be a positive linear functional on $C(\Omega)$, with $E(1) = 1$. There is a unique probability measure P on $\mathcal{B}(\Omega)$ such that

- (a) $E(f) = \int_{\Omega} f dP$ for all $f \in C(\Omega)$, and
- (b) for all $A \in \mathcal{B}(\Omega)$,

$$P(A) = \inf\{P(V) : V \supset A, \quad V \text{ open}\}$$

or equivalently,

$$P(A) = \sup\{P(K) : K \subset A, \quad K \text{ compact}\}.$$

(“Compact” may be replaced by “closed” since Ω is compact Hausdorff.)

PROOF. Let $L = C(\Omega)$. If $\{f_n, n \in D\}$ is a net in L and $f_n \downarrow 0$, then, as in 4.3.9, for any $\delta > 0$ we have $\Omega = \bigcup_{j \in F} \{f_j < \delta\}$ for some finite set $F \subset D$. If $N \in D$ and $N \geq j$ for all $j \in F$, then $\Omega = \{f_N < \delta\}$ by monotonicity of the net. Thus $n \geq N$ implies $0 \leq f_n \leq f_N < \delta$, and it follows as in 4.3.9 that $E(f_n) \downarrow 0$.

By 4.2.10 there is a probability measure P on $\sigma(L') = \sigma(\mathcal{G})$ such that $E(f) = \int_{\Omega} f dP$ for all $f \in L$. But in fact \mathcal{G} is the class of open sets, so that $\sigma(\mathcal{G}) = \mathcal{B}(\Omega)$. For if $f \in L'$, there is a net of nonnegative continuous functions $f_n \uparrow f$; hence for each real a , $\{f > a\} = \bigcup_n \{f_n > a\}$ is an open set. Thus if $G \in \mathcal{G}$, then $I_G \in L'$, so that $G = \{I_G > 0\}$ is open. Conversely if G is open and $\omega \in G$, there is a continuous function $f_{\omega} : \Omega \rightarrow [0, 1]$ such that $f_{\omega}(\omega) = 1$ and $f_{\omega} = 0$ on G^c . Thus $I_G = \sup_{\omega \in G} f_{\omega}$, so that if for each finite set $F \subset G$ we define $g_F = \max\{f_{\omega} : \omega \in F\}$, and direct the sets F by inclusion, we obtain a monotone net of nonnegative continuous functions increasing to I_G . Therefore $I_G \in L'$, so that $G \in \mathcal{G}$.

Thus we have established the existence of a probability measure P on $\mathcal{B}(\Omega)$ satisfying part (a) of 4.3.10; part (b) follows since $P = \mu^*$ on $\sigma(\mathcal{G})$, and \mathcal{G} is the class of open sets.

To prove uniqueness, let P' be another probability measure satisfying (a) and (b) of 4.3.10. If we can show that P' satisfies 4.2.10(b), it will follow from the uniqueness part of 4.2.10 that $P' = P$. Thus let $\{G_n\}$ be a net of open sets with $G_n \uparrow G$; since G is the union of the G_n , G is open. By hypothesis, given $\delta > 0$, there is a compact $K \subset G$ such that $P'(G) \leq P'(K) + \delta$. Now $G_n \cup K^c \uparrow G \cup K^c = \Omega$; hence by compactness and the monotonicity of $\{G_n\}$, $G_m \cup K^c = \Omega$ for some m , so that $K \subset G_m$. Consequently,

$$P'(G) \leq P'(K) + \delta \leq P'(G_m) + \delta \leq \lim_n P'(G_n) + \delta$$

and it follows that $P'(G_n) \uparrow P'(G)$. ■

Property (b) of 4.3.10 is often referred to as the “regularity” of P . Since the word “regular” is used in so many different ways in the literature, let us state exactly what it will mean for us.

4.3.11 Definitions. If μ is a measure on $\mathcal{B}(\Omega)$, where Ω is a normal topological space, μ is said to be *regular* iff for each $A \in \mathcal{B}(\Omega)$,

$$\mu(A) = \inf\{\mu(V): V \supset A, \quad V \text{ open}\}$$

and

$$\mu(A) = \sup\{\mu(C): C \subset A, \quad C \text{ closed}\}.$$

Either one of these conditions implies the other if μ is finite, and if in addition, Ω is a compact Hausdorff space, we obtain property (b) of 4.3.10.

If $\mu = \mu^+ - \mu^-$ is a finite signed measure on $\mathcal{B}(\Omega)$, Ω normal, we say that μ is regular iff μ^+ and μ^- are regular (equivalently, iff the total variation $|\mu|$ is regular).

The following result connects 4.3.9 and 4.3.10.

4.3.12 Theorem. If P is a probability measure on $\mathcal{A}(\Omega)$, Ω compact Hausdorff, then P has a unique extension to a regular probability measure on $\mathcal{B}(\Omega)$.

PROOF. Let $E(f) = \int_{\Omega} f dP$, $f \in C(\Omega)$. Then E is a positive linear functional on $L = C(\Omega)$, and thus (see the proof of 4.3.10) if $\{f_n\}$ is a net in L decreasing to 0, then $E(f_n) \downarrow 0$. By 4.3.10 there is a unique regular probability measure P' on $\mathcal{B}(\Omega)$ such that $\int_{\Omega} f dP = \int_{\Omega} f dP'$ for all $f \in L$. But each f in L is measurable: $(\Omega, \mathcal{A}) \rightarrow (R, \mathcal{B}(R))$, hence by 1.5.5, $\int_{\Omega} f dP'$ is determined by the values of P' on Baire sets. Thus the condition that $\int_{\Omega} f dP = \int_{\Omega} f dP'$ for all $f \in L$ is equivalent to $P = P'$ on $\mathcal{A}(\Omega)$, by the uniqueness part of 4.3.9. ■

In 4.3.9 and 4.3.10, the assumption $E(1) = 1$ is just a normalization, and if it is dropped, the results are the same, except that “unique probability measure” is replaced by “unique finite measure.” Similarly, 4.3.12 applies equally well to finite measures.

Now let Ω be a compact Hausdorff space, and consider $L = C(\Omega)$ as a vector space over the reals; L is a Banach space with the sup norm. If E is a positive linear functional on L , we can show that E is continuous, and this will allow us to generalize 4.3.9 and 4.3.10 by giving representation theorems for continuous linear functionals on $C(\Omega)$. To prove continuity of E , note that if $f \in L$ and $\|f\| \leq 1$, then $-1 \leq f(\omega) \leq 1$ for all ω ; hence $-E(1) \leq E(f) \leq E(1)$.

$E(1)$, that is, $|E(f)| \leq E(1)$. Therefore $\|E\| \leq E(1)$; in fact $\|E\| = E(1)$, as may be seen by considering the function that is identically 1.

The representation theorem we are about to prove will involve integration with respect to a finite signed measure $\mu = \mu^+ - \mu^-$; the integral is defined in the obvious way, namely,

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f \, d\mu^+ - \int_{\Omega} f \, d\mu^-,$$

assuming the right side is well-defined.

4.3.13 Theorem. Let E be a continuous linear functional on $C(\Omega)$, Ω compact Hausdorff.

- (a) There is a unique finite signed measure μ on $\mathcal{A}(\Omega)$ such that $E(f) = \int_{\Omega} f \, d\mu$ for all $f \in C(\Omega)$.
- (b) There is a unique regular finite signed measure λ on $\mathcal{B}(\Omega)$ such that $E(f) = \int_{\Omega} f \, d\lambda$ for all $f \in C(\Omega)$.

Furthermore, any finite signed measure on $\mathcal{A}(\Omega)$ has a unique extension to a regular finite signed measure on $\mathcal{B}(\Omega)$; in particular λ is the unique extension of μ .

PROOF. The existence of the desired signed measures μ and λ will follow from 4.3.9 and 4.3.10 if we show that E is the difference of two positive linear functionals E^+ and E^- . If $f \geq 0, f \in C(\Omega)$, define

$$E^+(f) = \sup\{E(g) : 0 \leq g \leq f, g \in C(\Omega)\}.$$

If $f \leq M$, then $E^+(f) \leq M\|E\| < \infty$ by continuity of E ; hence E^+ is finite. If $0 \leq g_i \leq f_i, i = 1, 2$, with all functions in $C(\Omega)$, then

$$E(g_1) + E(g_2) = E(g_1 + g_2) \leq E^+(f_1 + f_2).$$

Take the sup over g_1 and g_2 to obtain $E^+(f_1) + E^+(f_2) \leq E^+(f_1 + f_2)$.

Now if $0 \leq g \leq f_1 + f_2, f_1, f_2 \geq 0$, define $g_1 = g \wedge f_1, g_2 = g \vee (g - f_1)$. Then $0 \leq g_1 \leq f_1, 0 \leq g_2 \leq f_2$, and $g = g_1 + g_2$. Thus $E(g) = E(g_1) + E(g_2) \leq E^+(f_1) + E^+(f_2)$; hence $E^+(f_1 + f_2) \leq E^+(f_1) + E^+(f_2)$, and consequently

$$E^+(f_1 + f_2) = E^+(f_1) + E^+(f_2). \tag{1}$$

Clearly $E^+(af) = aE^+(f)$ if $a \geq 0$. Thus E^+ extends to a positive linear functional on $C(\Omega)$. Specifically, if $f = f^+ - f^-$, take $E^+(f) = E^+(f^+) - E^+(f^-)$. If f can also be represented as $g - h$, where g and h are nonnegative

functions in $C(\Omega)$, then $E^+(f^+) - E^+(f^-) = E^+(g) - E^+(h)$ by (1), so the extension is well-defined. If $f \geq 0$, $f \in C(\Omega)$, define

$$E^-(f) = -\inf\{E(g) : 0 \leq g \leq f, g \in C(\Omega)\} = (-E)^+(f).$$

By the above argument, E^- extends to a positive linear functional on $C(\Omega)$. Now

$$\begin{aligned} E^-(f) &= -\inf\{E(f) - E(f-g) : 0 \leq g \leq f, g \in C(\Omega)\} \\ &= -E(f) + \sup\{E(h) : 0 \leq h \leq f, h \in C(\Omega)\} \\ &= -E(f) + E^+(f). \end{aligned}$$

Thus $E = E^+ - E^-$.

To prove that μ is unique, assume that τ is a finite signed measure on $\mathcal{A}(\Omega)$ such that $\int_{\Omega} f d\tau = 0$ for all $f \in C(\Omega)$. Then $\int_{\Omega} f d\tau^+ = \int_{\Omega} f d\tau^-$ for all $f \in C(\Omega)$, hence $\tau^+ = \tau^-$ (so $\tau = 0$) by 4.3.9. Uniqueness of λ is proved similarly, using 4.3.10.

Now if μ is any finite signed measure on $\mathcal{A}(\Omega)$, $E(f) = \int_{\Omega} f d\mu$ defines a continuous linear functional on $C(\Omega)$, so by what we have just proved there is a unique regular finite signed measure λ on $\mathcal{B}(\Omega)$ such that $\int_{\Omega} f d\mu = \int_{\Omega} f d\lambda$ for all $f \in C(\Omega)$. But as in 4.3.12, this condition is equivalent to $\mu = \lambda$ on $\mathcal{A}(\Omega)$, so μ has a unique extension to a regular finite signed measure λ on $\mathcal{B}(\Omega)$. ■

Theorems 4.3.9, 4.3.10, and 4.3.13 are referred to as versions of the *Riesz representation theorem*. [This name is also given to the somewhat different result 3.3.4(a).] A result of this type for complex-valued functions and complex measures is given in Problem 6.

We are going to show that $\|E\| = |\mu|(\Omega) = |\lambda|(\Omega)$ in 4.3.13. To do this we need a result on the approximation of Borel measurable functions by continuous functions.

4.3.14 Theorem. Consider the measure space $(\Omega, \mathcal{F}, \mu)$, where Ω is a normal topological space, $\mathcal{F} = \mathcal{B}(\Omega)$, and μ is a regular measure on \mathcal{F} . If $0 < p < \infty$, $\varepsilon > 0$, and $f \in L^p(\Omega, \mathcal{F}, \mu)$, there is a continuous complex-valued function $g \in L^p(\Omega, \mathcal{F}, \mu)$ such that $\|f - g\|_p < \varepsilon$; furthermore, g can be chosen so that $\sup|g| \leq \sup|f|$. Thus the continuous functions are dense in L^p .

PROOF. This is done exactly as in 2.4.14, except that the application of Problem 12, Section 1.4, is now replaced by the hypothesis that μ is regular. ■

4.3.15 Theorem. In the Riesz representation theorem 4.3.13, $\|E\| = |\mu|(\Omega) = |\lambda|(\Omega)$.

PROOF. If $f \in C(\Omega)$, $E(f) = \int_{\Omega} f d\lambda = \int_{\Omega} f d\lambda^+ - \int_{\Omega} f d\lambda^-$; hence

$$|E(f)| \leq \int_{\Omega} |f| d\lambda^+ + \int_{\Omega} |f| d\lambda^- = \int_{\Omega} |f| d|\lambda| \leq \|f\| |\lambda|(\Omega),$$

where $\| \cdot \|$ is the sup norm. Thus $\|E\| \leq |\lambda|(\Omega)$.

Now let A_1, \dots, A_n be disjoint measurable subsets of Ω , and define $f = \sum_{j=1}^n x_j I_{A_j}$, where $x_j = 1$ if $\lambda(A_j) \geq 0$, and $x_j = -1$ if $\lambda(A_j) < 0$. By 4.3.14, if $\varepsilon > 0$, there is a continuous function g such that $\|g\| \leq 1$ and $\int_{\Omega} |f - g| d|\lambda| < \varepsilon$; hence $E(g) = \int_{\Omega} g d\lambda \geq \int_{\Omega} f d\lambda - \varepsilon$. But

$$\int_{\Omega} f d\lambda = \sum_{j=1}^n |\lambda(A_j)|;$$

hence

$$\|E\| \geq E(g) \geq \sum_{j=1}^n |\lambda(A_j)| - \varepsilon.$$

But $\sum_{j=1}^n |\lambda(A_j)|$ may be taken arbitrarily close to $|\lambda|(\Omega)$ (see Problem 4, Section 2.1); hence $\|E\| \geq |\lambda|(\Omega)$. Since $|\lambda|(\Omega) = |\mu|(\Omega)$, the result follows. ■

If Ω is compact Hausdorff and $M(\Omega)$ is the collection of finite signed measures on $\mathcal{A}(\Omega)$ (or equally well the regular finite signed measures on $\mathcal{B}(\Omega)$), Theorems 4.3.13 and 4.3.15 show that the map $E \rightarrow \mu$ (or $E \rightarrow \lambda$) is an isometric isomorphism of the conjugate space of $C(\Omega)$ and $M(\Omega)$, where the norm of an element $\mu \in M(\Omega)$ is taken as $|\mu|(\Omega)$.

We close this section with some further results on approximation by continuous functions.

4.3.16 Theorem. Let μ be a regular finite measure on $\mathcal{B}(\Omega)$, Ω normal. If f is a complex-valued Borel measurable function on Ω and $\delta > 0$, there is a continuous complex-valued function g on Ω such that

$$\mu\{\omega : f(\omega) \neq g(\omega)\} < \delta.$$

Furthermore, it is possible to choose g so that $\sup|g| \leq \sup|f|$.

PROOF. First assume f is real-valued and $0 \leq f < 1$. If $h_n(\omega) = (k-1)2^{-n}$ when $(k-1)2^{-n} \leq f(\omega) < k2^{-n}$, $k = 1, 2, \dots, n2^n$, $h_n(\omega) = n$ when $f(\omega) \geq n$, the h_n are nonnegative simple functions increasing to f . Let $f_n = h_n - h_{n-1}$,

$n = 1, 2, \dots$ (with $h_0 = 0$), so that $f = \sum_{n=1}^{\infty} f_n$. Note that f_n has only two possible values, 0 and 2^{-n} . If $A_n = \{f_n \neq 0\}$, let C_n be a closed subset of A_n and V_n an open overset of A_n such that $\mu(V_n - C_n) < \delta 2^{-n}$. Since Ω is normal, there is a continuous $g_n: \Omega \rightarrow [0, 1]$ such that $g_n = 1$ on C_n and $g_n = 0$ off V_n . If $g = \sum_{n=1}^{\infty} 2^{-n} g_n$, then by the Weierstrass M-test, g is a continuous map of Ω into $[0, 1]$. We claim that if $\omega \notin \bigcup_{n=1}^{\infty} (V_n - C_n)$, a set of measure less than δ , then $f(\omega) = g(\omega)$. To see this, observe that for each n , $\omega \in C_n$ or $\omega \notin V_n$. If $\omega \in C_n \subset A_n$, then $2^{-n} g_n(\omega) = 2^{-n} = f_n(\omega)$, and if $\omega \notin V_n$, then $2^{-n} g_n(\omega) = 0 = f_n(\omega)$ since $\omega \notin A_n$.

This proves the existence of g when $0 \leq f < 1$; the extension to a complex-valued bounded f is immediate. If f is unbounded, write $f = f I_{\{|f| < n\}} + f I_{\{|f| \geq n\}} = f_1 + f_2$, where f_1 is bounded and $\mu\{f_2 \neq 0\} = \mu\{|f| \geq n\}$, which can be made less than $\delta/2$ for sufficiently large n . If g is continuous and $\mu\{f_1 \neq g\} < \delta/2$, then $\mu\{f \neq g\} < \delta$, as desired.

Finally, if $|f| \leq M < \infty$, and g approximates f as above, define $g_1(\omega) = g(\omega)$ if $|g(\omega)| \leq M$, $g_1(\omega) = Mg(\omega)/|g(\omega)|$ if $|g(\omega)| > M$. Then g_1 is continuous, $|g_1| \leq M$, and $f(\omega) = g(\omega)$ implies $|g(\omega)| \leq M$; hence $g_1(\omega) = g(\omega) = f(\omega)$. Therefore $\mu\{f \neq g_1\} \leq \mu\{f \neq g\} < \delta$, completing the proof. ■

4.3.17 Corollaries.

Assume the hypothesis of 4.3.16.

(a) There is a sequence of continuous complex-valued functions f_n on Ω converging to f a.e. [μ], with $|f_n| \leq \sup |f|$ for all n .

(b) Given $\varepsilon > 0$, there is a closed set $C \subset \Omega$ and a continuous complex-valued function g on Ω such that $\mu(C) \geq \mu(\Omega) - \varepsilon$ and $f = g$ on C , hence the restriction of f to C is continuous. If μ has the additional property that $\mu(A) = \sup\{\mu(K): K \subset A, K \text{ compact}\}$ for each $A \in \mathcal{B}(\Omega)$, then C may be taken as compact.

PROOF. (a) By 4.3.16, there is a continuous function f_n such that $|f_n| \leq M = \sup |f|$ and $\mu\{f_n \neq f\} < 2^{-n}$. If $A_n = \{f_n \neq f\}$ and $A = \lim \sup_n A_n$, then $\mu(A) = 0$ by the Borel-Cantelli lemma. But if $\omega \notin A$, then $f_n(\omega) = f(\omega)$ for sufficiently large n .

(b) By 4.3.16, there is a continuous g such that $\mu\{f \neq g\} < \varepsilon/2$. By regularity of μ , there is a closed set $C \subset \{f = g\}$ with $\mu(C) \geq \mu\{f = g\} - \varepsilon/2$. The set C has the desired properties. The proof under the assumption of approximation by compact subsets is the same, with C compact rather than closed. ■

Corollary 4.3.17(b) is called *Lusin's theorem*.

Problems

- Let F be a closed subset of the metric space Ω . Define $f_n(\omega) = e^{-nd(\omega, F)}$ where $d(\omega, F) = \inf\{d(\omega, y) : y \in F\}$. Show that the f_n are continuous and $f_n \downarrow I_F$. Use this to give a direct proof (avoiding 4.3.2) that in a metric space, the Baire and Borel sets coincide.
- Give an example of a measure space $(\Omega, \mathcal{F}, \mu)$, where Ω is a metric space and $\mathcal{F} = \mathcal{B}(\Omega)$, such that for some $A \in \mathcal{F}$, $\mu(A) \neq \sup\{\mu(K) : K \subset A, K \text{ compact}\}$.
- In 4.3.14, assume in addition that Ω is locally compact, and that $\mu(A) = \sup\{\mu(K) : K \text{ compact subset of } A\}$ for all Borel sets A . Show that the continuous functions with compact support (that is, the continuous functions that vanish outside a compact subset) are dense in $L^p(\Omega, \mathcal{F}, \mu)$, $0 < p < \infty$. Also, as in 4.3.14, if $f \in L^p$ is approximated by the continuous function g with compact support, g may be chosen so that $\sup|g| \leq \sup|f|$.
- (a) Let Ω be a normal topological space, and let H be the smallest class of real-valued functions on Ω that contains the continuous functions and is closed under pointwise limits of monotone sequences. Show that H is the class of Baire measurable functions, that is, H consists of all $f: (\Omega, \mathcal{A}) \rightarrow (R, \mathcal{B}(R))$ (use 4.1.4).
(b) If H is as in part (a) and $\sigma(H)$ is the smallest σ -field \mathcal{G} of subsets of Ω making all functions in H measurable (relative to \mathcal{G} and $\mathcal{B}(R)$), show that $\sigma(H) = \mathcal{A}(\Omega)$; hence $\sigma(H)$ is the same as $\sigma(C(\Omega))$.
- Let Ω be a normal topological space, and let K_0 be the class of all continuous real-valued functions on Ω . Having defined K_β for all ordinals β less than the ordinal α , define

$$K_\alpha = \bigcup\{K_\beta : \beta < \alpha\}',$$

where, for any class D of real-valued functions, D' means the class of all real-valued functions that are pointwise limits of monotone sequences in D .

Let

$$K = \bigcup\{K_\alpha : \alpha < \beta_1\},$$

where β_1 is the first uncountable ordinal. Show that K is the class of Baire measurable functions.

- Let E be a continuous linear functional on the space $C(\Omega)$ of all continuous complex-valued functions on the compact Hausdorff space Ω . Show that there is a unique complex measure μ on $\mathcal{A}(\Omega)$ such that $E(f) = \int_{\Omega} f \, d\mu$ for all $f \in C(\Omega)$, and a unique regular complex measure λ on $\mathcal{B}(\Omega)$ such that $E(f) = \int_{\Omega} f \, d\lambda$ for all $f \in C(\Omega)$. Furthermore, $\|E\| =$

$|\mu|(\Omega) = |\lambda|(\Omega)$, so that $C(\Omega)^*$ is isometrically isomorphic to the space of complex measures on $\mathcal{A}(\Omega)$, or equally well to the space of regular complex measures on $\mathcal{B}(\Omega)$. (If $\mu = \mu_1 + i\mu_2$ is a complex measure, $\int_{\Omega} f d\mu$ is defined as $\int_{\Omega} f d\mu_1 + i \int_{\Omega} f d\mu_2$, provided f is integrable with respect to μ_1 and μ_2 . Also, μ is said to be regular iff μ_1 and μ_2 are regular; since $|\mu_1|, |\mu_2| \leq |\mu| \leq |\mu_1| + |\mu_2|$, this is equivalent to regularity of $|\mu|$. See Section 2.2, Problem 6, and Section 2.4, Problems 10 and 11 for the basic properties of complex measures.)

7. Let $M(\Omega, \mathcal{F})$ be the collection of finite signed measures on a σ -field \mathcal{F} of subsets of Ω . Show that if we take $\|\mu\| = |\mu|(\Omega)$, $\mu \in M(\Omega, \mathcal{F})$, $M(\Omega, \mathcal{F})$ becomes a Banach space. (A similar result holds for the collection of complex measures.)

4.4 Measures on Uncountably Infinite Product Spaces

In Section 2.7 we considered probability measures on countably infinite product spaces. The result may be extended to uncountable products if certain topological assumptions are made about the individual factor spaces.

The product of uncountably many σ -fields is formed in essentially the same way as in the countable case.

4.4.1 Definitions and Comments. For t in the arbitrary index set T , let $(\Omega_t, \mathcal{F}_t)$ be a measurable space. Let $\prod_{t \in T} \Omega_t$ be the set of all functions $\omega = (\omega(t), t \in T)$ on T such that $\omega(t) \in \Omega_t$ for each $t \in T$. If $t_1, \dots, t_n \in T$ and $B^n \subset \prod_{i=1}^n \Omega_{t_i}$, we define the set $B^n(t_1, \dots, t_n)$ as $\{\omega \in \prod_{t \in T} \Omega_t: (\omega(t_1), \dots, \omega(t_n)) \in B^n\}$. We call $B^n(t_1, \dots, t_n)$ the *cylinder* with *base* B^n at (t_1, \dots, t_n) ; the cylinder is said to be *measurable* iff $B^n \in \prod_{i=1}^n \mathcal{F}_{t_i}$. If $B^n = B_1 \times \dots \times B_n$, the cylinder is called a *rectangle*, a *measurable rectangle* iff $B_i \in \mathcal{F}_{t_i}$, $i = 1, \dots, n$. Note that if all $\Omega_t = \Omega$, then $\prod_{t \in T} \Omega_t = \Omega^T$, the set of all functions from T to Ω .

For example, let $T = [0, 1]$, $\Omega_t = \mathbb{R}$ for all $t \in T$, $B^2 = \{(u, v): u > 3, 1 < v < 2\}$. Then

$$B^2(\frac{1}{2}, \frac{3}{4}) = \{x \in \mathbb{R}^T: x(\frac{1}{2}) > 3, 1 < x(\frac{3}{4}) < 2\}$$

[see Fig. 4.1, where $x_1 \in B^2(\frac{1}{2}, \frac{3}{4})$ and $x_2 \notin B^2(\frac{1}{2}, \frac{3}{4})$].

Exactly as in 2.7.1, the measurable cylinders form a field, as do the finite disjoint unions of measurable rectangles. The minimal σ -field over the measurable cylinders is denoted by $\prod_{t \in T} \mathcal{F}_t$, and called the *product* of the σ -fields \mathcal{F}_t . If $\Omega_t = S$ and $\mathcal{F}_t = \mathcal{S}$ for all t , $\prod_{t \in T} \mathcal{F}_t$ is denoted by \mathcal{S}^T . Again as in 2.7.1, $\prod_{t \in T} \mathcal{F}_t$ is also the minimal σ -field over the measurable rectangles.

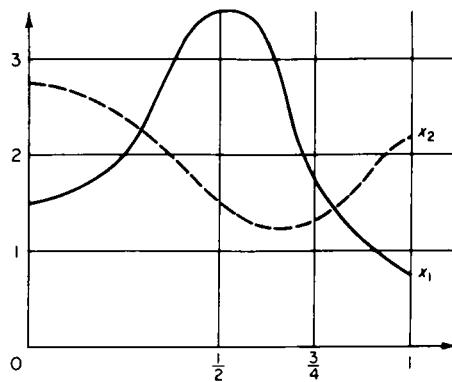


Figure 4.1.

We now consider the problem of constructing probability measures on $\prod_{t \in T} \mathcal{F}_t$. The approach will be as follows: Let $v = \{t_1, \dots, t_n\}$ be a finite subset of T , where $t_1 < t_2 < \dots < t_n$. (If T is not a subset of \mathbb{R} , some fixed total ordering is put on T .) Assume that for each such v we are given a probability measure P_v on $\prod_{i=1}^n \mathcal{F}_{t_i}$; $P_v(B)$ is to represent $P\{\omega \in \prod_{t \in T} \Omega_t : (\omega(t_1), \dots, \omega(t_n)) \in B\}$. We shall require that the P_v be "consistent"; to see what kind of consistency is needed, consider an example.

Suppose T is the set of positive integers and $\Omega_t = \mathbb{R}$, $\mathcal{F}_t = \mathcal{B}(\mathbb{R})$ for all t . Suppose we know $P_{12345}(B^5) = P\{\omega : (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) \in B^5\}$ for all $B^5 \in \mathcal{B}(\mathbb{R}^5)$. Then $P\{\omega : (\omega_2, \omega_3) \in B^2\} = P\{\omega : (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) \in R \times B^2 \times R^2\} = P_{12345}(R \times B^2 \times R^2)$, $B^2 \in \mathcal{B}(\mathbb{R}^2)$. Thus once probabilities of sets involving the first five coordinates are specified, probabilities of sets involving (ω_2, ω_3) [as well as $(\omega_1, \omega_3, \omega_4)$, and so on], are determined. Thus the original specification of P_{23} must agree with the measure induced from P_{12345} . We are going to show that under appropriate topological assumptions, a consistent family of probability measures P_v determines a unique probability measure on $\prod_{t \in T} \mathcal{F}_t$.

Now to formalize: If $v = \{t_1, \dots, t_n\}$, $t_1 < \dots < t_n$, the space $(\prod_{i=1}^n \Omega_{t_i}, \prod_{i=1}^n \mathcal{F}_{t_i})$ is denoted by $(\Omega_v, \mathcal{F}_v)$. If $u = \{t_{i1}, \dots, t_{ik}\}$ is a nonempty subset of v and $y = (y(t_1), \dots, y(t_n)) \in \Omega_v$, the k -tuple $(y(t_{i1}), \dots, y(t_{ik}))$ is denoted by y_u . Similarly if $\omega = (\omega(t), t \in T)$ belongs to $\prod_{t \in T} \Omega_t$, the notation ω_v will be used for $(\omega(t_1), \dots, \omega(t_n))$.

If P_v is a probability measure on \mathcal{F}_v , the projection of P_v on \mathcal{F}_u is the probability measure $\pi_u(P_v)$ on \mathcal{F}_u defined by

$$[\pi_u(P_v)](B) = P_v\{\omega \in \Omega_v : y_u \in B\}, \quad B \in \mathcal{F}_u.$$

Similarly, if Q is a probability measure on $\prod_{t \in T} \mathcal{F}_t$, the projection of Q on \mathcal{F}_v is defined by

$$[\pi_v(Q)](B) = Q\left\{\omega \in \prod_{t \in T} \Omega_t : \omega_v \in B\right\} = Q(B(v)), \quad B \in \mathcal{F}_v.$$

We need one preliminary result.

4.4.2 Theorem. For each $n = 1, 2, \dots$, suppose that \mathcal{F}_n is the class of Borel sets of a separable metric space Ω_n . Let $\Omega = \prod_n \Omega_n$, with the product topology, and let $\mathcal{F} = \mathcal{B}(\Omega)$. [Note that Ω is metrizable, so that the Baire and Borel sets of Ω coincide; we may take

$$d(x, y) = \sum_n \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)},$$

where d_n is the metric of Ω_n . Also, if each Ω_n is complete, so is Ω .]

Then \mathcal{F} is the product σ -field $\prod_n \mathcal{F}_n$.

PROOF. The sets $\{\omega \in \Omega : \omega_1 \in A_1, \dots, \omega_n \in A_n\}$, $n = 1, 2, \dots$, where the A_i range over the countable base for Ω_i (recall that separability and second countability are equivalent in metric spaces), form a countable base for Ω . Since the sets are measurable rectangles, it follows that every open subset of Ω belongs to $\prod_n \mathcal{F}_n$; hence $\mathcal{F} \subset \prod_n \mathcal{F}_n$. On the other hand, for a fixed positive integer i let $\mathcal{C} = \{B \in \mathcal{B}(\Omega_i) : \{\omega \in \Omega : \omega_i \in B\} \in \mathcal{F}\}$. Then \mathcal{C} is a σ -field containing the open sets of Ω_i , hence $\mathcal{C} = \mathcal{B}(\Omega_i)$. In other words, every measurable rectangle with one-dimensional base belongs to \mathcal{F} . Since an arbitrary measurable rectangle is a finite intersection of such sets, it follows that $\prod_n \mathcal{F}_n \subset \mathcal{F}$. ■

We are now ready for the main result.

4.4.3 Kolmogorov Extension Theorem. For each t in the arbitrary index set T , let Ω_t be a complete, separable metric space, and \mathcal{F}_t the Borel sets of Ω_t . Assume that for each finite nonempty subset v of T , we are given a probability measure P_v on \mathcal{F}_v . Assume the P_v are consistent, that is, $\pi_u(P_v) = P_u$ for each nonempty $u \subset v$.

Then there is a unique probability measure P on $\mathcal{F} = \prod_{t \in T} \mathcal{F}_t$, such that $\pi_v(P) = P_v$ for all v .

PROOF. We define the hoped-for measure on measurable cylinders by $P(B^n(v)) = P_v(B^n)$, $B^n \in \mathcal{F}_v$.

We must show that this definition makes sense since a given measurable cylinder can be represented in several ways. For example, if all $\Omega_t = R$ and $B^2 = (-\infty, 3) \times (4, 5)$, then

$$\begin{aligned} B^2(t_1, t_2) &= \{\omega : \omega(t_1) < 3, \quad 4 < \omega(t_2) < 5\} \\ &= \{\omega : \omega(t_1) < 3, \quad 4 < \omega(t_2) < 5, \quad \omega(t_3) \in R\} \\ &= B^3(t_1, t_2, t_3) \quad \text{where } B^3 = (-\infty, 3) \times (4, 5) \times R. \end{aligned}$$

It is sufficient to consider dual representation of the same measurable cylinder in the form $B^n(v) = B^k(u)$ where $k < n$ and $u \subset v$. But then

$$\begin{aligned} P_u(B^k) &= [\pi_u(P_v)](B^k) \quad \text{by the consistency hypothesis} \\ &= P_v\{y \in \Omega_v : y_u \in B^k\} \quad \text{by definition of projection.} \end{aligned}$$

But the assumption $B^n(v) = B^k(u)$ implies that if $y \in \Omega_v$, then $y \in B^n$ iff $y_u \in B^k$, hence $P_u(B^k) = P_v(B^n)$, as desired.

Thus, P is well-defined on measurable cylinders; the class \mathcal{F}_0 of measurable cylinders forms a field, and $\sigma(\mathcal{F}_0) = \mathcal{F}$.

Now if A_1, \dots, A_m are disjoint sets in \mathcal{F}_0 , we may write (by introducing extra factors as in the above example) $A_i = B_i^n(v)$, $i = 1, \dots, m$, where $v = \{t_1, \dots, t_n\}$ is fixed and the B_i^n , $i = 1, \dots, m$, are disjoint sets in \mathcal{F}_v . Thus

$$\begin{aligned} P\left(\bigcup_{i=1}^m A_i\right) &= P\left(\bigcup_{i=1}^m B_i^n(v)\right) \\ &= P_v\left(\bigcup_{i=1}^m B_i^n\right) \quad \text{by definition of } P \\ &= \sum_{i=1}^m P_v(B_i^n) \quad \text{since } P_v \text{ is a measure} \\ &= \sum_{i=1}^m P(A_i) \quad \text{again by definition of } P. \end{aligned}$$

Therefore P is finitely additive on \mathcal{F}_0 . To show that P is countably additive on \mathcal{F}_0 , we must verify that P is continuous from above at \emptyset and invoke 1.2.8(b). The Carathéodory extension theorem (1.3.10) then extends P to \mathcal{F} .

Let A_k , $k = 1, 2, \dots$ be a sequence of measurable cylinders decreasing to \emptyset . If $P(A_k)$ does not approach 0, we have, for some $\varepsilon > 0$, $P(A_k) \geq \varepsilon > 0$ for all k . Suppose $A_k = B^{n_k}(v_k)$; by tacking on extra factors, we may assume that the numbers n_k and the sets v_k increase with k .

By 4.4.2, each Ω_{v_k} is a complete, separable metric space and $\mathcal{F}_{v_k} = \mathcal{B}(\Omega_{v_k})$. It follows from 4.3.8 that we can find a compact set $C^{n_k} \subset B^{n_k}$ such that $P_{v_k}(B^{n_k} - C^{n_k}) < \varepsilon/2^{k+1}$. Define $A'_k = C^{n_k}(v_k) \subset A_k$. Then $P(A_k - A'_k) =$

$P_{v_k}(B^{n_k} - C^{n_k}) < \varepsilon/2^{k+1}$. In this way we approximate the given cylinders by cylinders with compact bases.

Now take

$$D_k = A_1' \cap A_2' \cap \cdots \cap A_k' \subset A_1 \cap A_2 \cap \cdots \cap A_k = A_k.$$

Then

$$\begin{aligned} P(A_k - D_k) &= P\left(A_k \cap \bigcup_{i=1}^k A_i'^c\right) \leq \sum_{i=1}^k P(A_k \cap A_i'^c) \\ &\leq \sum_{i=1}^k P(A_i - A_i') < \sum_{i=1}^k \varepsilon/2^{i+1} < \varepsilon/2. \end{aligned}$$

Since $D_k \subset A_k' \subset A_k$, $P(A_k - D_k) = P(A_k) - P(D_k)$, consequently $P(D_k) > P(A_k) - \varepsilon/2$. In particular, D_k is not empty.

Now pick $x^k \in D_k$, $k = 1, 2, \dots$. Say $A_1' = C^{n_1}(t_1, \dots, t_{n_1}) = C^{n_1}(v_1)$ (note all $D_k \subset A_1'$). Consider the sequence

$$(x_{t_1}^1, \dots, x_{t_{n_1}}^1), \quad (x_{t_1}^2, \dots, x_{t_{n_1}}^2), \quad (x_{t_1}^3, \dots, x_{t_{n_1}}^3), \quad \dots,$$

that is, $x_{v_1}^1, x_{v_1}^2, x_{v_1}^3, \dots$.

Since the $x_{v_1}^n$ belong to C^{n_1} , a compact subset of Ω_{v_1} , we have a convergent subsequence $x_{v_1}^{r_{1n}}$ approaching some $x_{v_1} \in C^{n_1}$. If $A_2' = C^{n_2}(v_2)$ (so $D_k \subset A_2'$ for $k \geq 2$), consider the sequence $x_{v_2}^{r_{21}}, x_{v_2}^{r_{22}}, x_{v_2}^{r_{23}}, \dots \in C^{n_2}$ (eventually), and extract a convergent subsequence $x_{v_2}^{r_{2n}} \rightarrow x_{v_2} \in C^{n_2}$.

Note that $(x_{v_2}^{r_{2n}})_{v_1} = x_{v_1}^{r_{2n}}$; as $n \rightarrow \infty$, the left side approaches $(x_{v_2})_{v_1}$, and since $\{r_{2n}\}$ is a subsequence of $\{r_{1n}\}$, the right side approaches x_{v_1} . Hence $(x_{v_2})_{v_1} = x_{v_1}$.

Continue in this fashion; at step i we have a subsequence

$$x_{v_i}^{r_{in}} \rightarrow x_{v_i} \in C^{n_i}, \quad \text{and} \quad (x_{v_i})_{v_j} = x_{v_j} \quad \text{for } j < i.$$

Pick any $\omega \in \prod_{t \in T} \Omega_t$ such that $\omega_{v_j} = x_{v_j}$ for all $j = 1, 2, \dots$ (such a choice is possible since $(x_{v_i})_{v_j} = x_{v_j}$, $j < i$). Then $\omega_{v_j} \in C^{n_j}$ for each j ; hence

$$\omega \in \bigcap_{j=1}^{\infty} A_j' \subset \bigcap_{j=1}^{\infty} A_j = \emptyset,$$

a contradiction. Thus P extends to a measure on \mathcal{F} , and by construction, $\pi_v(P) = P_v$ for all v .

Finally, if P and Q are two probability measures on \mathcal{F} such that $\pi_v(P) = \pi_v(Q)$ for all finite $v \subset T$, then for any $B^n \in \mathcal{F}_v$,

$$P(B^n(v)) = [\pi_v(P)](B^n) = [\pi_v(Q)](B^n) = Q(B^n(v)).$$

Thus P and Q agree on measurable cylinders, and hence on \mathcal{F} by the uniqueness part of the Carathéodory extension theorem. ■

Problems

In Problems 1–7, $(\Omega_t, \mathcal{F}_t)$ is a measurable space for each $t \in T$, and $(\Omega, \mathcal{F}) = (\prod_{t \in T} \Omega_t, \prod_{t \in T} \mathcal{F}_t)$.

1. Let p_t be the projection map of Ω onto Ω_t , that is, $p_t(\omega) = \omega(t)$. If (S, \mathcal{S}) is a measurable space and $f: S \rightarrow \Omega$, show that f is measurable iff $p_t \circ f$ is measurable for all t .
2. If all $\Omega_t = R$, all $\mathcal{F}_t = \mathcal{B}(R)$, and T is a (nonempty) subset of R , how many sets are there in \mathcal{F} ?
3. Show that if $B \in \mathcal{F}$, then membership in B is determined by a countable number of coordinates, that is, there is a countable set $T_0 \subset T$ and a set $B_0 \in \mathcal{F}_{T_0} = \prod_{t \in T_0} \mathcal{F}_t$ such that $\omega \in B$ iff $\omega_{T_0} \in B_0$, where $\omega_{T_0} = (\omega(t), t \in T_0)$.
4. If T is an open interval of reals and $\Omega_t = R$ (or \bar{R}), $\mathcal{F}_t = \mathcal{B}(R)$ (or $\mathcal{B}(\bar{R})$) for all t , use Problem 3 to show that the following sets do *not* belong to \mathcal{F} :
 - (a) $\{\omega: \omega \text{ is continuous at } t_0\}$, where t_0 is a fixed element of T .
 - (b) $\{\omega: \sup_{a \leq t \leq b} \omega(t) < c\}$, where $c \in R$ and $[a, b] \subset T$.
5. Assume each Ω_t is a compact metric space, with \mathcal{F}_t the Baire (= Borel) subsets of Ω_t . Then by the Tychonoff theorem, Ω is compact in the topology of pointwise convergence. Show that \mathcal{F} is the class of Baire sets of Ω , in other words, $\mathcal{A}(\prod_{t \in T} \Omega_t) = \prod_{t \in T} \mathcal{A}(\Omega_t)$, as follows:
 - (a) If $A = \{\omega \in \Omega: \omega(t) \in F\}$, F a closed subset of Ω_t , show that $A = \{\omega \in \Omega: f(\omega) = 0\}$ for some $f \in C(\Omega)$. Conclude that $\mathcal{F} \subset \mathcal{A}(\Omega)$.
 - (b) Use the Stone–Weierstrass theorem to show that the functions $f \in C(\Omega)$ depending on only one coordinate [that is, $f(\omega) = g_t(\omega(t))$ for some t , where $g_t \in C(\Omega_t)$] generate an algebra that is dense in $C(\Omega)$. Conclude that $\mathcal{A}(\Omega) \subset \mathcal{F}$.
6. Assume T is an open interval of reals, and $(\Omega_t, \mathcal{F}_t) = (\bar{R}, \mathcal{B}(\bar{R}))$ for all t ; thus $\Omega = \bar{R}^T$, $\mathcal{F} = \mathcal{B}(\bar{R})^T$. Let $A = \{\omega \in \Omega: \omega \text{ is continuous at } t_0\}$, where t_0 is a fixed point of T .
 - (a) Show that A is an $F_{\sigma\delta}$ (a countable intersection of F_σ sets); hence $A \in \mathcal{B}(\Omega)$.
 - (b) Show that $A \notin \mathcal{A}(\Omega)$, so that we have an example of a compact Hausdorff space in which the Baire and Borel sets do not coincide.
7. (Alternative proof of the Kolmogorov extension theorem†) Assume the hypothesis of 4.4.3, with the stronger condition that each Ω_t is a compact metric space. Put the topology of pointwise convergence on Ω .
 - (a) Let A be the set of functions in $C(\Omega)$ that depend on only a finite number of coordinates; that is, there is a finite set $v \subset T$ and a con-

† Nelson. E., *Ann. of Math.* **69**, 630 (1959).

tinuous $g: \Omega_v \rightarrow R$ such that $f(\omega) = g(\omega_v)$, $\omega \in \Omega$. Use the Stone-Weierstrass theorem to show that A is dense in $C(\Omega)$.

- (b) If $f \in A$, define $E(f) = \int_{\Omega} f dP_v$. Show that E is well-defined and extends uniquely to a positive linear functional on $C(\Omega)$. Since $E(1) = 1$, the Riesz representation theorem 4.3.9 yields a unique probability measure P on $\mathcal{A}(\Omega)$ ($= \prod_{t \in T} \mathcal{F}_t$ by Problem 5) such that $E(f) = \int_{\Omega} f dP$ for all $f \in C(\Omega)$.
 - (c) Let v be a fixed finite subset of T , and let H be the collection of all functions $g: (\Omega_v, \mathcal{F}_v) \rightarrow (R, \mathcal{B}(R))$ such that if $f(\omega) = g(\omega_v)$, $\omega \in \Omega$, then $\int_{\Omega} f dP = \int_{\Omega} g dP_v$. Show that $I_A \in H$ for each open set $A \subset \Omega_v$, and then show that H contains all bounded Borel measurable functions on $(\Omega_v, \mathcal{F}_v)$. Conclude that $\pi_v(P) = P_v$. (Uniqueness of P is proved as in 4.4.3.)
8. The metric space Ω is said to be *Borel equivalent* to a subset of the metric space Ω' iff there is a one-to-one map $f: \Omega \rightarrow \Omega'$ such that $E = f(\Omega) \in \mathcal{B}(\Omega')$ and f and f^{-1} are Borel measurable. [Measurability of f^{-1} means that $f^{-1}: (E, \mathcal{B}(E)) \rightarrow (\Omega, \mathcal{B}(\Omega))$.]
- (a) Let Ω be a complete separable metric space with metric d . [It may be assumed without loss of generality that $d(x, y) \leq 1$ for all x, y , since the metric $d' = d/(1 + d)$ induces the same topology as d and is also complete.] Denote by $[0, 1]^\infty$ the space of all sequences of real numbers with components in $[0, 1]$, with the topology of pointwise convergence. (This space is metrizable; explicitly, we may take the metric as

$$d_0(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

If $D = \{\omega_1, \omega_2, \dots\}$ is a countable dense subset of Ω , define $f: \Omega \rightarrow [0, 1]^\infty$ by $f(\omega) = \{d(\omega, \omega_n), n = 1, 2, \dots\}$. Show that f is continuous and one-to-one, with a continuous inverse.

- (b) Show that $f(\Omega)$ is a Borel subset of $[0, 1]^\infty$; thus Ω is Borel equivalent to a subset of $[0, 1]^\infty$.
- (c) Let $S = \{0, 1\}$, and let \mathcal{S} consist of all subsets of S . Show that there is a map $g: ([0, 1], \mathcal{B}[0, 1]) \rightarrow (S^\infty, \mathcal{S}^\infty)$ such that g is one-to-one, $g[0, 1] \in \mathcal{S}^\infty$, and both g and g^{-1} are measurable.
- (d) Show that $([0, 1], \mathcal{B}[0, 1])$ and $(S^\infty, \mathcal{S}^\infty)$ are equivalent, that is, there is a map $h: [0, 1] \rightarrow S^\infty$ such that h is one-to-one onto, and h and h^{-1} are measurable.
- (e) If $(\Omega_n, \mathcal{F}_n)$ is equivalent to (S_n, \mathcal{S}_n) , with associated map h_n , $n = 1, 2, \dots$, show that $(\prod_n \Omega_n, \prod_n \mathcal{F}_n)$ is equivalent to $(\prod_n S_n, \prod_n \mathcal{S}_n)$. Thus by (d), $([0, 1]^\infty, \mathcal{B}[0, 1])^\infty$ is equivalent to $(S^\infty, \mathcal{S}^\infty)$.

Now by 4.4.2, $(\mathcal{B}[0, 1])^\infty$ is the minimal σ -field over the open sets of $[0, 1]^\infty$, that is, $(\mathcal{B}[0, 1])^\infty$ is the class of Borel sets $\mathcal{B}([0, 1]^\infty)$. It follows from these results that if Ω is a complete, separable metric space, Ω is Borel equivalent to a (Borel) subset of $[0, 1]$.

4.5 Weak Convergence of Measures

By the Riesz representation theorem, a continuous linear functional on $C(\Omega)$, where Ω is compact Hausdorff, may be identified with a regular finite signed measure on $\mathcal{B}(\Omega)$. Thus if $\{\mu_n\}$ is a sequence of such measures, weak* convergence of the sequence to the measure μ means that $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all $f \in C(\Omega)$. In this section we investigate this type of convergence in a somewhat different context; Ω will be a metric space, not necessarily compact, and all measures will be nonnegative. The results form the starting point for the study of the central limit theorem of probability.

4.5.1 Theorem. Let μ, μ_1, μ_2, \dots be finite measures on the Borel sets of a metric space Ω . The following conditions are equivalent:

- (a) $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all bounded continuous $f: \Omega \rightarrow \mathbb{R}$.
- (b) $\liminf_{n \rightarrow \infty} \int_{\Omega} f d\mu_n \geq \int_{\Omega} f d\mu$ for all bounded lower semicontinuous $f: \Omega \rightarrow \mathbb{R}$.
- (b') $\limsup_{n \rightarrow \infty} \int_{\Omega} f d\mu_n \leq \int_{\Omega} f d\mu$ for all bounded upper semicontinuous $f: \Omega \rightarrow \mathbb{R}$.
- (c) $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all bounded $f: (\Omega, \mathcal{B}(\Omega)) \rightarrow (R, \mathcal{B}(R))$ such that f is continuous a.e. $[\mu]$.
- (d) $\liminf_{n \rightarrow \infty} \mu_n(A) \geq \mu(A)$ for every open set $A \subset \Omega$, and $\mu_n(\Omega) \rightarrow \mu(\Omega)$.
- (d') $\limsup_{n \rightarrow \infty} \mu_n(A) \leq \mu(A)$ for every closed set $A \subset \Omega$, and $\mu_n(\Omega) \rightarrow \mu(\Omega)$.
- (e) $\mu_n(A) \rightarrow \mu(A)$ for every $A \in \mathcal{B}(\Omega)$ such that $\mu(\partial A) = 0$ (∂A denotes the boundary of A).

PROOF. (a) implies (b): If $g \leq f$ and g is bounded continuous,

$$\liminf_{n \rightarrow \infty} \int_{\Omega} f d\mu_n \geq \liminf_{n \rightarrow \infty} \int_{\Omega} g d\mu_n = \int_{\Omega} g d\mu \quad \text{by (a).}$$

But since f is lower semicontinuous (LSC), it is the limit of a sequence of continuous functions, and if $|f| \leq M$, all functions in the sequence can also be taken less than or equal to M in absolute value. (See the appendix on general topology, Section A6, for the basic properties of semicontinuous

functions.) Thus if we take the sup over g in the above equation, we obtain (b).

(b) is equivalent to (b'): Note that f is LSC iff $-f$ is upper semicontinuous (USC).

(b) implies (c): Let \bar{f} be the lower envelope of f (the sup of all LSC functions g such that $g \leq f$) and \tilde{f} the upper envelope (the inf of all USC functions g such that $g \geq f$). Since $\bar{f}(x) = \liminf_{y \rightarrow x} f(y)$ and $\tilde{f}(x) = \limsup_{y \rightarrow x} f(y)$, continuity of f at x implies $\bar{f}(x) = f(x) = \tilde{f}(x)$. Furthermore, \bar{f} is LSC and \tilde{f} is USC. Thus if f is bounded and continuous a.e. $[\mu]$,

$$\begin{aligned} \int_{\Omega} f d\mu &= \int_{\Omega} \bar{f} d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} \bar{f} d\mu_n \quad \text{by (b)} \\ &\leq \liminf_{n \rightarrow \infty} \int_{\Omega} f d\mu_n \quad \text{since } \bar{f} \leq f \\ &\leq \limsup_{n \rightarrow \infty} \int_{\Omega} f d\mu_n \\ &\leq \limsup_{n \rightarrow \infty} \int_{\Omega} \tilde{f} d\mu_n \quad \text{since } f \leq \tilde{f} \\ &\leq \int_{\Omega} \tilde{f} d\mu \quad \text{by (b')} \\ &= \int_{\Omega} f d\mu, \quad \text{proving (c).} \end{aligned}$$

(c) implies (d): Clearly (c) implies (a), which in turn implies (b). If A is open, then I_A is LSC, so by (b), $\liminf_{n \rightarrow \infty} \mu_n(A) \geq \mu(A)$. Now $I_{\Omega} \equiv 1$, so $\mu_n(\Omega) \rightarrow \mu(\Omega)$ by (c).

(d) is equivalent to (d'): Take complements.

(d) implies (e): Let A^0 be the interior of A , \bar{A} the closure of A . Then

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu_n(A) &\leq \limsup_{n \rightarrow \infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) \quad \text{by (d')} \\ &= \mu(A) \quad \text{by hypothesis.} \end{aligned}$$

Also, using (d),

$$\liminf_{n \rightarrow \infty} \mu_n(A) \geq \liminf_{n \rightarrow \infty} \mu_n(A^0) \geq \mu(A^0) = \mu(A).$$

(e) implies (a): Let f be a bounded continuous function on Ω . If $|f| < M$, let $A = \{c \in R: \mu(f^{-1}\{c\}) \neq 0\}$; A is countable since the $f^{-1}\{c\}$ are disjoint and μ is finite. Construct a partition of $[-M, M]$, say $-M = t_0 < t_1 < \dots < t_j = M$, with $t_i \notin A$, $i = 0, 1, \dots, j$ (M may be increased if necessary). If $B_i = \{x: t_i \leq f(x) < t_{i+1}\}$, $i = 0, 1, \dots, j-1$, it follows from (e) that

$$\sum_{i=0}^{j-1} t_i \mu_n(B_i) \rightarrow \sum_{i=0}^{j-1} t_i \mu(B_i).$$

[Since $f^{-1}(t_i, t_{i+1})$ is open, $\partial f^{-1}[t_i, t_{i+1}] \subset f^{-1}\{t_i, t_{i+1}\}$, and $\mu f^{-1}\{t_i, t_{i+1}\} = 0$ since $t_i, t_{i+1} \notin A$.] Now

$$\begin{aligned} \left| \int_{\Omega} f d\mu_n - \int_{\Omega} f d\mu \right| &\leq \left| \int_{\Omega} f d\mu_n - \sum_{i=0}^{j-1} t_i \mu_n(B_i) \right| \\ &\quad + \left| \sum_{i=0}^{j-1} t_i \mu_n(B_i) - \sum_{i=0}^{j-1} t_i \mu(B_i) \right| \\ &\quad + \left| \sum_{i=0}^{j-1} t_i \mu(B_i) - \int_{\Omega} f d\mu \right|. \end{aligned}$$

The first term on the right may be written as

$$\left| \sum_{i=0}^{j-1} \int_{B_i} (f(x) - t_i) d\mu_n(x) \right|$$

and this is bounded by $\max_i (t_{i+1} - t_i) \mu_n(\Omega)$, which can be made arbitrarily small by choice of the partition since $\mu_n(\Omega) \rightarrow \mu(\Omega) < \infty$ by (e). The third term on the right is bounded by $\max_i (t_{i+1} - t_i) \mu(\Omega)$, which can also be made arbitrarily small. The second term approaches 0 as $n \rightarrow \infty$, proving (a). ■

4.5.2 Comments. Another condition equivalent to those of 4.5.1 is that $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all bounded uniformly continuous $f: \Omega \rightarrow R$ (see Problem 1).

The proof of 4.5.1 works equally well if the sequence $\{\mu_n\}$ is replaced by a net.

The convergence described in 4.5.1 is sometimes called *weak* or *vague* convergence of measures. We shall write $\mu_n \xrightarrow{w} \mu$.

If the measures μ_n and μ are defined on $\mathcal{B}(R)$, there are corresponding distribution functions F_n and F on R . We may relate convergence of measures to convergence of distribution functions.

4.5.3 Definition. A *continuity point* of a distribution function F on R is a point $x \in R$ such that F is continuous at x , or $\pm\infty$ (thus by convention, ∞ and $-\infty$ are continuity points).

4.5.4 Theorem. Let μ, μ_1, μ_2, \dots be finite measures on $\mathcal{B}(R)$, with corresponding distribution functions F, F_1, F_2, \dots . The following are equivalent:

- (a) $\mu_n \xrightarrow{w} \mu$.
- (b) $F_n(a, b] \rightarrow F(a, b]$ at all continuity points a, b of F , where $F(a, b] = F(b) - F(a)$, $F(\infty) = \lim_{x \rightarrow \infty} F(x)$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$.

If all distribution functions are 0 at $-\infty$, condition (b) is equivalent to the statement that $F_n(x) \rightarrow F(x)$ at all points $x \in R$ at which F is continuous, and $F_n(\infty) \rightarrow F(\infty)$.

PROOF. (a) *implies* (b): If a and b are continuity points of F in R , then $(a, b]$ is a Borel set whose boundary has μ -measure 0. By 4.5.1(e), $\mu_n(a, b] \rightarrow \mu(a, b]$, that is, $F_n(a, b] \rightarrow F(a, b]$. If $a = -\infty$, the argument is the same, and if $b = \infty$, then (a, ∞) is a Borel set whose boundary has μ -measure 0, and the proof proceeds as before.

(b) *implies* (a): Let A be an open subset of R ; write A as the disjoint union of open intervals I_1, I_2, \dots . Then

$$\begin{aligned}\liminf_{n \rightarrow \infty} \mu_n(A) &= \liminf_{n \rightarrow \infty} \sum_k \mu_n(I_k) \\ &\geq \sum_k \liminf_{n \rightarrow \infty} \mu_n(I_k) \quad \text{by Fatou's lemma.}\end{aligned}$$

Let $\varepsilon > 0$ be given. For each k , let I'_k be a right-semiclosed subinterval of I_k such that the endpoints of I'_k are continuity points of F , and $\mu(I'_k) \geq \mu(I_k) - \varepsilon 2^{-k}$; the I'_k can be chosen since F has only countably many discontinuities. Then

$$\liminf_{n \rightarrow \infty} \mu_n(I_k) \geq \liminf_{n \rightarrow \infty} \mu_n(I'_k) = \mu(I'_k) \quad \text{by (b).}$$

Thus

$$\liminf_{n \rightarrow \infty} \mu_n(A) \geq \sum_k \mu(I'_k) \geq \sum_k \mu(I_k) - \varepsilon = \mu(A) - \varepsilon.$$

Since ε is arbitrary, we have $\mu_n \xrightarrow{w} \mu$ by 4.5.1(d). ■

Condition (b) of 4.5.4 is sometimes called weak convergence of the sequence $\{F_n\}$ to F , written $F_n \xrightarrow{w} F$.

Problems

1. (a) If F is a closed subset of the metric space Ω , show that I_F is the limit of a decreasing sequence of uniformly continuous functions f_n , with $0 \leq f_n \leq 1$ for all n .
 (b) Show that in 4.5.1, $\mu_n \xrightarrow{w} \mu$ iff $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all bounded uniformly continuous $f: \Omega \rightarrow R$.
2. Show that in 4.5.1, $\mu_n \xrightarrow{w} \mu$ iff $\mu_n(A) \rightarrow \mu(A)$ for all open sets A such that $\mu(\partial A) = 0$.
3. Let Ω be a locally compact metric space. If $A \subset \Omega$, A is said to be *bounded* iff $A \subset K$ for some compact set K .

If μ, μ_1, μ_2, \dots are measures on $\mathcal{B}(\Omega)$ that are finite on bounded Borel sets, we say that $\mu_n \xrightarrow{w} \mu$ iff $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all continuous functions $f: \Omega \rightarrow R$ with compact support. (The support of f is defined by $\text{supp } f = \{x: f(x) \neq 0\}$; note that f has compact support iff f vanishes outside a compact set.)

Show that the following conditions are equivalent:

- (a) $\mu_n \xrightarrow{w} \mu$.
- (b) $\mu_n(A) \rightarrow \mu(A)$ for all bounded Borel sets A with $\mu(\partial A) = 0$.
- (c) $\int_{\Omega} f d\mu_n \rightarrow \int_{\Omega} f d\mu$ for all bounded $f: \Omega \rightarrow R$ such that f has compact support and is continuous a.e. $[\mu]$.

4.6 References

The approach to the Daniell integral and the Riesz representation theorem given in Sections 4.2 and 4.3 is based on Neveu (1965). If the hypothesis that L contains the constant functions is dropped in the development of the Daniell integral, the situation becomes more complicated. However, a representation theorem somewhat similar to 4.2.9 may be proved under the assumption that $f \in L$ implies $1 \wedge f \in L$. For a detailed proof of this result (known as *Stone's theorem*), see Royden (1968, Chapter 13).

There are other versions of the Riesz representation theorem appropriate in a locally compact Hausdorff space Ω . If E is a positive linear functional on the set $C_c(\Omega)$ of continuous real-valued functions on Ω with compact support, there is a unique regular measure μ on the Borel sets such that $E(f) = \int_{\Omega} f d\mu$ for each f in $C_c(\Omega)$. Here, regularity means that μ is finite on compact sets, $\mu(A) = \inf\{\mu(V): V \supset A, V \text{ open}\}$ for all Borel sets A , and $\mu(A) = \sup\{\mu(K): K \subset A, K \text{ compact}\}$ for all Borel sets A which are either open or of finite measure. Also, if E is a continuous linear functional on the Banach space $C_0(\Omega)$ of continuous real-valued functions on Ω that vanish at ∞ (f vanishes at ∞ iff, given $\varepsilon > 0$, there is a compact set K such that $|f| < \varepsilon$ off K), there is a unique regular finite signed measure on the Borel sets such that $E(f) = \int_{\Omega} f d\mu$ for all $f \in C_0(\Omega)$; furthermore, $\|E\| = |\mu|(\Omega)$. As in the compact case, the result may be extended to complex-valued functions and complex measures, as in Problem 6, Section 4.3. Proofs are given in Rudin (1966).

In probability theory, there is usually no reason to replace compact by locally compact spaces in order to take advantage of a more general version of the Riesz representation theorem. For if we have a random experiment whose outcomes are represented by points in a locally compact space, we simply compactify the space.

For an account of the theory of weak convergence of measures, with applications to probability, see Billingsley (1968) and Parthasarathy (1967).

5

Basic Concepts of Probability

5.1 Introduction

The starting point for probability theory is a set Ω called the *sample space* whose points are in one-to-one correspondence with the possible outcomes of a given performance of a random experiment. For example, if two dice are tossed, we may take Ω to have 36 points, one for each ordered pair (i, j) , $i, j = 1, \dots, 6$. The sample space for a given experiment is not unique. For example, if two dice are tossed and N is the sum of the faces, we may take Ω to consist of 11 points, corresponding to the outcomes $N = 2, 3, \dots, 12$. The particular sample space to be used will be determined by the problem at hand. For example, in the dice-tossing example above, if we are interested in the result of the first toss, the sample space corresponding to $N = 2, 3, \dots, 12$ will not be of value.

An “event” in a random experiment corresponds to a question that can be answered “yes” or “no.” For example, in the dice-tossing example, let Ω be the set of ordered pairs (i, j) , $i, j = 1, \dots, 6$. We may ask the question “Is the maximum of the two coordinates i and j less than or equal to 2?” The subset of Ω associated with a “yes” answer is $A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$; the subset associated with a “no” answer is A^c .

Thus it is reasonable to define an event as a subset of Ω . However, in some situations not all subsets may be regarded as events. As an example,

admittedly somewhat artificial, suppose that a coin is tossed four times, and Ω consists of the 16 sequences of length 4 with components H and T . Assume that only the results of the first two tosses are written down. If A is the set of points of Ω corresponding to at least three heads, then A is not “measurable,” that is, the given information concerning ω is not sufficient to determine whether or not $\omega \in A$. More serious problems arise when Ω is R^n ; in this case we are almost always forced by mathematical consistency requirements to take the event class to be the Borel sets rather than the collection of all subsets of Ω .

The development of the mathematical theory will be facilitated if we require that the event class form a σ -field. Thus we may form countable unions, countable intersections, and complements of events and be assured that the resulting sets are also events.

Finally, we must talk about the probability $P(A)$ assigned to an event A . The basic physical requirement is that $P(A)$ correspond to the relative frequency of A in a very large number of independent repetitions of the random experiment. It follows that P should be a nonnegative, finitely additive set function, with $P(\Omega) = 1$. As indicated in the discussion of 1.2.3, there does not seem to be a convincing physical argument that P should be countably additive rather than simply finitely additive. On the other hand, there is no convincing physical argument against the assumption of countable additivity, and this hypothesis will allow a smooth development of the mathematical theory.

The above discussion may be summarized by saying that P is a probability measure on the σ -field \mathcal{F} . Thus the basic mathematical object we are to study is a probability space (Ω, \mathcal{F}, P) .

5.2 Discrete Probability Spaces

If the sample space Ω is a finite or countably infinite set, measure-theoretic difficulties do not arise. We take \mathcal{F} to consist of all subsets of Ω , and assign probabilities in the following canonical way. Let $\Omega = \{\omega_1, \omega_2, \dots\}$, and let p_1, p_2, \dots be nonnegative numbers whose sum is 1. If A is any subset of Ω , we define

$$P(A) = \sum_{\omega_i \in A} p_i.$$

In particular,

$$P\{\omega_i\} = p_i.$$

Then P is a probability measure, and the probability of the event A is computed simply by adding the probabilities of the individual points of A .

If Ω is countable, \mathcal{F} consists of all subsets of Ω , and P is defined as above, (Ω, \mathcal{F}, P) is called a *discrete probability space*.

Classical probability theory was concerned with the special case $\Omega = \{\omega_1, \dots, \omega_n\}$, $p_i = 1/n$, $i = 1, \dots, n$. In this case,

$$P(A) = \frac{\text{number of points in } A}{\text{number of points in } \Omega}.$$

Thus to find $P(A)$ we count the number of favorable outcomes and divide by the total number of outcomes.

Unless otherwise specified, if Ω is countable, we always take \mathcal{F} to contain all subsets of Ω . Thus all subsets of Ω are events, all functions on Ω are measurable, and measure-theoretic machinery is not needed.

5.3 Independence

Intuitively, two events A and B are independent if a statement concerning the occurrence or nonoccurrence of one of the events does not change the odds about the other event. Let us translate the physical requirement into mathematical terms. Suppose, for example, that $P(A) = 0.4$ and $P(B) = 0.3$. In a long sequence of repetitions of the random experiment, A will occur approximately 40 percent of the time. If B is to be independent of A , the occurrence of A will not influence the odds about B , so that if we examine only those trials on which A occurs, B will occur roughly 30 percent of the time; hence $P(A \cap B) = (0.4)(0.3) = P(A)P(B)$. Similarly, the nonoccurrence of A will not influence the odds about B ; hence $P(A^c \cap B) = (0.6)(0.3) = P(A^c)P(B)$.

Conversely, if $P(A \cap B) = P(A)P(B)$ and $P(A^c \cap B) = P(A^c)P(B)$, B will be independent of A . For if we examine only the trials on which A occurs, B must occur roughly 30 percent of the time in order to have $P(A \cap B) = P(A)P(B)$. Thus the occurrence of A does not influence the odds about B , and similarly, the occurrence of A^c does not change the odds about B .

The above discussion suggests that we call the event B independent of A if and only if $P(A \cap B) = P(A)P(B)$ and $P(A^c \cap B) = P(A^c)P(B)$. However, the first condition implies the second. For suppose $P(A \cap B) = P(A)P(B)$. Then

$$\begin{aligned} P(A^c \cap B) &= P(B - A) = P(B - A \cap B) \\ &= P(B) - P(A \cap B) \quad \text{since } A \cap B \subset B \\ &= P(B) - P(A)P(B) \quad \text{by hypothesis} \\ &= (1 - P(A))P(B) \\ &= P(A^c)P(B). \end{aligned}$$

Therefore B is independent of A if and only if $P(A \cap B) = P(A)P(B)$. But this condition is not altered if A and B are interchanged; hence B is independent of A if and only if A is independent of B . We may therefore formulate the definition of independence as follows.

5.3.1 Definition. Two events A and B are *independent* iff $P(A \cap B) = P(A)P(B)$.

We now consider independence of more than two events. If the events $A_i, i \in I$, are to be independent, and i_1, \dots, i_k are distinct indices, a statement about one or more of the events A_{i_1}, \dots, A_{i_k} should not change the odds about any of the remaining events. The physical discussion at the beginning of the section leads to the following requirement.

5.3.2 Definition. Let I be an arbitrary index set, and let $A_i, i \in I$, be events in a given probability space. The A_i are *independent* iff for all finite collections $\{i_1, \dots, i_k\}$ of distinct indices in I , we have

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

5.3.3 Comments. (a) If the events $A_i, i \in I$, are independent and any event is replaced by its complement, independence is maintained, that is,

$$P(B_{i_1} \cap \cdots \cap B_{i_k}) = P(B_{i_1}) \cdots P(B_{i_k}),$$

where i_1, \dots, i_k are distinct indices and B_{i_j} is either A_{i_j} or $A_{i_j}^c$, $j = 1, \dots, k$.

We have essentially proved this in the discussion preceding the definition of independence by showing that $P(A \cap B) = P(A)P(B)$ implies $P(A^c \cap B) = P(A^c)P(B)$.

(b) If A_1, \dots, A_n are events such that A_{i_1}, \dots, A_{i_k} are independent for all distinct indices $i_1, \dots, i_k, k = 2, \dots, n - 1$, it does not follow that A_1, \dots, A_n are independent. For example, let a coin be tossed twice, and assign probability $\frac{1}{4}$ to each of the outcomes HH, HT, TH, TT. Let $A = \{\text{first toss is a head}\}$, $B = \{\text{second toss is a head}\}$, $C = \{\text{first toss} = \text{second toss}\} = \{\text{HH, TT}\}$. Then

$$P(A \cap B) = \frac{1}{4} = P(A)P(B),$$

$$P(A \cap C) = \frac{1}{4} = P(A)P(C),$$

$$P(B \cap C) = \frac{1}{4} = P(B)P(C).$$

But

$$P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C).$$

Thus A and B are independent, as are A and C , and also B and C , but A , B , and C are not independent.

Conversely, if $P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$, it does not follow that $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$ when $k < n$. For example, suppose that two dice are tossed, and let Ω be all ordered pairs (i, j) , $i, j = 1, 2, \dots, 6$, with probability $\frac{1}{36}$ assigned to each point. Let

$$\begin{aligned}A &= \{\text{second die is } 1, 2 \text{ or } 5\}, \\B &= \{\text{second die is } 4, 5 \text{ or } 6\}, \\C &= \{\text{the sum of the faces is } 9\}.\end{aligned}$$

Then

$$\begin{aligned}P(A \cap B) &= \frac{1}{6} \neq P(A)P(B) = \frac{1}{4}, \\P(A \cap C) &= \frac{1}{36} \neq P(A)P(C) = \frac{1}{18}, \\P(B \cap C) &= \frac{1}{12} \neq P(B)P(C) = \frac{1}{18},\end{aligned}$$

but

$$P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C).$$

5.4 Bernoulli Trials

A sequence of n *Bernoulli trials* consists of n independent observations, with the property that each observation has only two possible results, called “success” and “failure.” The probability of success on a given trial is p and the probability of failure is $q = 1 - p$.

To construct a probability space that meets the given physical requirements, we take Ω to be all 2^n ordered sequences of length n with components 1 and 0, with 1 indicating success and 0 failure. Consider a typical sample point ω with ones in positions i_1, \dots, i_k and zeros in positions i_{k+1}, \dots, i_n . If A_i is the event of obtaining a success on trial i , so that $A_i = \{\omega: \text{the } i\text{th coordinate of } \omega \text{ is } 1\}$, we have

$$\{\omega\} = A_{i_1} \cap \cdots \cap A_{i_k} \cap A_{i_{k+1}}^c \cap \cdots \cap A_{i_n}^c.$$

Since the trials are independent and $P(A_i) = p$ for all i , the probability assigned to ω is determined; it must be

$$P\{\omega\} = P(A_{i_1}) \cdots P(A_{i_k})P(A_{i_{k+1}}^c) \cdots P(A_{i_n}^c) = p^k q^{n-k}.$$

Now there are $\binom{n}{k} = n!/k!(n-k)!$ sequences in Ω having exactly k ones, because such a sequence is determined by selecting k positions out of n for the ones to occur. Thus the probability of obtaining exactly k successes is

$$p(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n. \tag{1}$$

By the binomial theorem, $\sum_{k=0}^n p(k) = (p + q)^n = 1$; therefore the sum of the probabilities assigned to all points is 1, and we have a legitimate probability measure.

A sequence of n *generalized Bernoulli trials* consists of n independent observations, such that each observation has k possible outcomes ($k > 2$). If the k outcomes are labeled b_1, \dots, b_k , the probability that b_i will occur on a given trial is p_i , where the p_i are nonnegative and $\sum_{i=1}^k p_i = 1$.

To construct an appropriate probability space, we take Ω to be all k^n ordered sequences of length n with components b_1, \dots, b_k . If ω is a sample point having n_i occurrences of b_i , $i = 1, \dots, k$, the independence of the trials and the assumption that the probability of obtaining b_i on a given trial is p_i leads us to assign to ω the probability $p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$.

Now to find the number of sequences in Ω in which b_i occurs exactly n_i times, $i = 1, \dots, k$, we reason as follows. Such a sequence is determined by selecting n_1 positions out of n to be occupied by b_1 's, then n_2 positions from the remaining $n - n_1$ for the b_2 's, and so on. Thus the total number of sequences is

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}} \binom{n_k}{n_k} \\ = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

The total probability assigned to all points is

$$\sum \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

where the sum is taken over all nonnegative integers n_1, \dots, n_k whose sum is n . But this is $(p_1 + \cdots + p_k)^n = 1$, using the multinomial theorem.

The probability that b_1 will occur n_1 times, b_2 will occur n_2 times, \dots , and b_k will occur n_k times, is

$$p(n_1, \dots, n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}, \quad (2)$$

where $n_1, \dots, n_k = 0, 1, \dots, n_1 + \dots + n_k = n$.

5.5 Conditional Probability

If two events A and B are independent, a statement about the occurrence or nonoccurrence of one of the events does not change the odds about the other. In the absence of independence, the odds are altered, and the concept of conditional probability gives a quantitative measure of the change.

For example, suppose that the probability of A is 0.4 and the probability of $A \cap B$ is 0.1. If we repeat the experiment independently a large number of times and examine only the trials on which A has occurred, B will occur roughly 25 percent of the time. In general, the ratio $P(A \cap B)/P(A)$ is a measure of the probability of B under the condition that A is known to have occurred.

We therefore define the *conditional probability of B given A* , as

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (1)$$

provided $P(A) > 0$.

In the next chapter, we shall discuss in detail the concept of conditional probability $P(B|A)$ when the event A has probability 0. This is not a degenerate case; there are many natural and intuitive examples. Of course, the definition (1) no longer makes sense, and the approach will be somewhat indirect. At this point we shall only derive a few consequences of (1).

5.5.1 Theorem. (a) If $P(A) > 0$, A and B are independent iff $P(B|A) = P(B)$. [Similarly, independence is equivalent to $P(A|B) = P(A)$ if $P(B) > 0$.]

(b) If $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

PROOF. Part (a) follows from the definitions of independence and conditional probability. To prove (b), observe that $P(A_1 \cap \cdots \cap A_{n-1}) > 0$ implies that $P(A_1), P(A_1 \cap A_2), \dots, P(A_1 \cap \cdots \cap A_{n-2}) > 0$, so all conditional probabilities are well defined. Now by the definition of conditional probability,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1 \cap \cdots \cap A_{n-1})P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

An induction argument completes the proof. ■

The following result will be quite useful.

5.5.2 Theorem of Total Probability. Let B_1, B_2, \dots form a finite or countably infinite family of *mutually exclusive and exhaustive events*, that is, the B_i are disjoint and their union is Ω .

(a) If A is any event, then $P(A) = \sum_i P(A \cap B_i)$. Thus $P(A)$ is calculated by making a list of mutually exclusive, exhaustive ways in which A can happen, and adding the individual probabilities.

(b) $P(A) = \sum_i P(B_i)P(A|B_i)$, where the sum is taken over those i for which $P(B_i) > 0$. Thus $P(A)$ is a weighted average of the conditional probabilities $P(A|B_i)$.

PROOF.

- (a) $P(A) = P(A \cap \Omega) = P(A \cap \bigcup_i B_i) = P(\bigcup_i (A \cap B_i)) = \sum_i P(A \cap B_i)$.
- (b) This follows from (a) and the fact that $P(A \cap B_i) = 0$ if $P(B_i) = 0$, and equals $P(B_i)P(A|B_i)$ if $P(B_i) > 0$. ■

5.5.3 Example. A positive integer I is selected, with $P\{I = n\} = (\frac{1}{2})^n$, $n = 1, 2, \dots$. If I takes the value n , a coin with probability e^{-n} of heads is tossed once. Find the probability that the resulting toss is a head.

Here we have specified $P(B_n) = (\frac{1}{2})^n$, where $B_n = \{I = n\}$, $n = 1, 2, \dots$. If A is the event that the coin comes up heads, we have specified $P(A|B_n) = e^{-n}$. By the theorem of total probability, this is enough to determine $P(A)$. Formally, we may take Ω to consist of all ordered pairs (n, m) , $n = 1, 2, \dots$, $m = 0$ (tail) or 1 (head). We assign to the point $(n, 1)$ the probability $(\frac{1}{2})^n e^{-n}$, and to $(n, 0)$ the probability $(\frac{1}{2})^n (1 - e^{-n})$. We may then verify that $P(B_n) = (\frac{1}{2})^n$, $P(A|B_n) = e^{-n}$. Hence

$$P(A) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n e^{-n} = \frac{e^{-1}/2}{1 - (e^{-1}/2)}.$$

5.6 Random Variables

Intuitively, a random variable is a quantity that is measured in connection with a random experiment. If (Ω, \mathcal{F}, P) is a probability space and the outcome of the experiment corresponds to the point $\omega \in \Omega$, a measuring process is carried out to obtain a number $X(\omega)$. Thus X is a function from the sample space Ω to the reals (or the extended reals). For example, if (Ω, \mathcal{F}, P) corresponds to a sequence of four Bernoulli trials (Section 5.4) and X is the number of successes, then $X(1 0 1 1) = 3$, $X(0 1 0 0) = 1$, and so on.

If we are interested in a random variable X defined on a given probability space, we generally want to know the probability of events involving X ; for example, the probability that in a given performance of the experiment the value of X will belong to B , where B is a set of real numbers. Thus we are looking for $P\{\omega: X(\omega) \in B\}$, and for this to make sense, $\{\omega: X(\omega) \in B\}$ must be an event, in other words, must belong to the σ -field \mathcal{F} . Also, in order to make use of the integration theory that we have available, we restrict B to the Borel subsets.

5.6.1 Definitions. A random variable X on a probability space (Ω, \mathcal{F}, P) is a Borel measurable function from Ω to R . In the terminology of Section 1.5, we have $X: (\Omega, \mathcal{F}) \rightarrow (R, \mathcal{B}(R))$. In many situations it is convenient to allow

X to take on the values $\pm\infty$; X is said to be an *extended random variable* iff X is a Borel measurable function from Ω to \bar{R} , that is, $X: (\Omega, \mathcal{F}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$.

If X is a random variable on (Ω, \mathcal{F}, P) the *probability measure induced by X* is the probability measure P_X on $\mathcal{B}(R)$ given by

$$P_X(B) = P\{\omega: X(\omega) \in B\}, \quad B \in \mathcal{B}(R).$$

The numbers $P_X(B)$, $B \in \mathcal{B}(R)$, completely characterize the random variable X in the sense that they provide the probabilities of all events involving X . It is useful to know that this information may be captured by a single function from R to R .

5.6.2 Definition. The *distribution function* of a random variable X is the function $F = F_X$ from R to $[0, 1]$ given by

$$F(x) = P\{\omega: X(\omega) \leq x\}, \quad x \text{ real.}$$

Since, for $a < b$, $F(b) - F(a) = P\{\omega: a < X(\omega) \leq b\} = P_X(a, b]$, F is a distribution function corresponding to the Lebesgue-Stieltjes measure P_X (Section 1.4). In particular, F is increasing and right-continuous. By 1.2.7, $F(x) \rightarrow 1$ as $x \rightarrow \infty$ and $F(x) \rightarrow 0$ as $x \rightarrow -\infty$. Thus among all distribution functions corresponding to P_X , we choose the one with $F(\infty) = 1$, $F(-\infty) = 0$.

Very often, the following statement is made: “Let X be a random variable with distribution function F ,” where F is a given function from R to $[0, 1]$ that is increasing and right-continuous, with $F(\infty) = 1$, $F(-\infty) = 0$. There is no reference to the underlying probability space (Ω, \mathcal{F}, P) , and actually the nature of the underlying space is not important. For the distribution function F determines the probability measure P_X , which in turn determines the probability of all events involving X . The only thing we have to check is that there be at least one (Ω, \mathcal{F}, P) on which a random variable X with distribution function F can be defined. In fact we can always supply the probability space in a canonical way; take $\Omega = R$, $\mathcal{F} = \mathcal{B}(R)$, with P the Lebesgue-Stieltjes measure corresponding to F , and define $X(\omega) = \omega$, $\omega \in \Omega$, that is, X is the identity map. Since $P_X(B) = P\{\omega: X(\omega) \in B\} = P(B)$, X has induced probability measure P and therefore distribution function F .

Thus if $F: R \rightarrow [0, 1]$ is increasing and right-continuous, with $F(\infty) = 1$, $F(-\infty) = 0$, then F is the distribution function of some random variable.

We isolate some particularly important classes of random variables.

5.6.3 Definitions and Comments. Let X be a random variable on (Ω, \mathcal{F}, P) . We say that X is *simple* iff X can take on only finitely many possible values, *discrete* iff the set of values of X is finite or countably infinite. (Any random variable on a discrete probability space is discrete, since Ω is countable.)

If X is discrete, and the values $\{x_n\}$ of X can be arranged so that $x_n < x_{n+1}$ for all n , then the distribution function F is a step function with a discontinuity at each x_n , of magnitude $p_n = P\{X = x_n\}$; F is constant between the x_n and takes the upper value at each discontinuity. To see this, observe that if $x_{n-1} < a < x_n \leq b < x_{n+1}$, then $F(b) - F(a) = P\{a < X \leq b\} = p_n$; if $x_n \leq c < d < x_{n+1}$, then $F(d) - F(c) = 0$.

If X is an arbitrary discrete random variable, the properties of X are completely determined by the *probability function* p_X , defined by

$$p_X(x) = P\{X = x\}, \quad x \in R.$$

Explicitly,

$$P_X(B) = \sum_{x \in B} p_X(x).$$

This is a countable sum since p_X is 0 except at the x_n .

Thus a discrete random variable may be specified by giving a countable set $\{x_n, n = 1, 2, \dots\}$ of real numbers and a set of probabilities $\{p_n, n = 1, 2, \dots\}$ ($p_n \geq 0, \sum_n p_n = 1$), where p_n is to serve as $P\{X = x_n\}$. The probability that X belongs to B is found by summing the p_n for those x_n which belong to B .

The random variable X is said to be *absolutely continuous* iff there is a nonnegative real-valued Borel measurable function f on R such that

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in R.$$

We call f the *density* or *density function* of X ; since $F(x) \rightarrow 1$ as $x \rightarrow \infty$, we have $\int_{-\infty}^{\infty} f(x) dx = 1$.

If X is absolutely continuous with density f , it follows that

$$P_X(B) = \int_B f(x) dx \quad \text{for each } B \in \mathcal{B}(R).$$

For the measure μ defined by $\mu(B) = \int_B f(x) dx$, $B \in \mathcal{B}(R)$, satisfies $\mu(a, b] = F(b) - F(a)$, $a < b$. Thus μ is the Lebesgue-Stieltjes measure corresponding to F ; hence $\mu = P_X$.

Thus absolute continuity of X means that $P_X \ll \text{Lebesgue measure}$, or equivalently, by 2.3.1, F_X is an absolutely continuous function.

Any nonnegative Borel measurable function f on R with $\int_{-\infty}^{\infty} f(x) dx = 1$ is the density of some absolutely continuous random variable X . For let $F(x) = \int_{-\infty}^x f(t) dt$; F is clearly increasing, and by 2.3.4, F is absolutely continuous, hence continuous, on R . Since $F(\infty) = 1$, $F(-\infty) = 0$, F is the distribution function of some random variable X , and X must have density f .

The random variable X is said to be *continuous* iff its distribution function F is continuous on all of R . Equivalently, X is continuous iff $P\{X = x\} = 0$ for all x [see (5) of 1.4.5].

We have seen that absolute continuity implies continuity. If F is the Cantor function (Problem 3, Section 2.3), extended so that $F(x) = 1$ for $x \geq 1$, and $F(x) = 0$ for $x < 0$, a random variable with distribution function F is continuous but not absolutely continuous.

Some typical examples of density functions are:

- (1) Uniform density on $[a, b]$:

$$f(x) = \begin{cases} (b-a)^{-1}, & a \leq x \leq b, \\ 0, & \text{elsewhere.} \end{cases}$$

- (2) Exponential density:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\lambda > 0$.

- (3) Two-sided exponential density:

$$f(x) = \frac{1}{2} \lambda e^{-\lambda|x|}, \quad \lambda > 0.$$

- (4) Normal density:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \quad \sigma > 0, \quad m \text{ real.}$$

- (5) Cauchy density:

$$f(x) = \frac{\theta}{\pi(x^2 + \theta^2)}, \quad \theta > 0.$$

Finally, some remarks on terminology. We shall often abbreviate $\{\omega: X(\omega) \in B\}$ by $\{X \in B\}$; note that this set is also $X^{-1}(B)$, the preimage of B under the mapping X . Similarly, $\{\omega: a < X(\omega) \leq b\}$ will be abbreviated by $\{a < X \leq b\}$. The letter \mathcal{B} will always stand for the Borel sets of an appropriate space. Thus $f: (R^n, \mathcal{B}) \rightarrow (R, \mathcal{B})$ means that $f^{-1}(B) \in \mathcal{B}(R^n)$ for each $B \in \mathcal{B}(R)$. The phrase “almost surely,” abbreviated a.s., is often used in the literature. It means “almost everywhere” with respect to a specified probability measure.

5.7 Random Vectors

We now consider situations involving more than one random variable associated with the same experiment.

An n -dimensional *random vector* on a probability space (Ω, \mathcal{F}, P) is a Borel measurable map from Ω to R^n .

If $X: \Omega \rightarrow R^n$ and $X_i = p_i \circ X$, where p_i is the projection of R^n onto the i th coordinate space, then X is Borel measurable iff each X_i is Borel measurable (see 1.5.8). Thus a random vector may be regarded as an n -tuple (X_1, \dots, X_n) of random variables.

Much of the development of the previous section carries over. As before, the *probability measure induced by the random vector X* is defined by

$$P_X(B) = P\{\omega: X(\omega) \in B\}, \quad B \in \mathcal{B}(R^n).$$

The *distribution function* of X is the function $F = F_X$ from R^n to $[0, 1]$ defined by

$$F(x) = P_X(-\infty, x] = P\{\omega: X_i(\omega) \leq x_i, \quad i = 1, \dots, n\};$$

F is also called the *joint distribution function* of X_1, \dots, X_n . F is increasing and right-continuous on R^n , and P_X is the Lebesgue–Stieltjes measure determined by F (see Section 1.4).

By 1.2.7, we have

$$F(x_1, \dots, x_n) \rightarrow \begin{cases} 1 & \text{as } x_i \uparrow \infty \text{ for all } i; \\ 0 & \text{as } x_i \downarrow -\infty \text{ for any particular } i \\ & \text{(with all other coordinates fixed).} \end{cases} \quad (1)$$

If F is a distribution function on R^n that satisfies (1), then F is the distribution function of some random vector X , and the underlying probability space can be constructed in a canonical way. Take $\Omega = R^n$, $\mathcal{F} = \mathcal{B}(R^n)$, with P the Lebesgue–Stieltjes measure determined by F , and X the identity function on Ω . Since $P_X(B) = P\{\omega: X(\omega) \in B\} = P(B)$, X has induced probability measure P . Furthermore, the distribution function of X is

$$\begin{aligned} F_X(x) &= P_X(-\infty, x] = P(-\infty, x] \\ &= \lim_{a \downarrow -\infty} P(a, x] \quad \text{by 1.2.7} \\ &= \lim_{a \downarrow -\infty} F(a, x] \quad \text{since } P \text{ is the Lebesgue–Stieltjes} \\ &\quad \text{measure determined by } F \\ &= F(x) \quad \text{by 1.4.8(b) and the hypothesis} \\ &\quad \text{that } F(x) \rightarrow 0 \text{ if any } x_i \downarrow -\infty. \end{aligned}$$

[The hypothesis that $F(x) \rightarrow 1$ as $x \uparrow \infty$ implies that P is actually a probability measure.]

The random vector X is said to be *discrete* iff the set of values of X is finite or countably infinite, or equivalently, iff the component random variables X_1, \dots, X_n are all discrete. In this case, the properties of X are

determined by the *probability function* p , given by $p(x) = P\{X = x\}$; explicitly, $P\{X \in B\} = \sum_{x \in B} p(x)$.

The random vector X is said to be *absolutely continuous* iff there is a nonnegative Borel measurable function f on R^n such that

$$F(x) = \int_{(-\infty, x]} f(t) dt, \quad x \in R^n.$$

It follows that

$$P_X(B) = \int_B f(x) dx \quad \text{for all } B \in \mathcal{B}(R^n).$$

For the measure μ defined by $\mu(B) = \int_B f(x) dx$, $B \in \mathcal{B}(R^n)$, satisfies $\mu(a, b] = \int_{(a, b]} f(x) dx = F(b) - F(a)$ [see 1.4.10(b)]. Thus μ is the Lebesgue-Stieltjes measure determined by F ; hence $\mu = P_X$.

Just as in the previous section, any nonnegative Borel measurable function f on R^n with $\int_{R^n} f(x) dx = 1$ is the density of some absolutely continuous random variable X .

5.8 Independent Random Variables

We have talked previously about independence of events; we now consider independent random variables. Intuitively, independence of X_1, \dots, X_n means that a statement about one or more of the X_i does not affect the odds concerning the remaining X_i . Now a statement about X_i corresponds to an event of the form $A_i = \{X_i \in B_i\}$; thus the events A_1, \dots, A_n will be independent. The formal definition is as follows.

5.8.1 Definition. Let X_1, \dots, X_n be random variables on (Ω, \mathcal{F}, P) . The X_1, \dots, X_n are said to be *independent* iff for all sets $B_1, \dots, B_n \in \mathcal{B}(R)$, we have

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\}.$$

By 2.6.8(b), independence of X_1, \dots, X_n may be expressed by saying that if $X = (X_1, \dots, X_n)$, then P_X is the product of the P_{X_i} , $i = 1, \dots, n$.

5.8.2 Comments. (a) If X_1, \dots, X_n are independent, so are X_1, \dots, X_k for $k < n$. To see this, let $B_1, \dots, B_k \in \mathcal{B}(R)$. Then

$$\begin{aligned} P\{X_1 \in B_1, \dots, X_k \in B_k\} &= P\{X_1 \in B_1, \dots, X_k \in B_k, X_{k+1} \in R, \dots, X_n \in R\} \\ &= P\{X_1 \in B_1\} \cdots P\{X_k \in B_k\}. \end{aligned}$$

Thus it is not necessary to check all subfamilies of the collection of events $\{X_i \in B_i\}$, $i = 1, \dots, n$, as in the definition of independent events in Section 5.3.

(b) Independence of extended random variables is defined exactly as above, with $\mathcal{B}(\bar{R})$ replacing $\mathcal{B}(R)$. In fact, suppose that each X_i is a *random object*; that is, a map $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$, where Ω_i is an arbitrary set and \mathcal{F}_i is a σ -field of subsets of Ω_i . Then the X_1, \dots, X_n are said to be independent iff for all sets $B_1 \in \mathcal{F}_1, \dots, B_n \in \mathcal{F}_n$ we have

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\}.$$

(c) Let X_i , $i \in I$ (an arbitrary index set) be an arbitrary family of random objects. The X_i are said to be independent iff X_{i_1}, \dots, X_{i_n} are independent for each finite set $\{i_1, \dots, i_n\}$ of distinct indices in I .

(d) If the $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$ are independent random objects, and $g_i : (\Omega_i, \mathcal{F}_i) \rightarrow (\Omega'_i, \mathcal{F}'_i)$, then the random objects $g_i \circ X_i$, $i \in I$, are independent. ("Functions of independent random objects are independent.") This follows from the definition of independence and the fact that

$$\{g_i \circ X_i \in B_i\} = \{X_i \in g_i^{-1}(B_i)\}.$$

Independence of random variables may be characterized in terms of distribution functions as follows.

5.8.3 Theorem. Let X_1, \dots, X_n be random variables on (Ω, \mathcal{F}, P) . Let F_i be the distribution function of X_i , $i = 1, \dots, n$, and F the distribution function of $X = (X_1, \dots, X_n)$. Then X_1, \dots, X_n are independent iff

$$F(x_1, x_2, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n) \quad \text{for all real } x_1, \dots, x_n.$$

PROOF. If X_1, \dots, X_n are independent, then

$$F(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \prod_{i=1}^n P\{X_i \leq x_i\} = \prod_{i=1}^n F_i(x_i).$$

Conversely, assume $F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i)$ for all x_1, \dots, x_n . Then

$$P_X(a, b] = F(a, b] = \prod_{i=1}^n [F_i(b_i) - F_i(a_i)] = \prod_{i=1}^n P_{X_i}(a_i, b_i]$$

[see 1.4.10(a)]. Thus

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\} \tag{1}$$

when the B_i are right-semiclosed intervals of reals. Now fix the intervals B_2, \dots, B_n . The collection \mathcal{C} of sets $B_1 \in \mathcal{B}(R)$ for which (1) holds is a monotone class including the field of finite disjoint unions of right-semiclosed intervals, and therefore $\mathcal{C} = \mathcal{B}(R)$ by the monotone class theorem. Applying

the same reasoning to each coordinate in turn, we obtain the independence of the X_i . (Explicitly, we prove by induction that if B_1, \dots, B_i are arbitrary Borel sets and B_{i+1}, \dots, B_n are right-semiclosed intervals, then (1) holds for B_1, \dots, B_n .) ■

We may also characterize independence in terms of densities.

5.8.4 Theorem. If $X = (X_1, \dots, X_n)$ has a density f , then each X_i has a density f_i . Furthermore, in this case X_1, \dots, X_n are independent iff $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ for all (x_1, \dots, x_n) except possibly for a Borel subset of R^n with Lebesgue measure zero.

PROOF.

$$\begin{aligned} F_1(x_1) &= P\{X_1 \leq x_1\} = P\{X_1 \leq x_1, X_2 \in R, \dots, X_n \in R\} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \cdots dt_n. \end{aligned}$$

By definition of absolute continuity, X_1 has a density given by

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \cdots dx_n$$

(in other words, we integrate out the unwanted variables). Borel measurability of f_1 follows from Fubini's theorem. Similarly, each X_i has a density f_i , obtained by integrating out all variables except x_i .

Now if $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ a.e., then

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n = F_1(x_1) \cdots F_n(x_n),$$

so by 5.8.3, the X_i are independent. Conversely, if the X_i are independent, then

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(t_1) \cdots f_n(t_n) dt_1 \cdots dt_n. \quad (1)$$

Thus (see the end of Section 5.7) if $g(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$, then

$$P_X(B) = \int_B g(x) dx, \quad B \in \mathcal{B}(R^n).$$

But $P_X(B) = \int_B f(x) dx$, and it follows that $f = g$ a.e. (Lebesgue measure) by 1.6.11. ■

5.8.5 Corollary. If X_1, \dots, X_n are independent and X_i has density f_i , $i = 1, \dots, n$, then X has a density f given by $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$.

PROOF. Equation (1) of 5.8.4 applies. ■

If X_1, \dots, X_n each have a density, it does not follow that (X_1, \dots, X_n) has a density; thus 5.8.5 is false without the independence hypothesis (see Problem 1).

Problems

1. Give an example of random variables X and Y (on the same probability space) such that X and Y each have densities, but (X, Y) does not.
 2. Give an example to show that even if (X, Y) has a density, it is not determined by the individual densities of X and Y .
 3. Let X_1, \dots, X_n be discrete random variables. Show that the X_i are independent iff
- $$P\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^n P\{X_i = x_i\} \quad \text{for all real } x_1, \dots, x_n.$$
4. Let (Ω, \mathcal{F}, P) be a probability space. The classes \mathcal{C}_i , $i \in I$, of sets in \mathcal{F} , are said to be *independent* iff given any choice of $C_i \in \mathcal{C}_i$, $i \in I$, the events C_i are independent. (Thus the random objects $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega'_i, \mathcal{F}'_i)$ are independent iff the classes $X_i^{-1}(\mathcal{F}'_i)$ are independent.) If the \mathcal{C}_i , $i \in I$, are independent, show that if the following sets are added to each \mathcal{C}_i , the enlarged classes still remain independent.
 - (a) Proper differences $A - B$, $A, B \in \mathcal{C}_i$, $B \subset A$.
 - (b) The sets \emptyset, Ω .
 - (c) Countable disjoint unions of sets in \mathcal{C}_i .
 - (d) Limits of monotone sequences in \mathcal{C}_i .

Give an example to show that finite intersections *cannot* be added.

Show that if the \mathcal{C}_i are independent classes, each closed under finite intersection, the minimal σ -fields over the \mathcal{C}_i are also independent.

5.9 Some Examples from Basic Probability

In this section we give a few illustrations to show how some of the computations done in elementary probability courses fit in with the present measure-theoretic framework.

5.9.1 Example. Two numbers X and Y are picked at random between 0 and 1. Assume that X and Y are independent and that each is uniformly distributed (that is, X and Y have densities f_1 and f_2 given $f_1(x) = f_2(x) = 1$, $0 \leq x \leq 1$, and 0 elsewhere). Let Z be the product XY , and let us find the distribution function of Z .

We take $\Omega = \mathbb{R}^2$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^2)$, $X(x, y) = x$, $Y(x, y) = y$. By 5.8.5, (X, Y) must have density $f(x, y) = f_1(x)f_2(y)$; hence

$$P\{(X, Y) \in B\} = \iint_B f(x, y) dx dy.$$

Thus we take our probability measure to be

$$P(B) = \iint_B f_1(x)f_2(y) dx dy, \quad B \in \mathcal{B}(\mathbb{R}^2).$$

Now

$$F(z) = P\{Z \leq z\} = P\{(x, y): xy \leq z\} = \iint_{xy \leq z} f_1(x)f_2(y) dx dy.$$

Since X and Y are between 0 and 1 (with probability 1), $F(z) = 1$ for $z \geq 1$ and $F(z) = 0$ for $z \leq 0$. Since $f_1(x)f_2(y)$ is 1 for $0 \leq x \leq 1$, $0 \leq y \leq 1$, and 0 elsewhere,

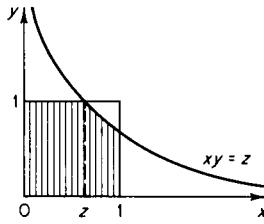


Figure 5.1.

where, $F(z)$ ($0 < z < 1$) is the shaded area in Fig. 5.1.; that is, $z + \int_z^1 (z/x) dx = z - z \ln z$. Thus Z has a density

$$f(z) = F'(z) = \begin{cases} -\ln z, & 0 < z < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Note that although f is unbounded, its *integral*, namely F , is always finite.

5.9.2 Example. Let X , Y , and Z be independent random variables, each normally distributed with $m = 0$, $\sigma = 1$; that is, X , Y , and Z each have the normal density

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-m)^2}{2\sigma^2}\right] \quad \text{with } m = 0, \quad \sigma = 1.$$

Let $W = (X^2 + Y^2 + Z^2)^{1/2}$ (take the positive square root so that $W \geq 0$). Find the distribution function of W .

We take $\Omega = R^3$, $\mathcal{F} = \mathcal{B}(R^3)$, $X(x, y, z) = x$, $Y(x, y, z) = y$, $Z(x, y, z) = z$, and

$$P(B) = \iiint_B f(x, y, z) dx dy dz, \quad B \in \mathcal{B}(R^3),$$

where

$$\begin{aligned} f(x, y, z) &= f_1(x)f_2(y)f_3(z) = g(x)g(y)g(z) \\ &= (2\pi)^{-3/2} \exp[-\frac{1}{2}(x^2 + y^2 + z^2)]. \end{aligned}$$

Thus

$$F(w) = P\{W \leq w\} = P\{X^2 + Y^2 + Z^2 \leq w^2\}$$

If $w \geq 0$,

$$F(w) = \iiint_{x^2+y^2+z^2 \leq w^2} (2\pi)^{-3/2} \exp[-\frac{1}{2}(x^2 + y^2 + z^2)] dx dy dz$$

or in spherical coordinates,

$$\begin{aligned} F(w) &= \int_0^{2\pi} d\theta \int_0^\pi d\phi \int_0^w (2\pi)^{-3/2} \exp[-\frac{1}{2}r^2] r^2 \sin \phi dr \\ &= (2\pi)^{-3/2} (2\pi) (2) \int_0^w r^2 \exp[-\frac{1}{2}r^2] dr. \end{aligned}$$

Thus W is absolutely continuous, with density

$$f(w) = \begin{cases} \frac{2}{\sqrt{2\pi}} w^2 \exp[-\frac{1}{2}w^2], & w \geq 0, \\ 0, & w < 0. \end{cases}$$

5.9.3 Example. Let X_1, \dots, X_n be independent random variables, each with density f and distribution function F ; that is, $\Omega = R^n$, $\mathcal{F} = \mathcal{B}(R^n)$, $X_i(x_1, \dots, x_n) = x_i$, $1 \leq i \leq n$,

$$P(B) = \int_B f(x_1) \cdots f(x_n) dx_1 \cdots dx_n, \quad B \in \mathcal{B}(R^n).$$

Let T_k be the k th smallest of the X_i ; for example, if $n = 4$, $X_1(\omega) = 2$, $X_2(\omega) = 1.4$, $X_3(\omega) = -7$, $X_4(\omega) = 8$, then $T_1(\omega) = \min_i X_i(\omega) = X_3(\omega) = -7$, $T_2(\omega) = X_2(\omega) = 1.4$, $T_3(\omega) = X_1(\omega) = 2$, $T_4(\omega) = \max_i X_i(\omega) = X_4(\omega) = 8$. [Note that

$$P\{X_i = X_j\} = \iint_{x_i=x_j} f(x_i)f(x_j) dx_i dx_j = 0 \quad \text{for } i \neq j,$$

and therefore

$$P\{X_i = X_j \text{ for at least one } i \neq j\} \leq \sum_{i \neq j} P\{X_i = X_j\} = 0.$$

Thus ties occur with probability 0 and can be ignored.]

Find the individual distribution functions of the T_k , and the joint distribution function of (T_1, \dots, T_n) .

Now

$$P\{T_k \leq x\} = \sum_{i=1}^n P\{T_k \leq x, T_k = X_i\} \quad \text{by 5.5.2} \quad (1)$$

and, for example,

$$\begin{aligned} P\{T_k \leq x, T_k = X_1\} &= P\{X_1 \leq x, \text{ exactly } k-1 \text{ of the random} \\ &\quad \text{variables } X_2, \dots, X_n \text{ are less than } X_1 \\ &\quad \text{and the remaining } n-k \text{ random variables} \\ &\quad \text{are greater than } X_1\}. \end{aligned} \quad (2)$$

But, using Fubini's theorem,

$$\begin{aligned} P\{X_1 \leq x, X_2 < X_1, \dots, X_k < X_1, X_{k+1} > X_1, \dots, X_n > X_1\} \\ &= \int_{x_1=-\infty}^x \int_{x_2=-\infty}^{x_1} \cdots \int_{x_k=-\infty}^{x_1} \int_{x_{k+1}=x_1}^{\infty} \cdots \int_{x_n=x_1}^{\infty} \\ &\quad f(x_1) \cdots f(x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^x f(x_1)(F(x_1))^{k-1}(1-F(x_1))^{n-k} dx_1. \end{aligned} \quad (3)$$

Now by symmetry, (2) is the sum of $\binom{n-1}{k-1}$ terms, each of which has the same value as (3), since we may select the $k-1$ random variables to be less than X_1 in $\binom{n-1}{k-1}$ ways. Also, each term in the summation (1) has the same value as (2). Thus

$$P\{T_k \leq x\} = \int_{-\infty}^x n \binom{n-1}{k-1} f(x_1)(F(x_1))^{k-1}(1-F(x_1))^{n-k} dx_1$$

so that T_k is absolutely continuous, with density

$$f_{T_k}(x) = n \binom{n-1}{k-1} f(x)(F(x))^{k-1}(1-F(x))^{n-k}.$$

We now find the joint distribution function of T_1, \dots, T_n . Let $b_1 < b_2 < \dots < b_n$. Then

$$\begin{aligned} P\{T_1 \leq b_1, \dots, T_n \leq b_n\} \\ = n! P\{T_1 \leq b_1, \dots, T_n \leq b_n, X_1 < X_2 < \dots < X_n\} &\quad \text{by symmetry} \\ = n! P\{X_1 \leq b_1, X_1 < X_2 \leq b_2, X_2 < X_3 \leq b_3, \dots, X_{n-1} < X_n \leq b_n\} \\ = n! \int_{-\infty}^{b_1} f(x_1) dx_1 \int_{x_1}^{b_2} f(x_2) dx_2 \cdots \int_{x_{n-1}}^{b_n} f(x_n) dx_n \\ = \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} g(x_1, \dots, x_n) dx_1 \cdots dx_n, \end{aligned}$$

where

$$g(x_1, \dots, x_n) = \begin{cases} n! f(x_1) \cdots f(x_n), & x_1 < x_2 < \cdots < x_n, \\ 0 & \text{elsewhere.} \end{cases}$$

Thus (T_1, \dots, T_n) is absolutely continuous with density g . (Note that f_{T_k} can be found from g (see 5.8.4), but the calculation is not any simpler than the direct method we have used above.)

5.9.4 Example. Let X be an absolutely continuous random variable with density f , assumed to be piecewise-continuous. Let D be a Borel subset of R such that D includes the range of X , and let g be a Borel measurable function from D to R .

If $Y = g \circ X$, we wish to find the distribution of Y . [*Distribution* is a generic term; to say that we know the distribution of Y means that we know how to calculate $P\{Y \in B\}$ for all Borel sets B . Thus the distribution may be specified by giving the induced probability measure P_Y or the distribution function F_Y . If Y is absolutely continuous, its density is adequate, and if Y is discrete, the probability function suffices. If $Y: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ is an arbitrary random object, the distribution of Y means the probability measure P_Y , defined by $P_Y(B) = P\{Y \in B\}, B \in \mathcal{F}'$.]

Assume that D is an open interval I , and g is either strictly increasing or strictly decreasing, with inverse h . Assume also that g has a continuous nonzero derivative (hence so does h). We show that Y is absolutely continuous with density

$$f_2(y) = \begin{cases} f_1(h(y)) |h'(y)|, & y \in g(I), \\ 0 & \text{elsewhere.} \end{cases}$$

We compute, for $y \in g(I)$,

$$\begin{aligned} F_2(y) &= P\{Y \leq y\} = P\{\omega : g(X(\omega)) \leq y\} \\ &= P\{X \leq h(y)\} \quad \text{if } g \text{ is increasing} \\ &= P\{X \geq h(y)\} \quad \text{if } g \text{ is decreasing} \end{aligned}$$

(see Fig. 5.2).

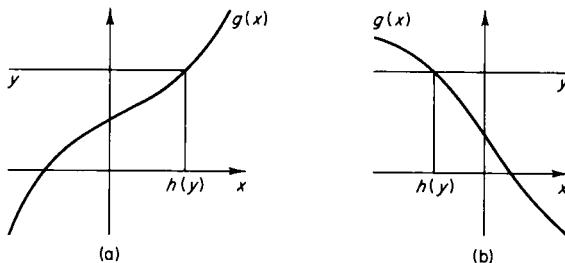


Figure 5.2. (a) g strictly increasing, (b) g strictly decreasing.

Thus

$$F_2(y) = \begin{cases} \int_{-\infty}^{h(y)} f_1(x) dx & \text{if } g \text{ is increasing,} \\ \int_{h(y)}^{\infty} f_1(x) dx & \text{if } g \text{ is decreasing.} \end{cases}$$

Therefore

$$\frac{dF_2(y)}{dy} = \begin{cases} f_1(h(y))h'(y) & \text{if } g \text{ is increasing,} \\ f_1(h(y))(-h'(y)) & \text{if } g \text{ is decreasing,} \\ f_1(h(y))|h'(y)| & \text{in either case.} \end{cases}$$

Now F_2 is continuous everywhere, and has a piecewise-continuous derivative; it follows that F_2 is the integral of its derivative:

$$F_2(y) = \int_{-\infty}^y \frac{dF_2(z)}{dz} dz$$

(apply the fundamental theorem of the calculus). Thus Y is absolutely continuous with density $f_1(h(y))|h'(y)|$.

Examples of this type in which g is more complicated will be considered in the problems.

Problems

1. Consider Example 5.9.4, but weaken the hypothesis so that the domain of g is the union of closed intervals I_1, \dots, I_n , such that on the interior of each I_j , g has a continuous nonzero derivative and is either strictly increasing or strictly decreasing, with inverse h_j . Show that Y is absolutely continuous with density

$$f_2(y) = \sum_{j=1}^n f_1(h_j(y))|h_j'(y)|,$$

where $f_1(h_j(y))|h_j'(y)|$ is interpreted as 0 if y does not belong to the domain of h_j .

2. Let X be an absolutely continuous random variable with density $f_1(x) = x^3/64$, $0 \leq x \leq 4$; $f_1(x) = 0$ elsewhere. Define a random variable Y by $Y = \min(\sqrt{X}, 2 - \sqrt{X})$. Find the density of Y .
3. Let X , Y , and Z be independent random variables, each uniformly distributed between 0 and 1. Find the probability that $Z^2 \leq XY$.
4. Let X be a random n -vector with density f_1 , and let $Y = g \circ X$, where $g: R^n \rightarrow R^n$ (or $g: D \rightarrow R^n$, where D is open in R^n and $P\{X \in D\} = 1$). Assume g to be one-to-one and continuously differentiable with a non-zero Jacobian J_g (hence g has a continuously differentiable inverse h). Show that Y is absolutely continuous with density

$$f_2(y) = f_1(h(y))|J_h(y)| = \frac{f_1(h(y))}{|J_g(x)|_{x=h(y)}}.$$

5. Let X and Y be independent random variables, each normally distributed with $m = 0$ and the same σ . Define random variables R and Θ by $X = R \cos \Theta$, $Y = R \sin \Theta$. Show that R and Θ are independent, and find their density functions (use Problem 4).
6. Let X_1, \dots, X_n be independent random variables, each with density f . Let X_0 be the number of random variables among X_1, \dots, X_n that exceed the smallest of the X_i by more than 2. Find $\int_{\Omega} X_0 dP$ (Leave the answer in the form of an integral on the real line.) Hint: Express X_0 as a sum of indicators.
7. Let $\Omega = [0, 1]$, \mathcal{F} = Borel sets, P = Lebesgue measure. Show that (Ω, \mathcal{F}, P) is a universal probability space in the sense that if F is any proper distribution function on R (“proper” means that $F(\infty) = 1$, $F(-\infty) = 0$), there is a random variable X on (Ω, \mathcal{F}, P) with distribution function F . (Define $F^{-1}(y) = \sup\{x: F(x) < y\}$, $0 < y < 1$, and take $X(\omega) = F^{-1}(\omega)$, with $X(0)$ and $X(1)$ arbitrary.)

5.10 Expectation

Let X be a simple random variable on (Ω, \mathcal{F}, P) , taking the values x_1, \dots, x_n with probabilities p_1, \dots, p_n . If the random experiment is repeated independently N times, N very large, X will take the value x_i roughly Np_i times, so the arithmetic average of the values of X in the N observations is roughly

$$\frac{1}{N} [Np_1x_1 + Np_2x_2 + \dots + Np_nx_n] = \sum_{i=1}^n p_i x_i.$$

This is a reasonable figure for the average value of X . If X is represented as $\sum_{i=1}^n x_i I_{B_i}$, where the B_i are disjoint sets in \mathcal{F} , the average value may be expressed as $\sum_{i=1}^n x_i P(B_i)$, which is $\int_{\Omega} X dP$. Since arbitrary random variables are ultimately built up from simple ones, it is reasonable to take $\int_{\Omega} X dP$ as the definition of the average value (henceforth to be called the “expectation”) of X .

5.10.1 Definition. If X is a random variable on (Ω, \mathcal{F}, P) , the *expectation* of X is defined by

$$E(X) = \int_{\Omega} X dP$$

provided the integral exists. Thus $E(X)$ is the integral of the Borel measurable function X with respect to the probability measure P , so that all the results of integration theory are applicable. The same definition is used if X is an extended random variable.

In many situations it is inconvenient to compute $E(X)$ by integrating over Ω ; the following result expresses $E(X)$ as an integral with respect to the induced probability measure P_X , which in turn is determined by the distribution function F .

First, a word about notation. If F is a distribution function on R^n with corresponding Lebesgue-Stieltjes measure μ , and $g: (R^n, \mathcal{B}) \rightarrow (R, \mathcal{B})$, then $\int_{R^n} g(x) dF(x)$ means $\int_{R^n} g d\mu$; it is *not* a Riemann-Stieltjes integral.

5.10.2 Theorem. Let X be a random variable on (Ω, \mathcal{F}, P) , with distribution function F . Let g be a Borel measurable function from R to R .

If $Y = g \circ X$, then

$$E(Y) = \int_R g(x) dF(x) \quad \left(= \int_R g dP_X \right)$$

in the sense that if either of the two sides exists, so does the other, and the two sides are equal.

PROOF. We use the basic technique of starting with indicators and proceeding to more complicated functions.

Let g be an indicator I_B , $B \in \mathcal{B}(R)$. Then

$$E(Y) = E(I_B \circ X) = E(I_{\{X \in B\}}) = P_X(B) = \int_R g dP_X$$

so that $E(Y)$ and $\int_R g dP_X$ exist and are equal.

Now let g be a nonnegative simple function, say $g(x) = \sum_{j=1}^n x_j I_{B_j}(x)$, the B_j disjoint sets in $\mathcal{B}(R)$. Then

$$\begin{aligned} E(Y) &= \sum_{j=1}^n x_j E(I_{B_j} \circ X) \\ &= \sum_{j=1}^n x_j \int_R I_{B_j} dP_X \quad \text{by what we have just proved} \\ &= \int_R \left(\sum_{j=1}^n x_j I_{B_j} \right) dP_X \quad \text{since } g \geq 0 \\ &= \int_R g dP_X. \end{aligned}$$

Again, both integrals exist and are equal.

If g is a nonnegative Borel measurable function, let g_1, g_2, \dots be non-negative simple functions with $g_n \uparrow g$. We have just proved that

$$E(g_n \circ X) = \int_R g_n dP_X;$$

hence by the monotone convergence theorem,

$$E(g \circ X) = \int_R g dP_X,$$

and again both integrals exist and are equal.

Finally, if $g = g^+ - g^-$ is an arbitrary Borel measurable function and $Y = g \circ X$, we have

$$\begin{aligned} E(Y) &= E(Y^+) - E(Y^-) \\ &= E(g^+ \circ X) - E(g^- \circ X) \\ &= \int_R g^+ dP_X - \int_R g^- dP_X \quad \text{by what we have already proved} \\ &= \int_R g dP_X. \end{aligned}$$

If $E(Y)$ exists and, say, $E(Y^-)$ is finite, then $\int_R g^- dP_X$ is finite, and hence $\int_R g dP_X$ exists; by the same reasoning, the existence of $\int_R g dP_X$ implies that of $E(Y)$. ■

5.10.3 Corollaries and Extensions. (a) Let X be a random vector on (Ω, \mathcal{F}, P) , and let g be a Borel measurable function from R^n to R . Then $E(g \circ X) = \int_{R^n} g(x) dF(x)$ in the sense that if either integral exists, so does the other, and the two are equal.

The proof is exactly as in 5.10.2, with R replaced by R^n .

(b) More generally, let X be a random object on (Ω, \mathcal{F}, P) , that is, $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, where (Ω', \mathcal{F}') is an arbitrary measurable space. Let g be a Borel measurable real (or extended real) valued function on (Ω', \mathcal{F}') . Let P_X be the probability measure induced by X :

$$P_X(B) = P\{\omega: X(\omega) \in B\}, \quad B \in \mathcal{F}'.$$

Then

$$E(g \circ X) = \int_{\Omega'} g dP_X$$

in the sense that if either integral exists, so does the other, and the two are equal.

Again, the proof is just as in 5.10.2, with R replaced by Ω' and $\mathcal{B}(R)$ by \mathcal{F}' .

(c) If X is a random variable (or random vector) with density f , then

$$\int g(x) dF(x) = \int g(x) f(x) dx$$

(integration over R in the case of a random variable, and over R^n in the case of a random vector) in the sense that if either integral exists, then so does the other, and the two are equal.

When g is an indicator I_B , this says that $P_X(B) = \int_B f(x) dx$, which holds for any Borel set B . The proof is completed by passing in turn to nonnegative simple functions, nonnegative Borel measurable functions, and arbitrary Borel measurable functions.

(d) If X is a discrete random variable with probability function p , then

$$\int g(x) dF(x) = \sum_x g(x)p(x),$$

where the series is interpreted as $\sum_x g^+(x)p(x) - \sum_x g^-(x)p(x)$, and again the interpretation is that the integral exists iff the sum exists (that is,

$$\sum_x g^+(x)p(x) < \infty \quad \text{or} \quad \sum_x g^-(x)p(x) < \infty),$$

and in this case the two are equal.

This is proved by starting with indicators as before.

Expectations of certain functions of X are of special interest.

5.10.4 Definitions. Let X be a random variable on (Ω, \mathcal{F}, P) . If $k > 0$, the number $E(X^k)$ is called the k th moment of X ; $E[|X|^k]$ is called the k th *absolute moment* of X . $E[(X - E(X))^k]$ is called the k th *central moment*, $E[|X - E(X)|^k]$ the k th *absolute central moment*; central moments are defined only when $E(X)$ is finite.

The first moment ($k = 1$) is $E(X)$, sometimes called the *mean* of X , and the first central moment (if it exists) is always 0. The second central moment $\sigma^2 = E[(X - E(X))^2]$ is called the *variance* of X , sometimes written $\text{Var } X$, and the positive square root σ the *standard deviation*.

Note that $E(X^k)$ is finite iff $E[|X|^k]$ is finite, by 1.6.4(b). Also, finiteness of the k th moments implies finiteness of lower moments, as we now prove.

5.10.5 Lemma. If $k > 0$ and $E(X^k)$ is finite, then $E(X^j)$ is finite for $0 < j < k$.

FIRST PROOF.

$$\begin{aligned} E[|X|^j] &= \int_{\Omega} |X|^j dP \\ &= \int_{\{|X|^j < 1\}} |X|^j dP + \int_{\{|X|^j \geq 1\}} |X|^j dP \\ &\leq P\{|X|^j < 1\} + \int_{\Omega} |X|^k dP < \infty. \blacksquare \end{aligned}$$

SECOND PROOF. We have $\|X\|_j \leq \|X\|_k$ for $0 < j < k$ (Problem 8, Section 2.4). \blacksquare

Central moments of integral order can be obtained from moments, as follows.

5.10.6 Lemma. If n is a positive integer greater than 1, $E(X^{n-1})$ is finite, and $E(X^n)$ exists, then $E[(X - E(X))^n] = \sum_{k=0}^n \binom{n}{k} [-E(X)]^{n-k} E(X^k)$. In particular, if $E(X)$ is finite [$E(X^2)$ always exists since $X^2 \geq 0$], then

$$\text{Var } X = E(X^2) - [E(X)]^2.$$

PROOF. Use the binomial theorem and the additivity theorem for integrals (1.6.3). \blacksquare

A similar formula expresses moments in terms of central moments. (Write $X^n = (X - E(X) + E(X))^n$ and use the binomial theorem.)

We now restate a result proved earlier in a measure-theoretic context.

5.10.7 Chebyshev's Inequality. (a) If X is a nonnegative random variable, $0 < p < \infty$ and $0 < \varepsilon < \infty$,

$$P\{X \geq \varepsilon\} \leq \frac{E(X^p)}{\varepsilon^p}.$$

(b) If X is a random variable with finite mean m and variance σ^2 , and $0 < k < \infty$,

$$P\{|X - m| \geq k\sigma\} \leq \frac{1}{k^2}.$$

This is a quantitative result to the effect that a random variable with small variance is likely to be close to its mean.

PROOF. See 2.4.9. ■

A normally distributed random variable has the useful property that the distribution is completely determined by the mean and variance. Specifically, if X has the normal density, that is,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right],$$

then $m = E(X)$ and $\sigma^2 = \text{Var } X$; the computation is straightforward, using the standard integrals

$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi} \quad \text{and} \quad \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx = \frac{1}{2}\sqrt{\pi}.$$

The phrase “normal (m, σ^2) ” is used for a random variable that is normally distributed with mean m and variance σ^2 .

The following result on the expectation of a product of independent random variables is a direct consequence of Fubini's theorem.

5.10.8 Theorem. Let X_1, \dots, X_n be independent random variables on (Ω, \mathcal{F}, P) . If all X_i are nonnegative or if $E(X_i)$ is finite for all i , then $E(X_1 \cdots X_n)$ exists and equals $E(X_1)E(X_2) \cdots E(X_n)$.

PROOF. If all $X_i \geq 0$, then by 5.10.3(a),

$$E(X_1 \cdots X_n) = \int_R x_1 \cdots x_n dP_X(x_1, \dots, x_n) \quad \text{where } X = (X_1, \dots, X_n).$$

Since P_X is the product of the P_{X_i} (see 5.8.1), Fubini's theorem yields

$$E(X_1 \cdots X_n) = \int_R x_1 dP_{X_1}(x_1) \cdots \int_R x_n dP_{X_n}(x_n) = E(X_1) \cdots E(X_n).$$

(This can also be proved without Fubini's theorem by starting with indicators and proceeding to nonnegative simple functions and then nonnegative measurable functions, but the present proof is faster. Note also that the result holds for extended random variables, with the same proof.)

If all $E(X_i)$ are finite, the above argument shows that

$$E(|X_1 \cdots X_n|) = \prod_{i=1}^n E(|X_i|) < \infty,$$

and thus Fubini's theorem may be applied just as in the first part of the proof. ■

Theorem 5.10.8 can be extended to complex-valued random variables. Recall from Section 2.4 that a complex-valued random variable X on (Ω, \mathcal{F}, P) is a Borel measurable map from Ω to the complex plane, that is, X is simply a two-dimensional random vector. If $X_1 = \operatorname{Re} X$ and $X_2 = \operatorname{Im} X$, we define $E(X) = E(X_1) + iE(X_2)$ provided $E(X_1)$ and $E(X_2)$ are both finite.

5.10.9 Theorem. If X_1, \dots, X_n are independent complex-valued random variables and $E(X_i)$ is finite for all i , then $E(X_1 \cdots X_n)$ is finite and equals $E(X_1) \cdots E(X_n)$.

PROOF. First let $n = 2$, $X_1 = Y_1 + iZ_1$, $X_2 = Y_2 + iZ_2$. By 5.8.2(d), Y_1 and Y_2 are independent, as are Z_1 and Z_2 , Z_1 and Y_2 , and Z_1 and Z_2 . By 5.10.8,

$$\begin{aligned} E(X_1 X_2) &= E(Y_1)E(Y_2) - E(Z_1)E(Z_2) + iE(Y_1)E(Z_2) + iE(Z_1)E(Y_2) \\ &= (E(Y_1) + iE(Z_1))(E(Y_2) + iE(Z_2)) \\ &= E(X_1)E(X_2). \end{aligned}$$

Now let $n > 2$, $X_j = Y_j + iZ_j$, $j = 1, \dots, n$, and assume the result has been established for $n - 1$ random variables. If $V = (Y_1, Z_1, Y_2, Z_2, \dots, Y_{n-1}, Z_{n-1})$ and $W = (Y_n, Z_n)$, we claim that V and W are independent. For by independence of X_1, \dots, X_n , $P\{V \in A, W \in B\} = P\{V \in A\}P\{W \in B\}$ when A and B are measurable rectangles. Pass from measurable rectangles to finite disjoint unions of measurable rectangles, and then, by means of the monotone class theorem, to arbitrary Borel sets.

The independence of V and W implies, by 5.8.2(d), that $X_1 \cdots X_{n-1}$ and X_n are independent, so that $E(X_1 \cdots X_n) = E(X_1 \cdots X_{n-1})E(X_n) = E(X_1) \cdots E(X_n)$ by the induction hypothesis. ■

Theorem 5.10.8 implies that the variance of a sum of independent random variables is the sum of the variances. Actually, a somewhat more general result may be derived. We first introduce some new terminology.

5.10.10 Definitions and Comments. Let X and Y be random variables with finite expectation, and assume $E(XY)$ is also finite. (In particular, if X and Y have finite second moments, $E(XY)$ is finite by the Cauchy–Schwarz inequality.) The *covariance* of X and Y is defined by

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

If X and Y are independent, then $\text{Cov}(X, Y) = 0$ by 5.10.8; however, the converse is not true (consider $X = \cos \theta$, $Y = \sin \theta$, where θ is uniformly distributed between 0 and 2π).

If the variances σ_X^2 and σ_Y^2 are finite and greater than 0, the *correlation coefficient* between X and Y is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

By the Cauchy–Schwarz inequality applied to $X - E(X)$ and $Y - E(Y)$, $-1 \leq \rho(X, Y) \leq 1$. Furthermore (see Problem 6, Section 2.4), $|\rho(X, Y)| = 1$ iff $X' = X - E(X)$ and $Y' = Y - E(Y)$ are linearly dependent, that is, $P\{aX' + bY' = 0\} = 1$ for some real numbers a and b , not both 0.

We now look at the variance of a sum.

5.10.11 Theorem. If X_1, \dots, X_n are random variables with finite expectation, and $E(X_i X_j)$ is finite for all i, j with $i \neq j$,

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var } X_i + 2 \sum_{\substack{i,j=1 \\ i < j}}^n \text{Cov}(X_i, X_j).$$

Thus if X_1, \dots, X_n are mutually uncorrelated, that is, $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$; in particular, if X_1, \dots, X_n are independent, then

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var } X_i.$$

PROOF.

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= E[(X_1 + \dots + X_n - E(X_1) - \dots - E(X_n))^2] \\ &= \sum_{i=1}^n E[(X_i - E(X_i))^2] + 2 \sum_{\substack{i,j=1 \\ i < j}}^n E[(X_i - E(X_i))(X_j - E(X_j))]. \quad \blacksquare \end{aligned}$$

5.10.12 Corollary. Under the hypothesis of 5.10.11, if a_1, \dots, a_n, b are arbitrary real numbers,

$$\text{Var}(a_1 X_1 + \cdots + a_n X_n + b) = \sum_{i=1}^n a_i^2 \text{Var } X_i + 2 \sum_{\substack{i,j=1 \\ i < j}}^n a_i a_j \text{Cov}(X_i, X_j).$$

PROOF. This follows from 5.10.11, along with the observations, which may be verified from the definitions, that $\text{Var}(aX + b) = a^2 \text{Var } X$ and $\text{Cov}(a_i X_i, a_j X_j) = a_i a_j \text{Cov}(X_i, X_j)$. ■

Problems

- Let X be a discrete random variable, with $P\{X = n\} = (\frac{1}{2})^n$, $n = 1, 2, \dots$. Let $Y = g \circ X$, where $g(n) = (-1)^{n+1} 2^n/n$. Show that $E(Y)$ does not exist, although the series $\sum_{n=1}^{\infty} g(n) P\{X = n\}$ is conditionally convergent.
- Let X be a random variable with the distribution function shown in Fig. P5.1. Compute $E(X^2)$.

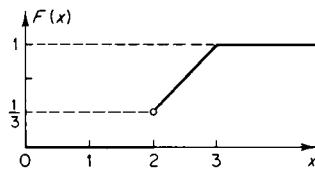


Figure P5.1.

- Suppose X is a random variable with distribution function F_X , and $Y = g \circ X$, $g : R \rightarrow R$, Borel measurable. It is desired to find the expectation of Y . One student evaluates $\int_{-\infty}^{\infty} g(x) dF_X(x)$; another first finds the distribution function of Y , that is,

$$F_Y(y) = P\{Y \leq y\} = P\{X \in g^{-1}(-\infty, y]\}$$

and then evaluates $\int_{-\infty}^{\infty} y dF_Y(y)$. Will the answers be the same?

- Show that the random variables X_1, \dots, X_n are independent iff for all Borel measurable $g_i : R \rightarrow R$ such that $g_i \geq 0$, we have

$$E[g_1(X_1) \cdots g_n(X_n)] = \prod_{i=1}^n E[g_i(X_i)], \quad \text{where } g_i(X_i) = g_i \circ X_i. \quad (1)$$

Also show that X_1, \dots, X_n are independent iff (1) holds for all Borel measurable $g_i : R \rightarrow R$ such that $E[g_i(X_i)]$ is finite for each i .

- Let X be an extended random variable on (Ω, \mathcal{F}, P) and let g be a Borel measurable function from \bar{R} to R . Define $F(x) = P\{-\infty < X \leq x\}$, $x \in R$; thus $F(\infty) - F(-\infty) = P\{X \text{ finite}\} \leq 1$. Show that if $E[g \circ X]$ exists, $E[g \circ X] = \int_{-\infty}^{\infty} g(x) dF(x) + g(\infty)P\{X = \infty\} + g(-\infty)P\{X = -\infty\}$.

5.11 Infinite Sequences of Random Variables

Very often we shall be interested in a sequence of random variables X_1, X_2, \dots, X_n , where n is arbitrarily large and not fixed in advance. Thus it is convenient to have a single probability space on which we can define an infinite sequence of random variables. The discussion of measures on infinite product spaces (Section 2.7) provides the necessary machinery. In particular, we can require that X_1, X_2, \dots be independent, with X_i having a specified distribution. In fact the X_i can be random objects with values in an arbitrary measurable space.

5.11.1 Theorem. Let $(\Omega_j, \mathcal{F}_j, P_j)$, $j = 1, 2, \dots$ be an arbitrary sequence of probability spaces. There exists a probability space (Ω, \mathcal{F}, P) and a sequence of independent random objects $X_j: (\Omega, \mathcal{F}) \rightarrow (\Omega_j, \mathcal{F}_j)$ such that

$$P\{X_j \in B\} = P_j(B) \quad \text{for all } B \in \mathcal{F}_j, \quad j = 1, 2, \dots$$

To obtain a sequence of independent random variables with specified distribution functions F_1, F_2, \dots , we take $\Omega_j = R$, $\mathcal{F}_j = \mathcal{B}(R)$, with P_j the Lebesgue–Stieltjes measure corresponding to F_j .

PROOF. Let $\Omega = \prod_{j=1}^{\infty} \Omega_j$, $\mathcal{F} = \prod_{j=1}^{\infty} \mathcal{F}_j$, $P = \prod_{j=1}^{\infty} P_j$ (see 2.7.3). If $\omega = (\omega_1, \omega_2, \dots) \in \Omega$, let $X_j(\omega) = \omega_j$, $j = 1, 2, \dots$. If $B \in \mathcal{F}_j$, then $\{\omega: X_j(\omega) \in B\} = \{\omega: \omega_j \in B\}$, a measurable rectangle. Thus $\{\omega: X_j(\omega) \in B\} \in \mathcal{F}$, so that the X_j are random objects. By 2.7.3,

$$P\{X_1 \in A_1, \dots, X_n \in A_n\} = \prod_{j=1}^n P_j(A_j) \quad \text{if } A_j \in \mathcal{F}_j, \quad 1 \leq j \leq n.$$

Take $A_i = \Omega_i$, $i \neq j$, to conclude that $P\{X_j \in A_j\} = P_j(A_j)$. Therefore $P\{X_1 \in A_1, \dots, X_n \in A_n\} = \prod_{j=1}^n P\{X_j \in A_j\}$, proving independence. ■

The results of Section 2.7 may also be used to provide the underlying probability space for a *Markov chain*. Let S be a finite or countably infinite set, called the *state space*; for convenience we may take S to be a subset of the integers. Let $\Pi = [p_{ij}]$, $i, j \in S$, be a *stochastic matrix*, that is, $p_{ij} \geq 0$ for all $i, j \in S$, and $\sum_j p_{ij} = 1$ for all $i \in S$. Let p_i , $i \in S$ be a set of nonnegative numbers adding to 1 (the *initial distribution*). We envision a process that starts at time $t = 0$ in an initial state X_0 , where $P\{X_0 = i\} = p_i$, $i \in S$, and makes transitions at times $t = 1, 2, \dots$ in accordance with the following rule. If X_n denotes the state at time n , and we know that $X_n = i$, then regardless of the past history, in other words, regardless of how the process happened to

arrive at state i , the probability of moving to state j at time $n + 1$ is p_{ij} . We expect from 5.5.1(b) that for all $i_0, i_1, \dots, i_n \in S$, $n = 0, 1, 2, \dots$,

$$P\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

We now show that it is possible to construct a sequence of random variables satisfying this requirement.

5.11.2 Theorem. For a given state space S , stochastic matrix $\Pi = [p_{ij}]$, $i, j \in S$, and initial distribution p_i , $i \in S$, there is a sequence of random variables X_0, X_1, \dots , all defined on the same probability space and taking values in S , such that

$$P\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \quad (1)$$

for all $i_0, \dots, i_n \in S$ and all $n = 0, 1, \dots$.

PROOF. Let \mathcal{S} consist of all subsets of S , and take $\Omega = S^\infty$, $\mathcal{F} = \mathcal{S}^\infty$. Define

$$\begin{aligned} P_0(B) &= \sum_{i \in B} p_i, \quad B \in \mathcal{S}, \\ P(i_0, \dots, i_{n-1}, B) &= \sum_{j \in B} p_{i_{n-1} j}, \quad B \in \mathcal{S}, \quad i_0, \dots, i_{n-1} \in S. \end{aligned}$$

Since \mathcal{S}^n consists of all subsets of S^n , the measurability requirements of Theorem 2.7.2 are automatically satisfied. If we define $X_n(\omega_0, \omega_1, \dots) = \omega_n$, $n = 0, 1, \dots$, we obtain

$$\begin{aligned} P\{X_0 = i_0, \dots, X_n = i_n\} &= \int_S P_0(d\omega_0) \int_S P(\omega_0, d\omega_1) \\ &\quad \cdots \int_S I_B(\omega_0, \dots, \omega_n) P(\omega_0, \dots, \omega_{n-1}, d\omega_n), \end{aligned}$$

where $B = \{(\omega_0, \dots, \omega_n) : \omega_0 = i_0, \dots, \omega_n = i_n\}$. Now if μ is a measure on \mathcal{S} with $\mu\{j\} = q_j$, $j \in S$, and $f: S \rightarrow R$, then $\int_S f d\mu = \sum_j q_j f(j)$ (see 2.4.12). Thus

$$\begin{aligned} \int_S I_B(\omega_0, \dots, \omega_n) P(\omega_0, \dots, \omega_{n-1}, d\omega_n) \\ &= I_B(\omega_0, \dots, \omega_{n-1}, i_n) P(\omega_0, \dots, \omega_{n-1}, \{i_n\}) \\ &= I_B(\omega_0, \dots, \omega_{n-1}, i_n) p_{\omega_{n-1} i_n}. \end{aligned}$$

We may continue this process to obtain

$$P\{X_0 = i_0, \dots, X_n = i_n\} = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \quad \blacksquare$$

A sequence of random variables $\{X_n\}$ satisfying (1) of 5.11.2 is called a *Markov chain* corresponding to the matrix Π and the initial distribution

$\{p_{ij}\}$; Π is called the *transition matrix* of the chain and the numbers p_{ij} are called *transition probabilities*.

The basic properties of Markov chains (and random walks, an important special case) are discussed by Ash (1970, Chapters 6 and 7).

If X_1, \dots, X_n are random variables on (Ω, \mathcal{F}, P) , the random vector $X = (X_1, \dots, X_n)$ is a Borel measurable map from Ω to R^n . The same interpretation is possible for an infinite sequence of random variables, as follows.

5.11.3 Theorem. Let $(\Omega_j, \mathcal{F}_j)$, $j = 1, 2, \dots$, be arbitrary measurable spaces, and let p_j be the projection of $\prod_{j=1}^{\infty} \Omega_j$ onto the j th coordinate space. If (Ω, \mathcal{F}) is a measurable space and $X: \Omega \rightarrow \prod_{j=1}^{\infty} \Omega_j$, let $X_j = p_j \circ X$, $j = 1, 2, \dots$. Then X is measurable (relative to \mathcal{F} and $\prod_{j=1}^{\infty} \mathcal{F}_j$) iff each X_j is measurable (relative to \mathcal{F} and \mathcal{F}_j). In particular, if X_1, X_2, \dots are random variables, then $X = (X_1, X_2, \dots)$ is measurable: $(\Omega, \mathcal{F}) \rightarrow (R^\infty, \mathcal{B}^\infty)$, $\mathcal{B} = \mathcal{B}(R)$. For this reason X is sometimes called a *random sequence*.

PROOF. If X is measurable, each X_j is a composition of measurable maps and is therefore measurable. Conversely, assume each X_j to be measurable. Let $B = \{(\omega_1, \omega_2, \dots): \omega_j \in A_j, j = 1, \dots, n\}$, the $A_j \in \mathcal{F}_j$, be a measurable rectangle in $\prod_{j=1}^{\infty} \mathcal{F}_j$. Then

$$\{\omega \in \Omega: X(\omega) \in B\} = \bigcap_{j=1}^n \{\omega \in \Omega: X_j(\omega) \in A_j\} \in \mathcal{F}. \blacksquare$$

The proof of 5.11.3 shows also that if $X: \Omega \rightarrow \prod_{j=1}^n \Omega_j$, then X is measurable iff each X_j , $1 \leq j \leq n$, is measurable.

A version of Theorem 5.11.3, valid in arbitrary (possibly uncountable) product spaces, was given in Section 4.4, Problem 1. The result is repeated here to make the present discussion independent of Chapter 4.

We close the chapter with an introduction to one of the basic limit theorems of probability.

5.11.4 Weak Law of Large Numbers. Let X_1, X_2, \dots be independent random variables (not necessarily with the same distribution), each with finite mean and variance. Assume the variances to be uniformly bounded by $M < \infty$. Let $S_n = X_1 + \dots + X_n$. Then $[S_n - E(S_n)]/n$ converges in probability to 0, that is, given $\varepsilon > 0$,

$$P \left\{ \left| \frac{S_n - E(S_n)}{n} \right| \geq \varepsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. By Chebyshev's inequality,

$$\begin{aligned}
 P\{|(S_n - ES_n)/n| \geq \varepsilon\} &\leq \frac{1}{\varepsilon^2} E\left[\left(\frac{S_n - ES_n}{n}\right)^2\right] \\
 &= \frac{1}{\varepsilon^2 n^2} \text{Var } S_n \\
 &= \frac{1}{\varepsilon^2 n^2} \sum_{k=1}^n \text{Var } X_k \quad \text{by 5.10.11} \\
 &\leq \frac{M}{\varepsilon^2 n} \rightarrow 0. \blacksquare
 \end{aligned}$$

There are two special cases of particular interest.

1. If $E(X_i) = m$ for all i , then $[S_n - E(S_n)]/n = (S_n/n) - m$; hence $S_n/n \rightarrow m$ in probability. Thus for large n , the arithmetic average of n independent random variables, each with finite expectation m (and with the variances uniformly bounded) is quite likely to be very close to m .
2. If X_1, X_2, \dots are independent, and for each i , $P\{X_i = 1\} = p$, $P\{X_i = 0\} = q = 1 - p$ (thus we have an infinite sequence of Bernoulli trials), then $X_1 + \dots + X_n$ is the number of successes in n trials, hence S_n/n is the relative frequency of successes. Since $EX_i = p$, we have $S_n/n \rightarrow p$ in probability. Thus for large n , the relative frequency of successes is quite likely to be very close to p .

Intuitively, the weak law of large numbers says the following. If we regard observation of X_1, \dots, X_n as one performance of an experiment, where n is very large but *fixed*, then if we repeat the experiment independently, $(S_n - ES_n)/n$ will be close to 0 a very high percentage of the time.

But physically we expect something more than this. If a coin with probability p of heads is tossed over and over again, we expect the relative frequency to approach p in the ordinary sense of convergence of a sequence of real numbers; in other words, given $\varepsilon > 0$, eventually the relative frequency S_n/n gets and *remains* within ε of p . Here we are considering observation of the infinite sequence X_1, X_2, \dots as one performance of the experiment, and what we must show is that $\lim_{n \rightarrow \infty} S_n(\omega)/n = p$ for almost every ω . A statement of this type is called a *strong law of large numbers*. This subject will be considered in detail later on.

Problems

1. (a) If Y_1, Y_2, \dots are independent random objects, show that for each n , (Y_1, \dots, Y_n) and $(Y_{n+1}, Y_{n+2}, \dots)$ are independent.

- (b) In part (a), if $Y_i: (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ for all i , and all Y_i have the same distribution, show that (Y_1, Y_2, \dots) and (Y_n, Y_{n+1}, \dots) have the same distribution for all n .
2. Consider the gambler's ruin problem, that is, the simple random walk with absorbing barriers at 0 and b . In this problem, Y_1, Y_2, \dots are independent random variables, with $P\{Y_i = 1\} = p$, $P\{Y_i = -1\} = q = 1 - p$. Let $X_n = \sum_{k=1}^n Y_k$, and let x be an arbitrary integer between 1 and $b - 1$. We wish to find the probability $h(x)$ of eventual ruin starting from x , in other words, the probability that $x + X_n$ will reach 0 before it reaches b . Intuitive reasoning based on the theorem of total probability leads to the result that

$$h(x) = ph(x + 1) + qh(x - 1).$$

Give a formal proof of this result. [For further details, see Ash (1970, Chapter 6).]

5.12 References

The general outline of this chapter is based on Ash (1970), which is a text for an undergraduate course in probability. Measure theory is not used in the book, although some of the underlying measure-theoretic ideas are sketched. Many additional examples and problems can be found in Feller (1950) and Parzen (1960).

6

Conditional Probability and Expectation

6.1 Introduction

In Chapter 5, we defined the conditional probability $P(B|A)$ only when $P(A) > 0$. However, conditional probabilities given events of probability zero are in no sense degenerate cases; they occur naturally in many problems. For example, consider the following two-stage random experiment. A random variable X is observed, where X has distribution function F . If X takes the value x , a random variable Y is observed, where the distribution of Y depends on x . (For example, if $0 \leq x \leq 1$, a coin with probability of heads x might be tossed independently n times, with Y the resulting number of heads.) Thus $P(x, B) = P\{Y \in B | X = x\}$ is prescribed in the statement of the problem, although the event $\{X = x\}$ may have probability zero for all values of x .

Let us try to construct a model for the above situation. Let $\Omega = \mathbb{R}^2$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^2)$, $X(x, y) = x$, $Y(x, y) = y$. Instead of specifying the joint distribution function of X and Y , we specify the distribution function of X , and thus the corresponding probability measure P_X ; also, for each x we are given a probability measure $P(x, \cdot)$ defined on $B(\mathbb{R})$; $P(x, B)$ is interpreted (informally for now) as $P\{Y \in B | X = x\}$.

We claim that the probability of any event of the form $\{(X, Y) \in C\}$ is determined. Reasoning intuitively, the probability that X falls into $(x, x + dx]$ is $dF(x)$. Given that this occurs, in other words (roughly), given $X = x$, (X, Y) will lie in C iff Y belongs to the section $C(x) = \{y : (x, y) \in C\}$. The

probability of this event is $P(x, C(x))$. The total probability that (X, Y) will belong to C is

$$P(C) = \int_{-\infty}^{\infty} P(x, C(x)) dF(x). \quad (1)$$

In the special case $C = \{(x, y) : x \in A, y \in B\} = A \times B$, $C(x) = B$ if $x \in A$ and $C(x) = \emptyset$ if $x \notin A$; therefore

$$P(C) = P(A \times B) = \int_A P(x, B) dF(x). \quad (2)$$

Now if $P(x, B)$ is Borel measurable in x for each fixed $B \in \mathcal{B}(R)$, then by the product measure theorem, there is a unique (probability) measure on $\mathcal{B}(R^2)$ satisfying (2) for all $A, B \in \mathcal{B}(R)$, namely, the measure given by (1). Thus in the mathematical formulation of the problem, we take the probability measure P on $\mathcal{F} = \mathcal{B}(R^2)$ to be the unique measure determined by P_x and the measures $P(x, \cdot)$, $x \in R$.

6.2 Applications

We apply the results of Section 6.1 to some typical situations in probability.

6.2.1 Example. Let X be uniformly distributed between 0 and 1. If $X = x$, a coin with probability x of heads is tossed independently n times. If Y is the resulting number of heads, find $P\{Y = k\}$, $k = 0, 1, \dots, n$.

Let us translate this into mathematical terms. Let $\Omega_1 = [0, 1]$, $\mathcal{F}_1 = \mathcal{B}[0, 1]$. We have specified $P_X(A) = \int_A dx$ = Lebesgue measure of A , $A \in \mathcal{F}_1$.

For each x , we are given $P(x, B)$, to be interpreted as the conditional probability that $Y \in B$, given $X = x$. We may take $\Omega_2 = \{0, 1, \dots, n\}$, \mathcal{F}_2 the class of all subsets of Ω_2 ; then $P(x, \{k\}) = \binom{n}{k} x^k (1-x)^{n-k}$, $k = 0, 1, \dots, n$ (this is Borel measurable in x). We take $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, P the unique probability measure determined by P_X and the $P(x, \cdot)$, namely,

$$P(C) = \int_0^1 P(x, C(x)) dP_X(x) = \int_0^1 P(x, C(x)) dx.$$

Now let $X(x, y) = x$, $Y(x, y) = y$. Then

$$\begin{aligned} P\{Y = k\} &= P(\Omega_1 \times \{k\}) = \int_0^1 P(x, \{k\}) dx \\ &= \int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \binom{n}{k} \beta(k+1, n-k+1), \end{aligned}$$

where $\beta(r, s) = \int_0^1 x^{r-1} (1-x)^{s-1} dx$, $r, s > 0$, is the *beta function*. We can

express $\beta(r, s)$ as $\Gamma(r)\Gamma(s)/\Gamma(r+s)$, where $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x} dx$, $r > 0$, is the *gamma function*. Since $\Gamma(n+1) = n!$, $n = 0, 1, \dots$, we have

$$P\{Y = k\} = \frac{\binom{n}{k} k! (n-k)!}{(n+1)!} = \frac{1}{n+1}, \quad k = 0, 1, \dots, n.$$

In doing a problem of this type, intuitive reasoning serves as a useful check on the formal development. Thus, the probability that X falls near x is dx ; given that $X = x$, the probability that k heads will be obtained is $\binom{n}{k} x^k (1-x)^{n-k}$. Integrate this from 0 to 1 to obtain the total probability.

The next example involves an n -stage random experiment.

6.2.2 Example. Let X_1 be uniformly distributed between 0 and 1. If $X_1 = x_1$, let X_2 be uniformly distributed between 0 and x_1 . In general, if $X_1 = x_1, \dots, X_k = x_k$, let X_{k+1} be uniformly distributed between 0 and x_k ($k = 1, \dots, n-1$). Find the expectation of X_n .

Here we have $\Omega_j = R$, $\mathcal{F}_j = \mathcal{B}(R)$, $\Omega = \prod_{j=1}^n \Omega_j$, $\mathcal{F} = \prod_{j=1}^n \mathcal{F}_j$, $X_j(x_1, \dots, x_n) = x_j$, $j = 1, \dots, n$.

Set P_1 = Lebesgue measure on $(0, 1)$, and for each $x_1 \in (0, 1)$,

$$P(x_1, \cdot) = \frac{1}{x_1} [\text{Lebesgue measure on } (0, x_1)],$$

that is,

$$P(x_1, B) = \frac{1}{x_1} \int_{B \cap (0, x_1)} dx_2.$$

In general, for each $x_1, \dots, x_k \in (0, 1)$, $k = 1, \dots, n-1$, take

$$P(x_1, \dots, x_k, \cdot) = \frac{1}{x_k} [\text{Lebesgue measure on } (0, x_k)].$$

(We use open intervals to avoid division by zero.)

Let P be the unique measure on \mathcal{F} determined by P_1 and the $P(x_1, \dots, x_k, \cdot)$. We may find the expectation of a Borel measurable function g from R^n to R by Fubini's theorem:

$$\int_{\Omega} g dP = \int_{\Omega_1} P_1(dx_1) \int_{\Omega_2} P(x_1, dx_2) \cdots \int_{\Omega_n} g(x_1, \dots, x_n) P(x_1, \dots, x_{n-1}, dx_n).$$

In the present case we have $g(x_1, \dots, x_n) = X_n(x_1, \dots, x_n) = x_n$. Thus

$$\begin{aligned} E(X_n) &= \int_0^1 dx_1 \int_0^{x_1} \frac{1}{x_1} dx_2 \cdots \int_0^{x_{n-2}} \frac{1}{x_{n-2}} dx_{n-1} \int_0^{x_{n-1}} \frac{x_n}{x_{n-1}} dx_n \\ &= \int_0^1 \frac{x_1}{2^{n-1}} dx_1 = 2^{-n}. \end{aligned}$$

This example has an alternative interpretation. Let Y_1, \dots, Y_n be independent random variables, each uniformly distributed between 0 and 1. Let Z_k be the product $Y_1 \cdots Y_k$, $1 \leq k \leq n$. It turns out (see Problem 2, Section 6.6) that (Z_1, \dots, Z_n) has the same distribution as (X_1, \dots, X_n) ; hence $E(X_n) = E(Z_n)$. Since the expectation of a product of independent nonnegative random variables is the product of the expectations, $E(Z_n) = \prod_{k=1}^n E(Y_k) = 2^{-n}$ as before.

6.2.3 Example. Let X be a discrete random variable, taking on the positive integer values 1, 2, ... with probabilities p_1, p_2, \dots ($p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1$). If $X = n$, a nonnegative number Y is selected according to the density f_n . Find the probability that $1 \leq X + Y \leq 3$.

Here we have

$$\Omega_1 = \{1, 2, \dots\}, \quad \mathcal{F}_1 = \text{the class of all subsets of } \Omega,$$

$$P_1\{k\} = p_k, \quad k = 1, 2, \dots,$$

$$\Omega_2 = R, \quad \mathcal{F}_2 = \mathcal{B}(R), \quad P(n, B) = \int_B f_n(x) dx,$$

$$X(\omega_1, \omega_2) = \omega_1, \quad Y(\omega_1, \omega_2) = \omega_2.$$

The measure determined by P_1 and the $P(n, \cdot)$ is given by

$$\begin{aligned} P(F) &= \int_{\Omega_1} P(\omega_1, F(\omega_1)) P_1(d\omega_1) \quad F \in \mathcal{F}_1 \times \mathcal{F}_2 \\ &= \sum_{n=1}^{\infty} P(n, F(n)) p_n \quad [\text{see 5.10.3(d)}]. \end{aligned}$$

If $F = \{(\omega_1, \omega_2): 1 \leq \omega_1 + \omega_2 \leq 3\}$, then $F(n) = \{\omega_2: 1 - n \leq \omega_2 \leq 3 - n\}$; hence

$$\begin{aligned} P(F) &= P\{1 \leq X + Y \leq 3\} = \sum_{n=1}^{\infty} p_n \int_{1-n}^{3-n} f_n(x) dx \\ &= p_1 \int_0^2 f_1(x) dx + p_2 \int_0^1 f_2(x) dx \quad \text{since } f_n(x) = 0 \text{ for } x < 0. \end{aligned}$$

Note that if $p_n > 0$, then $P\{Y \in B | X = n\}$ is defined and equals $P\{X = n, Y \in B\}/P\{X = n\}$. But $P\{X = n, Y \in B\} = P\{(\omega_1, \omega_2): \omega_1 = n, \omega_2 \in B\} = P(n, B)p_n$. Thus $P\{Y \in B | X = n\} = P(n, B)$, as we would expect intuitively.

For additional examples, see Ash (1970, Chapter 4).

6.3 The General Concept of Conditional Probability and Expectation

We have seen that specification of the distribution of a random variable X , together with $P(x, B)$, x real, $B \in \mathcal{B}(R)$, interpreted intuitively as the conditional probability that $Y \in B$ given $X = x$, determines a unique probability measure on $\mathcal{B}(R^2)$, so that there is only one reasonable joint distribution of X and Y consistent with the given data. However, this somewhat oblique approach has not resolved the difficulty of defining conditional probabilities given events with probability zero. For example, if X is a random object on (Ω, \mathcal{F}, P) , that is, $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, and $B \in \mathcal{F}$, we may ask whether it is possible to define in a meaningful way the conditional probability $P(x, B) = P(B|X = x)$, even though the event $\{X = x\}$ may have probability zero for some, in fact perhaps for all, x .

By the discussion in Section 6.1, if we have a reasonable conditional probability $P(x, B)$, it should satisfy

$$P(\{X \in A\} \cap B) = \int_A P(x, B) dP_X(x),$$

where P_X is the probability measure induced by X , namely,

$$P_X(A) = P\{\omega: X(\omega) \in A\}, A \in \mathcal{F}'.$$

In fact, this requirement determines $P(x, B)$ in the following sense.

6.3.1 Theorem. Let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ be a random object on (Ω, \mathcal{F}, P) , and let B be a fixed set in \mathcal{F} . Then there is a real-valued Borel measurable function g on (Ω', \mathcal{F}') such that for each $A \in \mathcal{F}'$,

$$P(\{X \in A\} \cap B) = \int_A g(x) dP_X(x).$$

Furthermore, if h is another such function then $g = h$ a.e. [P_X]. [We define $P(B|X = x)$ as $g(x)$; it is essentially unique for a given B .]

PROOF. Let $\lambda(A) = P(\{X \in A\} \cap B)$, $A \in \mathcal{F}'$. Then λ is a finite measure on \mathcal{F}' , absolutely continuous with respect to P_X [$P_X(A) = 0$ implies $\lambda(A) = 0$]. The result follows from the Radon–Nikodym theorem. ■

Let us verify that the conditional probability we have just introduced coincides with our intuition in simple cases.

6.3.2 Examples. (a) Let X take on only countably many values x_1, x_2, \dots , with

$$p_i = P\{X = x_i\} > 0, \quad \sum_{i=1}^{\infty} p_i = 1.$$

We claim that

$$g(x_i) = P(B | X = x_i) = \frac{P(B \cap \{X = x_i\})}{P\{X = x_i\}}, \quad i = 1, 2, \dots.$$

(Since P_X is concentrated on the x_i , we need not bother to specify $g(x)$ for x unequal to any of the x_i .) Thus the general definition reduces to the elementary definition in the discrete case. To prove this, let $\Omega' = \{x_1, x_2, \dots\}$, with \mathcal{F}' the collection of all subsets of Ω' . If $A \in \mathcal{F}'$ and g is defined as above, then

$$\begin{aligned} \int_A g(x) dP_X(x) &= \int_{\Omega'} g(x) I_A(x) dP_X(x) = \sum_{i=1}^{\infty} g(x_i) I_A(x_i) P_X\{x_i\} \\ &\quad [\text{see 5.10.3(d)}] \\ &= \sum_{x_i \in A} g(x_i) P\{X = x_i\} = \sum_{x_i \in A} P(B \cap \{X = x_i\}) \\ &= P(\{X \in A\} \cap B). \end{aligned}$$

Since there is essentially only one g satisfying

$$\int_A g dP_X = P(\{X \in A\} \cap B), \quad A \in \mathcal{F}',$$

the g we proposed must be correct.

(b) Let X and Y be random variables with joint density f

$$\begin{aligned} \Omega &= R^2, \quad \mathcal{F} = \mathcal{B}(R^2), \quad X(x, y) = x, \quad Y(x, y) = y, \\ P(A) &= \iint_A f(x, y) dx dy, \quad A \in \mathcal{F}. \end{aligned}$$

Now $\{X = x\}$ has probability zero for each x , but there is a reasonable approach to the conditional probability $P\{Y \in C | X = x\}$, as follows:

$$\begin{aligned} P\{Y \in C | x - h < X < x + h\} &= \frac{P\{x - h < X < x + h, Y \in C\}}{P\{x - h < X < x + h\}} \\ &= \frac{\int_{u=x-h}^{x+h} \int_{y \in C} f(u, y) du dy}{\int_{x-h}^{x+h} f_1(u) du}, \end{aligned}$$

where $f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$ is the density of X .

For small h , this is (hopefully) approximately

$$\frac{2h \int_C f(x, y) dy}{2hf_1(x)} = \int_C \frac{f(x, y)}{f_1(x)} dy.$$

We are led to define

$$h(y|x) = \frac{f(x, y)}{f_1(x)}$$

as the *conditional density* of Y , given $X = x$ (or for short, the *conditional density of Y given X*). Note that h is defined only when $f_1(x) \neq 0$; however, if $S = \{(x, y) : f_1(x) = 0\}$, then $P\{(X, Y) \in S\} = 0$, since

$$\begin{aligned} P\{(X, Y) \in S\} &= \iint_S f(x, y) dx dy = \int_{\{x : f_1(x) = 0\}} \left[\int_{y=-\infty}^{\infty} f(x, y) dy \right] dx \\ &= \int_{\{x : f_1(x) = 0\}} f_1(x) dx = 0. \end{aligned}$$

Thus we may essentially ignore those (x, y) for which the conditional density is not defined.

We expect that $P\{Y \in C | X = x\} = \int_C h(y|x) dy$. More generally, if $B \in \mathcal{F}$ and $X = x$, then B will occur iff $Y \in B(x)$. To find $P\{Y \in B(x) | X = x\}$, we integrate $h(y|x)$ over $y \in B(x)$. Thus we propose

$$g(x) = \int_{B(x)} h(y|x) dy, \quad B \in \mathcal{F}, \quad x \in R,$$

as the conditional probability of B given $X = x$. To prove this, first note that

$$g(x) = \int_{-\infty}^{\infty} I_B(x, y) h(y|x) dy;$$

hence g is Borel measurable by Fubini's theorem. Also, if $A \in \mathcal{B}(R)$,

$$\begin{aligned} P(\{X \in A\} \cap B) &= \iint_{\substack{x \in A \\ (x, y) \in B}} f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} I_B(x, y) h(y|x) dy \right] I_A(x) f_1(x) dx \\ &= \int_{x \in A} f_1(x) \int_{y \in B(x)} h(y|x) dy dx \\ &= \int_A g(x) f_1(x) dx \\ &= \int_A g(x) dP_X(x) \quad [\text{see 5.10.3(c)}]. \end{aligned}$$

Therefore $g(x) = P(B | X = x)$.

In this example we may look at the formula $f(x, y) = f_1(x)h(y|x)$ in two ways. If (X, Y) has density f , we have a notion of conditional probability: $P\{Y \in C | X = x\} = \int_C h(y|x) dy$. On the other hand, suppose that we specify that X has density f_1 , and whenever $X = x$, we select Y according to the density $h(\cdot|x)$; in other words, we specify $P(x, B) = \int_B h(y|x) dy$, $B \in \mathcal{B}(R)$. A unique measure P on $\mathcal{B}(R^2)$ is determined, satisfying, for $A, B \in \mathcal{B}(R)$,

$$\begin{aligned} P\{X \in A, Y \in B\} &= \int_A P(x, B) f_1(x) dx \\ &= \iint_{\substack{x \in A \\ y \in B}} f_1(x) h(y|x) dx dy. \end{aligned}$$

Therefore (X, Y) has density $f(x, y) = f_1(x)h(y|x)$.

Thus we have two points of view. We may regard the conditional density of Y given $X = x$ as ultimately derived from the joint density of X and Y . On the other hand, we may regard the observation of X and Y as a two stage random experiment, where the distribution of Y at stage 2 depends on the value of X at stage 1. The above discussion shows that the assignment of probabilities to events involving (X, Y) is the same in either case.

We may also define conditional densities in higher dimensions. For example, if X, Y, Z, W have joint density f , we define (say) the conditional density of (Z, W) , given (X, Y) , as

$$h(z, w|x, y) = \frac{f(x, y, z, w)}{f_{XY}(x, y)},$$

where $f_{XY}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z, w) dz dw$. If $B \in \mathcal{B}(R^4)$, then, exactly as before,

$$P(B | X = x, Y = y) = \iint_{B(x, y)} h(z, w|x, y) dz dw.$$

This is verified by proving that, for $A \in \mathcal{B}(R^2)$,

$$P(\{(X, Y) \in A\} \cap B) = \int_A P(B | X = x, Y = y) f_{XY}(x, y) dx dy.$$

(c) Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be given, with no probability defined as yet. Take $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $X(\omega_1, \omega_2) = \omega_1$, $Y(\omega_1, \omega_2) = \omega_2$. Assume that we are given P_X , a probability measure on $(\Omega_1, \mathcal{F}_1)$, and also that we are given $P(x, B)$, $x \in \Omega_1$, $B \in \mathcal{F}_2$, a probability measure in B for each fixed x , and a Borel measurable function of x for each fixed B . (We are specifying the distribution of X and the conditional distribution of Y , given

$X = x$.) By the product measure theorem, there is a unique measure P on \mathcal{F} such that

$$P\{X \in A, Y \in B\} = P(A \times B) = \int_A P(x, B) dP_X(x).$$

It follows that $P(x, B)$ is in fact the conditional probability $P\{Y \in B | X = x\}$.

We now consider conditional expectation. Let X and Y be random variables on (Ω, \mathcal{F}, P) ; we ask for a reasonable definition of the expectation of Y given that $X = x$, written $E(Y | X = x)$. Intuitively, $E(Y | X = x)$ should reflect the long run average value of Y in a sequence of independent trials when we look only at those observations on which $\{X = x\}$ has occurred.

If X and Y are discrete and we are given that $X = x$, the conditional probability of an event involving Y is governed by the set of conditional probabilities $p(y|x) = P\{Y = y | X = x\}$. Thus a reasonable figure for $E(Y | X = x)$ is $\sum_y y p(y|x)$. Similarly, if (X, Y) is absolutely continuous, and $h = h(y|x)$ is the conditional density of Y given $X = x$, we expect that $E(Y | X = x)$ should be $\int_{-\infty}^{\infty} y h(y|x) dy$. What we need is a general framework which includes these special cases.

Let Y be a random variable (or an extended random variable) on (Ω, \mathcal{F}, P) , and let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ be a random object. Our general definition of conditional probability hinges on a version of the theorem of total probability:

$$P(\{X \in A\} \cap B) = \int_A P(B | X = x) dP_X(x), \quad A \in \mathcal{F}', \quad B \in \mathcal{F}.$$

There is a closely related “theorem of total expectation,” which may be developed intuitively as follows. The probability that X falls near x is $dP_X(x)$; given that $X = x$, the average value of Y is what we are looking for, namely, $E(Y | X = x)$. It is reasonable to hope that the total expectation may be found by adding all the contributions:

$$E(Y) = \int_{\Omega'} E(Y | X = x) dP_X(x).$$

To develop this further, we replace Y by $YI_{\{X \in A\}}$, where $A \in \mathcal{F}'$. If $x \in A$, we expect that $E(YI_{\{X \in A\}} | X = x) = E(Y | X = x)$ since $X(\omega) = x \in A$ implies $I_{\{X \in A\}}(\omega) = 1$. If $x \notin A$, we expect that $E(YI_{\{X \in A\}} | X = x) = 0$. Replacing Y by $YI_{\{X \in A\}}$ in the above version of the theorem of total expectation, we obtain

$$E(YI_{\{X \in A\}}) = \int_{\Omega'} E(YI_{\{X \in A\}} | X = x) dP_X(x)$$

or

$$\int_{\{X \in A\}} Y dP = \int_A E(Y | X = x) dP_X(x).$$

In fact, this requirement essentially determines $E(Y | X = x)$.

6.3.3 Theorem. Let Y be an extended random variable on (Ω, \mathcal{F}, P) , and $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, a random object. If $E(Y)$ exists, there is a function $g: (\Omega', \mathcal{F}') \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$ such that for each $A \in \mathcal{F}'$,

$$\int_{\{X \in A\}} Y dP = \int_A g(x) dP_X(x).$$

(As usual, \mathcal{B} denotes the class of Borel sets.) Furthermore, if h is another such function, then $g = h$ a.e. [P_X]. [We define $E(Y|X=x)$ as $g(x)$; it is essentially unique for a given Y .]

PROOF. Let

$$\lambda(A) = \int_{\{X \in A\}} Y dP = \int_{X^{-1}(A)} Y dP, \quad A \in \mathcal{F}'.$$

Then λ is a countably additive set function on \mathcal{F}' by 1.6.1, and is absolutely continuous with respect to P_X since $P_X(A) = P\{X \in A\}$. The result follows from the Radon–Nikodym theorem. ■

Conditional expectation includes conditional probability as a special case, as we now prove.

6.3.4 Corollary. If X is a random object on (Ω, \mathcal{F}, P) and $B \in \mathcal{F}$, then

$$E(I_B | X = x) = P(B | X = x) \quad \text{a.e. } [P_X].$$

PROOF. In 6.3.3, set $Y = I_B$; the defining equation for conditional expectation becomes

$$P(\{X \in A\} \cap B) = \int_A E(I_B | X = x) dP_X(x).$$

The result now follows from 6.3.1. ■

Let us compare the general definition with the intuitive concept in special cases.

6.3.5 Examples. (a) Let X take on only countably many values x_1, x_2, \dots (assume all $P\{X = x_i\} > 0$). We have seen that

$$P(B | X = x_i) = \frac{P(B \cap \{X = x_i\})}{P\{X = x_i\}}, \quad B \in \mathcal{F}.$$

Thus we should expect that

$$E(I_B | X = x_i) = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} I_B dP.$$

Proceeding from indicators to nonnegative simple functions to nonnegative measurable functions to arbitrary measurable functions, we should like to believe that if $E(Y)$ exists,

$$E(Y | X = x_i) = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y dP, \quad i = 1, 2, \dots. \quad (1)$$

[We are not proving anything here since we do not yet know, for example, that

$$E\left(\sum_{i=1}^n Y_i \mid X = x\right) = \sum_{i=1}^n E(Y_i | X = x).$$

To establish (1), let

$$g(x_i) = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y dP, \quad i = 1, 2, \dots.$$

(We may assume $\Omega' = \{x_1, x_2, \dots\}$, with \mathcal{F}' the class of all subsets of Ω' .) Then

$$\begin{aligned} \int_{\{X \in A\}} Y dP &= \sum_{x_i \in A} P\{X = x_i\} \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y dP \\ &= \sum_{x_i \in A} P\{X = x_i\} g(x_i) = \int_A g(x) dP_X(x), \quad A \in \mathcal{F}', \end{aligned}$$

as desired.

In the special case when Y is discrete, (1) assumes a simpler form. If Y takes on the values y_1, y_2, \dots , we obtain (using countable additivity of the integral)

$$\begin{aligned} E(Y | X = x_i) &= \sum_j y_j \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} \\ &= \sum_j y_j P\{Y = y_j | X = x_i\}. \end{aligned} \quad (2)$$

(b) Let $B \in \mathcal{F}$, and assume $P(B) > 0$. If $E(Y)$ exists, we define the *conditional expectation of Y given B* , as follows. Let $X = I_B$, and set $E(Y | B) = E(Y | X = 1)$. This is a special case of (a); we obtain [see (1)]

$$E(Y | B) = \frac{1}{P(B)} \int_B Y dP,$$

in other words,

$$E(Y|B) = \frac{E(YI_B)}{P(B)}. \quad (3)$$

(c) Let X and Y be random variables having a joint density f , and let $h = h(y|x)$ be the conditional density of Y given X . We claim that if $E(Y)$ exists,

$$E(Y|X=x) = \int_{-\infty}^{\infty} yh(y|x) dy. \quad (4)$$

To prove this, note that

$$\begin{aligned} \int_{\{X \in A\}} Y dP &= \iint_{\{(x,y):x \in A\}} yf(x,y) dx dy \\ &= \int_{x \in A} f_1(x) \left[\int_{-\infty}^{\infty} yh(y|x) dy \right] dx \quad \text{by Fubini's theorem} \\ &= \int_A \left[\int_{-\infty}^{\infty} yh(y|x) dy \right] dP_X(x), \end{aligned}$$

proving (4).

Notice also that if q is a Borel measurable function from R to R and $E[q(Y)]$ exists, then

$$E(q(Y)|X=x) = \int_{-\infty}^{\infty} q(y)h(y|x) dy \quad (5)$$

by the same argument as above. Similarly, if X and Y are discrete [see (a), Eq. (2)] and $E[q(Y)]$ exists, then

$$E(q(Y)|X=x_i) = \sum_j q(y_j) P\{Y=y_j|X=x_i\}. \quad (6)$$

(d) Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be given, with no probability defined as yet. Let $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $X(x, y) = x$, $Y(x, y) = y$. Assume that a probability measure P_X on \mathcal{F}_1 is given, and also that we are given $P(x, B)$, $x \in \Omega_1$, $B \in \mathcal{F}_2$, a probability measure in B for each fixed x , and a Borel measurable function of x for each fixed B . Let P be the unique measure on \mathcal{F} determined by P_X and the $P(x, \cdot)$.

If $f: (\Omega_2, \mathcal{F}_2) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ and $E[f(Y)]$ exists, we claim that

$$E(f(Y)|X=x) = \int_{\Omega_2} f(y)P(x, dy). \quad (7)$$

To see this, we note, with the aid of Fubini's theorem, that

$$\begin{aligned} \int_{\{X \in A\}} f(Y) dP &= \int_{\Omega} f(Y) I_{\{X \in A\}} dP \\ &= \int_{\Omega_1} \int_{\Omega_2} f(Y(x, y)) I_A(x) P(x, dy) dP_X(x) \\ &= \int_A \left[\int_{\Omega_2} f(y) P(x, dy) \right] dP_X(x). \end{aligned}$$

Problems

1. Let X and Y be random variables with joint density $f(x, y)$. Indicate how to compute the following quantities.
 - (a) $E(g(X)|Y = y)$, where g is a Borel measurable function from R to R such that $E[g(X)]$ exists;
 - (b) $E(Y|A)$, where $A = \{X \in B\}$, $B \in \mathcal{B}(R)$;
 - (c) $E(X|A)$, where $A = \{X + Y \in B\}$, $B \in \mathcal{B}(R^2)$.
2. Let X be a random variable with density $f_0(\lambda)$. If $X = \lambda$, n independent observations X_1, \dots, X_n are taken, where each X_i has density $f_\lambda(x)$. Indicate how to compute the conditional expectation of $g(X)$, given $X_1 = x_1, \dots, X_n = x_n$.
3. Let X be a discrete random variable; if $X = x$, let Y have a conditional density $h(y|x)$. Show that

$$P\{X = x | Y = y\} = \frac{P\{X = x\}h(y|x)}{\sum_{x'} P\{X = x'\}h(y|x')}.$$

4. Let X be an absolutely continuous random variable. If $X = x$, let Y be discrete, with $P\{Y = y | X = x\} = p(y|x)$ specified. Show that there is a conditional density of X given Y , namely,

$$h(x|y) = \frac{f_X(x)p(y|x)}{P_Y(y)},$$

where

$$p_Y(y) = P\{Y = y\} = \int_{-\infty}^{\infty} f_X(x)p(y|x) dx.$$

5. (a) Let X be a discrete random variable: If $X = \lambda$, n independent observations X_1, \dots, X_n are taken, where each X_i has density $f_\lambda(x)$. Indicate how to compute $E(g(X)|X_1 = x_1, \dots, X_n = x_n)$.

 (b) Let X be an absolutely continuous random variable. If $X = \lambda$, n independent observations X_1, \dots, X_n are taken, where each X_i is discrete with probability function $p_\lambda(x)$. Indicate how to compute $E(g(X)|X_1 = x_1, \dots, X_n = x_n)$.

6. If X is a random vector with density f , and $A = \{X \in B_0\}$, $B_0 \in \mathcal{B}(R^n)$, show that there is a conditional density for X given A , namely,

$$f(x|A) = \begin{cases} \frac{f(x)}{P(A)} & \text{if } x \in B_0, \\ 0 & \text{if } x \notin B_0. \end{cases}$$

[The interpretation of the conditional density is that

$$P\{X \in B|A\} = \int_B f(x|A) dx, \quad B \in \mathcal{B}(R^n).]$$

7. Let B_1, B_2, \dots be mutually exclusive and exhaustive events with strictly positive probability. Establish the following version of the theorem of total expectation: If $E(X)$ exists, then

$$E(X) = \sum_{n=1}^{\infty} P(B_n)E(X|B_n).$$

8. Let X and Y be nonnegative random variables, such that (X, Y) has induced probability measure

$$P_{XY} = \frac{1}{4}P_1 + \frac{3}{4}P_2,$$

where, for $B \in \mathcal{B}([0, \infty) \times [0, \infty))$, P_1 = point mass at $(1, 2)$, that is,

$$P_1(B) = \begin{cases} 1 & \text{if } (1, 2) \in B, \\ 0 & \text{if } (1, 2) \notin B, \end{cases}$$

and

$$P_2(B) = \iint_B e^{-x}e^{-y} dx dy.$$

Thus with probability $\frac{1}{4}$, $(X, Y) = (1, 2)$; with probability $\frac{3}{4}$, X and Y are chosen independently, each with density e^{-x} , $x \geq 0$. Calculate $P\{Y \in B | X = x\}$, $x \geq 0$, $B \in \mathcal{B}[0, \infty)$. (Hint: think like a statistician; if you observe that $X = 1$, it is a moral certainty that you are operating under P_1 ; if $X \neq 1$, it is an absolute certainty that you are operating under P_2 .)

6.4 Conditional Expectation Given a σ -field

It will be very convenient to regard conditional expectations as functions defined on the sample space Ω . Let us first recall the main result of the previous section.

If Y is an extended random variable on (Ω, \mathcal{F}, P) whose expectation exists, and $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ is a random object, then $g(x) = E(Y|X=x)$ is characterized as the a.e. $[P_X]$ unique function: $(\Omega', \mathcal{F}') \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$ satisfying

$$\int_{\{X \in A\}} Y dP = \int_A E(Y|X=x) dP_X(x), \quad A \in \mathcal{F}'. \quad (1)$$

Now let $h(\omega) = g(X(\omega))$; then $h: (\Omega, \mathcal{F}) \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$ (see Fig. 6.1).

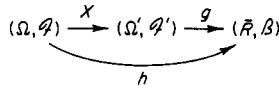


Figure 6.1.

Thus $h(\omega)$ is the conditional expectation of Y , given that X takes the value $x = X(\omega)$; consequently, h measures the average value of Y given X , but h is defined on Ω rather than Ω' .

It will be useful to have an analog on (1) for h . We claim that

$$\int_{\{X \in A\}} h dP = \int_{\{X \in A\}} Y dP \quad \text{for each } A \in \mathcal{F}'. \quad (2)$$

To prove this, note that

$$\begin{aligned} \int_{\{X \in A\}} h dP &= \int_{\Omega} g(X(\omega)) I_A(X(\omega)) dP(\omega) \\ &= \int_{\Omega'} g(x) I_A(x) dP_X(x) \quad [\text{by 5.10.3(b)}] \\ &= \int_A g(x) dP_X(x) \\ &= \int_{\{X \in A\}} Y dP \quad [\text{by (1)}]. \end{aligned}$$

Since $\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega: X(\omega) \in A\}$, we may express (2) as follows:

$$\int_C h dP = \int_C Y dP \quad \text{for each } C \in X^{-1}(\mathcal{F}'), \quad (3)$$

where $X^{-1}(\mathcal{F}') = \{X^{-1}(A): A \in \mathcal{F}'\}$.

The σ -field $X^{-1}(\mathcal{F}')$ will be very important for us, and we shall look at some of its properties before proceeding.

6.4.1 Definition. Let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ be a random object. The σ -field induced by X is given by

$$\mathcal{F}(X) = X^{-1}(\mathcal{F}').$$

Thus a set in $\mathcal{F}(X)$ is of the form $\{X \in A\}$ for some $A \in \mathcal{F}'$. In particular, if $X = (X_1, \dots, X_n)$, a random vector, $\mathcal{F}(X)$ consists of all sets $\{X \in B\}$, $B \in \mathcal{B}(R^n)$.

The induced σ -field has the following properties.

6.4.2 Theorem. Let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$.

- (a) The induced σ -field $\mathcal{F}(X)$ is the smallest σ -field \mathcal{G} of subsets of Ω making X measurable relative to \mathcal{G} and \mathcal{F}' .
- (b) If $\Omega' = \prod_j \Omega_j$, $\mathcal{F}' = \prod_j \mathcal{F}_j$ so that $X = (X_1, X_2, \dots)$, where $X_j: (\Omega, \mathcal{F}) \rightarrow (\Omega_j, \mathcal{F}_j)$, then $\mathcal{F}(X)$ is the smallest σ -field \mathcal{G} of subsets of Ω making each X_j measurable relative to \mathcal{G} and \mathcal{F}_j .
- (c) If $Z: (\Omega, \mathcal{F}(X)) \rightarrow (\bar{R}, \mathcal{B})$ (or (R, \mathcal{B})), then $Z = f \circ X$ for some $f: (\Omega', \mathcal{F}') \rightarrow (\bar{R}, \mathcal{B})$. Conversely, if $Z = f \circ X$ and $f: (\Omega', \mathcal{F}') \rightarrow (\bar{R}, \mathcal{B})$, then $Z: (\Omega, \mathcal{F}(X)) \rightarrow (\bar{R}, \mathcal{B})$.

PROOF. (a) If $A \in \mathcal{F}'$, then $X^{-1}(A) \in \mathcal{F}(X)$ by definition of $\mathcal{F}(X)$; hence $\mathcal{F}(X)$ makes X measurable. If \mathcal{G} is any σ -field making X measurable, $X^{-1}(A) \in \mathcal{G}$ for all $A \in \mathcal{F}'$; hence $\mathcal{F}(X) \subset \mathcal{G}$.

(b) By 5.11.3, X is measurable relative to \mathcal{G} and \mathcal{F}' iff each X_j is measurable relative to \mathcal{G} and \mathcal{F}_j . The result follows from (a).

(c) Assume $Z: (\Omega, \mathcal{F}(X)) \rightarrow (\bar{R}, \mathcal{B})$. If Z is an indicator I_C , $C \in \mathcal{F}(X)$, then $C = X^{-1}(A)$ for some $A \in \mathcal{F}'$. If $f = I_A$, then $f \circ X = I_{\{X \in A\}} = I_C = Z$. If $Z = \sum_{k=1}^n z_k I_{C_k}$ is a finite-valued simple function and $I_{C_k} = f_k \circ X$ as above, then $Z = f \circ X$, where $f = \sum_{k=1}^n z_k f_k$.

In general, let Z_1, Z_2, \dots be finite-valued simple functions such that $Z_n \rightarrow Z$. We can express $Z_n = f_n \circ X$ as above; define $f = \lim_{n \rightarrow \infty} f_n$ where the limit exists, and 0 elsewhere. Then

$$Z(\omega) = \lim_{n \rightarrow \infty} Z_n(\omega) = \lim_{n \rightarrow \infty} f_n(X(\omega)) = f(X(\omega)).$$

The converse holds because a composition of measurable functions is measurable. ■

Note that 6.4.2(b) holds equally well for uncountable product spaces, because, as we observed at the time, 5.11.3 extends to arbitrary products. For an extension of 6.4.2(c), see Problem 1.

Now let us return to Eq. (3) at the beginning of this section:

$$\int_C h dP = \int_C Y dP, \quad C \in \mathcal{F}(X),$$

where $h = g \circ X$, $g(x) = E(Y|X=x)$. Since $g: (\Omega', \mathcal{F}') \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$, we have $h: (\Omega, \mathcal{F}(X)) \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$ by 6.4.2(c). This fact, along with (3), characterizes h , and gives us the concept of conditional expectation given a σ -field.

6.4.3 Theorem. Let Y be an extended random variable on (Ω, \mathcal{F}, P) , \mathcal{G} a sub σ -field of \mathcal{F} . Assume that $E(Y)$ exists. Then there is a function $E(Y|\mathcal{G}): (\Omega, \mathcal{G}) \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$, called the *conditional expectation of Y given \mathcal{G}* , such that

$$\int_C Y dP = \int_C E(Y|\mathcal{G}) dP \quad \text{for each } C \in \mathcal{G}.$$

Any two such functions must coincide a.e. [P]. [Note that we cannot simply set $E(Y|\mathcal{G}) = Y$ since $E(Y|\mathcal{G})$ is required to be measurable relative to \mathcal{G} .]

PROOF. Let $\lambda(C) = \int_C Y dP$, $C \in \mathcal{G}$. Then λ is a countably additive set function on \mathcal{G} , absolutely continuous with respect to P ; the result follows from the Radon–Nikodym theorem. ■

6.4.4 Comment. If $g(x) = E(Y|X=x)$ and $h(\omega) = g(X(\omega))$, then by 6.4.3, $h = E(Y|\mathcal{G})$, where $\mathcal{G} = \mathcal{F}(X)$; for convenience we shall usually write

$$h = E(Y|X).$$

[For example, if $E(Y|X=x) = x^2$, then $E(Y|X) = X^2$.]

We have seen that the conditional expectation $g(x) = E(Y|X=x)$, $x \in \Omega'$, can be transferred to Ω by forming $h(\omega) = g(X(\omega))$. Conversely, any conditional expectation $E(Y|\mathcal{G})$, \mathcal{G} an arbitrary sub σ -field of \mathcal{F} , arises from a random object X in this way. Simply take $X: (\Omega, \mathcal{F}) \rightarrow (\Omega, \mathcal{G})$ to be the identity map: $X(\omega) = \omega$, $\omega \in \Omega$. Then $X^{-1}(\mathcal{F}') = X^{-1}(\mathcal{G}) = \mathcal{G}$, so if $g(x) = E(Y|X=x)$, then $h = E(Y|\mathcal{F}(X)) = E(Y|\mathcal{G})$.

Now intuitively, $E(Y|\mathcal{G}) = E(Y|X)$ is the average value of Y , given that X is known. But what does it mean to “know” $X: (\Omega, \mathcal{F}) \rightarrow (\Omega, \mathcal{G})$? The events involving X are sets of the form $\{X \in G\}$, $G \in \mathcal{G}$, and since X is the identity map, $\{X \in G\} = G$. Since an event corresponds to a question that has a yes or no answer, $E(Y|\mathcal{G})$ may be interpreted as the average value of $Y(\omega)$, given that we know, for each $G \in \mathcal{G}$, whether or not $\omega \in G$. Some examples may help to make this clear.

6.4.5 Examples. (a) Let X be discrete, with values x_1, x_2, \dots ; take $\Omega' = \{x_1, x_2, \dots\}$, with \mathcal{F}' the class of all subsets of Ω' , and assume $P\{X = x_i\} > 0$ for all i . We have seen in 6.3.5(a) that

$$g(x_i) = E(Y|X=x_i) = \frac{1}{P\{X=x_i\}} \int_{\{X=x_i\}} Y dP.$$

Let $h = E(Y|X)$, that is, $h(\omega) = g(X(\omega))$. Then h has the constant value $g(x_i)$ on the set $\{X = x_i\}$, and $\mathcal{G} = X^{-1}(\mathcal{F})$ consists of all unions of the sets $\{X = x_i\}$. Knowledge of the value of $X(\omega)$ is equivalent to knowledge, for each $G \in \mathcal{G}$, of the membership or nonmembership of $\omega \in G$.

(b) Let X and Y be random variables with a joint density f . Let $\Omega = R^2$, $\mathcal{F} = \mathcal{B}(R^2)$, $P(B) = \iint_B f(x, y) dx dy$, $B \in \mathcal{F}$; $X(x, y) = x$, $Y(x, y) = y$. Take $\Omega' = R$, $\mathcal{F}' = \mathcal{B}(R)$. We have seen in 6.3.5(c) that $g(x) = E(Y|X = x) = \int_{-\infty}^{\infty} y h_0(y|x) dy$, where h_0 is the conditional density of Y given X . Let $h = E(Y|X)$, that is, $h(\omega) = g(X(\omega))$ or $h(x, y) = g(x)$. Thus $E(Y|X)$ is constant on vertical strips; also, $X^{-1}(\mathcal{F})$ consists of all sets $B \times R$, $B \in \mathcal{B}(R)$. Since $x \in B$ iff $(x, y) \in B \times R$, information about $X(\omega)$ [obtained intuitively by asking questions of the form “does $X(\omega)$ belong to B ?”] is equivalent to information about membership of ω in sets of \mathcal{G} .

We also have a general concept of conditional probability given a σ -field.

6.4.6 Theorem. Let (Ω, \mathcal{F}, P) be a probability space, \mathcal{G} a sub σ -field of \mathcal{F} ; fix $B \in \mathcal{F}$. There is a function $P(B|\mathcal{G}): (\Omega, \mathcal{G}) \rightarrow (R, \mathcal{B})$, called the *conditional probability of B given \mathcal{G}* , such that

$$P(C \cap B) = \int_C P(B|\mathcal{G}) dP \quad \text{for each } C \in \mathcal{G}.$$

Any two such functions must coincide a.e. [P], and in fact

$$P(B|\mathcal{G}) = E(I_B|\mathcal{G}) \quad \text{a.e. [P].}$$

PROOF. Let $\lambda(C) = P(C \cap B)$, $C \in \mathcal{G}$. Then λ is a countably additive finite-valued set function on \mathcal{G} , absolutely continuous with respect to P ; the existence and a.e. [P] uniqueness of $P(B|\mathcal{G})$ follow from the Radon-Nikodym theorem. (Since λ is finite, the range of $P(B|\mathcal{G})$ may be taken as R rather than \bar{R} .) The connection between conditional probability and conditional expectation follows from 6.4.3 with $Y = I_B$. ■

6.4.7 Comment. If $g(x) = P(B|X = x)$, X a random object, then $g(x) = E(I_B|X = x)$ a.e. [P_X] by 6.3.4. If $h(\omega) = g(X(\omega))$, then $h = E(I_B|\mathcal{F}(X))$ by 6.4.4, hence $h = P(B|\mathcal{F}(X))$ by 6.4.6. To summarize: If $g(x) = P(B|X = x)$ and $h(\omega) = g(X(\omega))$, then

$$h = P(B|\mathcal{G}), \mathcal{G} = \mathcal{F}(X).$$

For convenience we shall usually write

$$h = P(B|X).$$

Problem

1. Let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ and $Z: (\Omega, \mathcal{F}(X)) \rightarrow (\Omega'', \mathcal{F}'')$. We investigate conditions under which $Z = f \circ X$ for some $f: (\Omega', \mathcal{F}') \rightarrow (\Omega'', \mathcal{F}'')$. By 6.4.2(c), such an f can be found if $\Omega'' = \bar{R}$, $\mathcal{F}'' = \mathcal{B}(\bar{R})$.
 - (a) Assume that \mathcal{F}'' separates points; in other words, if $a, b \in \Omega''$, $a \neq b$, there are disjoint sets $A, B \in \mathcal{F}''$ such that $a \in A$, $b \in B$ (this will always hold if \mathcal{F}'' contains all singletons). Show that there is a function $f: \Omega' \rightarrow \Omega''$ (not necessarily measurable) such that $Z = f \circ X$.
 - (b) Assume that $Z = f \circ X$, where $f: \Omega' \rightarrow \Omega''$. If $X(\Omega) \in \mathcal{F}'$ and $f(\Omega' - X(\Omega))$ consists of a single point, show that f is measurable relative to \mathcal{F}' and \mathcal{F}'' .

Thus by (a) and (b), a measurable f such that $Z = f \circ X$ can be found if \mathcal{F}'' separates points and $X(\Omega) \in \mathcal{F}'$.

6.5 Properties of Conditional Expectation

The conditional expectation $E(Y|X = x)$ is a more intuitive object than the conditional expectation $E(Y|\mathcal{G})$; however, $E(Y|\mathcal{G})$ is often easier to handle in formal arguments, and is almost universally preferred in the probability literature. For convenience, we develop the basic properties of conditional expectation in pairs, one argument for $E(Y|\mathcal{G})$ and another (usually very similar) for $E(Y|X = x)$. Theorems about conditional probabilities are obtained by replacing Y by I_B , and results concerning $E(Y|X)$ are obtained by setting $\mathcal{G} = \mathcal{F}(X)$.

In the discussion to follow, Y, Y_1, Y_2, \dots are extended random variables on (Ω, \mathcal{F}, P) , with all expectations assumed to exist; $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ is a random object, and \mathcal{G} is a sub σ -field of \mathcal{F} . The phrase “a.e.” with no measure specified will always mean a.e. $[P]$. If $Z: (\Omega, \mathcal{G}) \rightarrow (\bar{R}, \mathcal{B})$, we say that Z is \mathcal{G} -measurable, and if $g: (\Omega', \mathcal{F}') \rightarrow (\bar{R}, \mathcal{B})$, we say that g is \mathcal{F}' -measurable.

6.5.1 Theorem. If Y is a constant k a.e., then

- (a) $E(Y|\mathcal{G}) = k$ a.e.
- (a') $E(Y|X = x) = k$ a.e. $[P_X]$.

If $Y_1 \leq Y_2$ a.e., then

- (b) $E(Y_1|\mathcal{G}) \leq E(Y_2|\mathcal{G})$ a.e.
- (b') $E(Y_1|X = x) \leq E(Y_2|X = x)$ a.e. $[P_X]$.

[A statement such as $E(Y_1|\mathcal{G}) \leq E(Y_2|\mathcal{G})$ a.e. means that if Z_j is a version of $E(Y_j|\mathcal{G})$, in other words, Z_j satisfies the defining requirement 6.4.3, then $Z_1 \leq Z_2$ a.e.]

- (c) $|E(Y|\mathcal{G})| \leq E(|Y| |\mathcal{G})$ a.e.
- (c') $|E(Y|X=x)| \leq E(|Y| |X=x)$ a.e. $[P_X]$.

PROOF. (a) The function constant at k is \mathcal{G} -measurable and

$$\int_C Y dP = \int_C k dP, \quad C \in \mathcal{G}.$$

- (a') If $g(x) \equiv k$, $x \in \Omega'$, then g is \mathcal{F}' -measurable and

$$\int_{\{X \in A\}} Y dP = \int_A k dP_X.$$

- (b) $\int_C Y_1 dP \leq \int_C Y_2 dP$; hence

$$\int_C E(Y_1|\mathcal{G}) dP \leq \int_C E(Y_2|\mathcal{G}) dP \quad \text{for each } C \in \mathcal{G}.$$

The result follows from 1.6.11.

- (b') $\int_{\{X \in A\}} Y_1 dP \leq \int_{\{X \in A\}} Y_2 dP$; hence

$$\int_A E(Y_1|X=x) dP_X \leq \int_A E(Y_2|X=x) dP_X \quad \text{for each } A \in \mathcal{F}'.$$

The result follows from 1.6.11.

Parts (c) and (c') follow from (b) and (b'), along with the observation that $-|Y| \leq Y \leq |Y|$. ■

We now prove an additivity theorem for conditional expectations.

6.5.2 Theorem. (a) If $a, b \in R$, and $aE(Y_1) + bE(Y_2)$ is well-defined (not of the form $\infty - \infty$), then $E(aY_1 + bY_2|\mathcal{G}) = aE(Y_1|\mathcal{G}) + bE(Y_2|\mathcal{G})$ a.e.

- (a') If $a, b \in R$ and $aE(Y_1) + bE(Y_2)$ is well-defined, then

$$E(aY_1 + bY_2|X=x) = aE(Y_1|X=x) + bE(Y_2|X=x) \quad \text{a.e. } [P_X].$$

PROOF. (a) If $C \in \mathcal{G}$,

$$\begin{aligned} \int_C (aY_1 + bY_2) dP &= \int_C aY_1 dP + \int_C bY_2 dP \\ &\quad \text{by the additivity theorem for integrals} \\ &= \int_C aE(Y_1|\mathcal{G}) dP + \int_C bE(Y_2|\mathcal{G}) dP \\ &\quad \text{by definition of conditional expectation.} \end{aligned}$$

Thus $\int_C aE(Y_1|\mathcal{G}) dP + \int_C bE(Y_2|\mathcal{G}) dP$ is well-defined, so again by the additivity theorem for integrals,

$$\int_C (aY_1 + bY_2) dP = \int_C [aE(Y_1|\mathcal{G}) + bE(Y_2|\mathcal{G})] dP,$$

as desired.

(a') This is done as in (a), with C replaced by $\{X \in A\}$ and

$$\int_C E(Y_j|\mathcal{G}) dP \quad \text{by} \quad \int_A E(Y_j|X=x) dP_X.$$

In the future, we shall dispose of proofs of this type with a phrase such as “same as (a).” ■

The monotone convergence theorem and the fact that a nonnegative series can be integrated term by term have exact analogs for conditional expectations.

6.5.3 Theorem. If $Y_n \geq 0$ for all n and $Y_n \uparrow Y$ a.e., then

- (a) $E(Y_n|\mathcal{G}) \uparrow E(Y|\mathcal{G})$ a.e.
- (a') $E(Y_n|X=x) \uparrow E(Y|X=x)$ a.e. [P_X].

If all $Y_n \geq 0$, then

- (b) $E(\sum_{n=1}^{\infty} Y_n|\mathcal{G}) = \sum_{n=1}^{\infty} E(Y_n|\mathcal{G})$ a.e.
- (b') $E(\sum_{n=1}^{\infty} Y_n|X=x) = \sum_{n=1}^{\infty} E(Y_n|X=x)$ a.e. [P_X].

In particular, if B_1, B_2, \dots are disjoint sets in \mathcal{F} ,

- (c) $P(\bigcup_{n=1}^{\infty} B_n|\mathcal{G}) = \sum_{n=1}^{\infty} P(B_n|\mathcal{G})$ a.e.
- (c') $P(\bigcup_{n=1}^{\infty} B_n|X=x) = \sum_{n=1}^{\infty} P(B_n|X=x)$ a.e. [P_X].

PROOF. (a) $\int_C Y_n dP = \int_C E(Y_n|\mathcal{G}) dP$, $C \in \mathcal{G}$; by 6.5.1(b), the $E(Y_n|\mathcal{G})$ increase to a \mathcal{G} -measurable function h . By the monotone convergence theorem, $\int_C Y dP = \int_C h dP$; hence $h = E(Y|\mathcal{G})$ a.e.

(a') Same as (a).

(b) By 6.5.2(a), $E(\sum_{k=1}^n Y_k|\mathcal{G}) = \sum_{k=1}^n E(Y_k|\mathcal{G})$ a.e. Let $n \rightarrow \infty$ and apply part (a) to obtain the desired result.

(b') Same as (b).

Finally, (c) is a special case of (b), and (c') of (b'). ■

If we take the expectation of a conditional expectation, the result is the same as if we were to take the expectation directly. This is actually a special case of the defining equations of Theorems 6.4.3 and 6.3.3.

6.5.4 Theorem. (a) $E[E(Y|\mathcal{G})] = E(Y)$; hence if Y is integrable, so is $E(Y|\mathcal{G})$.

$$(a') \quad \int_{\Omega'} E(Y|X=x) dP_X(x) = E(Y).$$

PROOF. (a) $\int_{\Omega} Y dP = \int_{\Omega} E(Y|\mathcal{G}) dP$.

$$(a') \quad \int_{\Omega'} E(Y|X=x) dP_X(x) = \int_{\{X \in \Omega'\}} Y dP = \int_{\Omega} Y dP. \blacksquare$$

We now prove the dominated convergence theorem for conditional expectations.

6.5.5 Theorem. If $|Y_n| \leq Z$ for all n , with $E(Z)$ finite, and $Y_n \rightarrow Y$ a.e., then

$$(a) \quad E(Y_n|\mathcal{G}) \rightarrow E(Y|\mathcal{G}) \text{ a.e.}$$

$$(a') \quad E(Y_n|X=x) \rightarrow E(Y|X=x) \text{ a.e. } [P_X].$$

PROOF. (a) Let $Z_n = \sup_{k \geq n} |Y_k - Y|$; then $Z_n \downarrow 0$ a.e. Now $E(Y_n|\mathcal{G})$ and $E(Y|\mathcal{G})$ are a.e. finite by 6.5.4(a), and $|E(Y_n|\mathcal{G}) - E(Y|\mathcal{G})| \leq E(|Y_n - Y| |\mathcal{G})$ by 6.5.1(c) and 6.5.2(a); this is less than or equal to $E(Z_n|\mathcal{G})$ by 6.5.1(b). Thus it suffices to show that $E(Z_n|\mathcal{G}) \rightarrow 0$ a.e. By 6.5.1(b), $E(Z_n|\mathcal{G}) \downarrow h$ a.e. for some \mathcal{G} -measurable function h . Since $0 \leq Z_n \leq 2Z$, which is integrable, we have

$$0 \leq \int_{\Omega} h dP \leq \int_{\Omega} E(Z_n|\mathcal{G}) dP = \int_{\Omega} Z_n dP \rightarrow 0$$

by the dominated convergence theorem. Thus $h = 0$ a.e., as desired.

$$(a') \quad \text{Same as (a).} \blacksquare$$

The extended monotone convergence theorem and Fatou's lemma may be proved for conditional expectations, as follows.

6.5.6 Theorem. Assume $Y_n \geq Z$ for all n , where $E(Z) > -\infty$.

$$(a) \quad \text{If } Y_n \uparrow Y \text{ a.e., then } E(Y_n|\mathcal{G}) \uparrow E(Y|\mathcal{G}) \text{ a.e.}$$

$$(a') \quad \text{If } Y_n \uparrow Y \text{ a.e., then } E(Y_n|X=x) \uparrow E(Y|X=x) \text{ a.e. } [P_X].$$

$$(b) \quad \liminf_{n \rightarrow \infty} E(Y_n|\mathcal{G}) \geq E(\liminf_{n \rightarrow \infty} Y_n|\mathcal{G}) \text{ a.e.}$$

$$(b') \quad \liminf_{n \rightarrow \infty} E(Y_n|X=x) \geq E(\liminf_{n \rightarrow \infty} Y_n|X=x) \text{ a.e. } [P_X].$$

Now assume $Y_n \leq Z$ for all n , where $E(Z) < +\infty$.

$$(c) \quad \text{If } Y_n \downarrow Y \text{ a.e., then } E(Y_n|\mathcal{G}) \downarrow E(Y|\mathcal{G}) \text{ a.e.}$$

$$(c') \quad \text{If } Y_n \downarrow Y \text{ a.e., then } E(Y_n|X=x) \downarrow E(Y|X=x) \text{ a.e. } [P_X].$$

$$(d) \quad \limsup_{n \rightarrow \infty} E(Y_n|\mathcal{G}) \leq E(\limsup_{n \rightarrow \infty} Y_n|\mathcal{G}) \text{ a.e.}$$

$$(d') \quad \limsup_{n \rightarrow \infty} E(Y_n|X=x) \leq E(\limsup_{n \rightarrow \infty} Y_n|X=x) \text{ a.e. } [P_X].$$

PROOF. (a) We have $0 \leq Y_n^+ \uparrow Y^+$; hence $E(Y_n^+ | \mathcal{G}) \rightarrow E(Y^+ | \mathcal{G})$ a.e. by 6.5.3(a). Also, $0 \leq Y_n^- \leq Z^-$, which is integrable; therefore $E(Y_n^- | \mathcal{G}) \rightarrow E(Y^- | \mathcal{G})$, a.e. by 6.5.5(a). Now

$$\begin{aligned} E(Y_n | \mathcal{G}) &= E(Y_n^+ | \mathcal{G}) - E(Y_n^- | \mathcal{G}) \quad \text{by 6.5.2(a)} \\ &\quad (\text{note that } Y_n^- \text{ is integrable}) \\ &\rightarrow E(Y^+ | \mathcal{G}) - E(Y^- | \mathcal{G}) = E(Y | \mathcal{G}) \quad \text{by 6.5.2(a)} \\ &\quad (\text{note that } Y^- \text{ is integrable}). \end{aligned}$$

(b) Let $Y'_n = \inf_{k \geq n} Y_k$; we then have $Y'_n \uparrow Y' = \liminf_{n \rightarrow \infty} Y_n$. By (a), $E(Y'_n | \mathcal{G}) \uparrow E(Y' | \mathcal{G})$ a.e. But $Y'_n \leq Y_n$, so

$$\begin{aligned} E(Y' | \mathcal{G}) &= \lim_{n \rightarrow \infty} E(Y'_n | \mathcal{G}) = \liminf_{n \rightarrow \infty} E(Y'_n | \mathcal{G}) \\ &\leq \liminf_{n \rightarrow \infty} E(Y_n | \mathcal{G}) \quad \text{by 6.5.1(b)}. \end{aligned}$$

(c) This follows from (a) upon replacing all extended random variables by their negatives.

$$\begin{aligned} (d) \quad E(\limsup_{n \rightarrow \infty} Y_n | \mathcal{G}) &= -E(\liminf_{n \rightarrow \infty} (-Y_n) | \mathcal{G}) \\ &\geq -\liminf_{n \rightarrow \infty} E(-Y_n | \mathcal{G}) \quad \text{by (b)} \\ &= \limsup_{n \rightarrow \infty} E(Y_n | \mathcal{G}). \end{aligned}$$

The proofs of (a') to (d') are the same as the proofs of (a) to (d). ■

Thus far we have examined $E(Y | \mathcal{G})$ and $E(Y | X = x)$ under various hypotheses on Y ; now we impose conditions on \mathcal{G} and X .

6.5.7 Theorem. (a) $E(Y | \{\emptyset, \Omega\}) = E(Y)$ a.e.

- (a') If X is a constant b a.e., then $E(Y | X = x) = E(Y)$ a.e. [P_x].
- (b) $E(Y | \mathcal{F}) = Y$ a.e.
- (b') If $X: (\Omega, \mathcal{F}) \rightarrow (\Omega, \mathcal{F})$ is the identity map, then $E(Y | X = \omega) = Y(\omega)$ a.e. [P].

PROOF. (a) $\int_C Y dP = \int_C E(Y) dP$ if $C = \emptyset$ or Ω .

- (a') If $b \in A$,

$$\int_{\{X \in A\}} X dP = \int_{\Omega} Y dP = E(Y) = \int_A E(Y) dP_X.$$

If $b \notin A$,

$$\int_{\{X \in A\}} Y dP = \int_{\emptyset} Y dP = 0 = \int_A E(Y) dP_X.$$

- (b) $\int_C Y dP = \int_C Y dP$, $C \in \mathcal{F}$, and Y is \mathcal{F} -measurable.
 (b') $\int_{\{X \in A\}} Y dP = \int_A Y dP$, and Y is $\mathcal{F}' (= \mathcal{F})$ -measurable. ■

The following result is preparatory to the next theorem.

6.5.8 Lemma. If $f: (\Omega, \mathcal{G}) \rightarrow (\bar{R}, \mathcal{B})$, μ is a measure on \mathcal{G} , and B is an atom of \mathcal{G} relative to μ ; that is, $B \in \mathcal{G}$, $\mu(B) > 0$, and if $A \in \mathcal{G}$, $A \subset B$, then $\mu(A) = 0$ or $\mu(B - A) = 0$, then f is a.e. constant on B .

PROOF. If $x \in \bar{R}$ and $\mu\{\omega \in B: f(\omega) < x\} = 0$, then $\mu\{\omega \in B: f(\omega) < y\} = 0$ for all $y \leq x$. Let $k = \sup\{x \in \bar{R}: \mu\{\omega \in B: f(\omega) < x\} = 0\}$. Then

$$\mu\{\omega \in B: f(\omega) < k\} = \mu\left[\bigcup_{\substack{r \text{ rational} \\ r < k}} \{\omega \in B: f(\omega) < r\}\right] = 0.$$

If $x > k$, then $\mu\{\omega \in B: f(\omega) < x\} > 0$; hence $\mu\{\omega \in B: f(\omega) \geq x\} = 0$ since B is an atom. Thus

$$\mu\{\omega \in B: f(\omega) > k\} = \mu\left[\bigcup_{\substack{r \text{ rational} \\ r > k}} \{\omega \in B: f(\omega) \geq r\}\right] = 0.$$

It follows that $f = k$ a.e. on B . ■

We now show that conditional expectation is an “averaging” or “smoothing” operation; if B is an atom of \mathcal{G} , $E(Y|\mathcal{G}) = k$ a.e. on B , where k is the average value of Y on B .

6.5.9 Theorem. (a) Let B be an atom of \mathcal{G} relative to P . Then

$$E(Y|\mathcal{G}) = \frac{1}{P(B)} \int_B Y dP = \frac{E(YI_B)}{P(B)} \quad \text{a.e. on } B.$$

As a special case, let B_1, B_2, \dots be disjoint sets in \mathcal{F} whose union is Ω , with $P(B_n) > 0$ for all n . Let \mathcal{G} be the minimal σ -field over the B_n , so that \mathcal{G} is the collection of all unions formed from the B_n . Then

$$E(Y|\mathcal{G}) = \frac{1}{P(B_n)} \int_{B_n} Y dP \quad \text{a.e. on } B_n, \quad n = 1, 2, \dots$$

(a') If $B = \{X = x_0\}$ and $P(B) > 0$, then

$$E(Y|X = x_0) = \frac{1}{P(B)} \int_B Y dP.$$

PROOF. (a) By 6.5.8, $E(Y|\mathcal{G})$ is a constant k a.e. on B . Since $\int_A Y dP = \int_A E(Y|\mathcal{G}) dP$ for all $A \in \mathcal{G}$, in particular for $A = B$, we have $\int_B Y dP = kP(B)$, as asserted.

(a') We first show that B is an atom of $\mathcal{G} = \mathcal{F}(X)$. For assume $A \in \mathcal{F}'$ and $X^{-1}(A) \subset B = X^{-1}\{x_0\}$. If $x_0 \in A$, then $X^{-1}\{x_0\} \subset X^{-1}(A)$; hence $X^{-1}(A) = B$. If $x_0 \notin A$, then $X^{-1}(A) \cap X^{-1}\{x_0\} = \emptyset$; but $X^{-1}(A) \subset X^{-1}\{x_0\}$, so $X^{-1}(A) = \emptyset$.

Now let $g(x) = E(Y|X = x)$, $h(\omega) = g(X(\omega)) = E(Y|\mathcal{G})(\omega)$. By (a), $h(\omega) = k = E(YI_B)/P(B)$ a.e. on B . If ω is any point of B such that $h(\omega) = k$, then $g(X(\omega)) = g(x_0) = k$, the desired result. ■

We now consider successive conditioning relative to two σ -fields, one of which is coarser than (that is, a subset of) the other. The result is that no matter which conditioning operation is applied first, the result is the same as the conditioning with respect to the coarser σ -field alone. This is intuitively reasonable; for example, to find the average value of a real-valued function f defined on $[0, 3]$, we may compute a_1 , the average of f on $[0, 1]$, and a_2 , the average of f on $[1, 3]$; the average of f on $[0, 3]$, namely, $\frac{1}{3} \int_0^3 f(x) dx$, is then $\frac{1}{3}a_1 + \frac{2}{3}a_2$.

6.5.10 Theorem. (a) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E[E(Y|\mathcal{G}_2)|\mathcal{G}_1] = E(Y|\mathcal{G}_1)$ a.e.

(a') If $f: (\Omega', \mathcal{F}') \rightarrow (\Omega'', \mathcal{F}'')$, then $E[E(Y|X)|f \circ X] = E(Y|f \circ X)$ a.e.

(b) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E[E(Y|\mathcal{G}_1)|\mathcal{G}_2] = E(Y|\mathcal{G}_1)$ a.e.

(b') If $f: (\Omega', \mathcal{F}') \rightarrow (\Omega'', \mathcal{F}'')$, then $E[E(Y|f \circ X)|X] = E(Y|f \circ X)$ a.e.

PROOF. (a) If $C \in \mathcal{G}_1$, then $\int_C E(Y|\mathcal{G}_1) dP = \int_C Y dP = \int_C E(Y|\mathcal{G}_2) dP$ since $C \in \mathcal{G}_2$. Thus $E[E(Y|\mathcal{G}_2)|\mathcal{G}_1] = E(Y|\mathcal{G}_1)$ a.e. Alternatively, if $C \in \mathcal{G}_1$, then $\int_C E[E(Y|\mathcal{G}_2)|\mathcal{G}_1] dP = \int_C E(Y|\mathcal{G}_2) dP = \int_C Y dP$ since $C \in \mathcal{G}_2$; thus $E(Y|\mathcal{G}_1) = E[E(Y|\mathcal{G}_2)|\mathcal{G}_1]$ a.e.

(a') Let $\mathcal{G}_2 = X^{-1}(\mathcal{F}'')$, $\mathcal{G}_1 = [f \circ X]^{-1}(\mathcal{F}'') = X^{-1}(f^{-1}(\mathcal{F}''))$, and apply (a).

(b) $E(Y|\mathcal{G}_1)$ is \mathcal{G}_1 -measurable, hence \mathcal{G}_2 -measurable, and

$$\int_C E(Y|\mathcal{G}_1) dP = \int_C E(Y|\mathcal{G}_2) dP, \quad C \in \mathcal{G}_2.$$

(b') Take \mathcal{G}_1 and \mathcal{G}_2 as in (a'), and apply (b). ■

If we take the conditional expectation of a product of two random variables, under certain conditions one of the terms can be factored out, as follows.

6.5.11 Theorem. (a) If Z is \mathcal{G} -measurable and both Y and YZ are integrable, then $E(YZ|\mathcal{G}) = ZE(Y|\mathcal{G})$ a.e. In particular, $E(Z|\mathcal{G}) = Z$ a.e.

(a') If $f: (\Omega', \mathcal{F}') \rightarrow (\bar{\mathcal{R}}, \mathcal{B})$ and both Y and $Y(f \circ X)$ are integrable, then $E(Y(f \circ X)|X = x) = f(x)E(Y|X = x)$ a.e. $[P_X]$. In particular,

$$E(f \circ X|X = x) = f(x) \quad \text{a.e.} \quad [P_X].$$

PROOF. (a) If Z is an indicator I_B , $B \in \mathcal{G}$, and $C \in \mathcal{G}$, we have

$$\begin{aligned} \int_C YZ dP &= \int_{C \cap B} Y dP = \int_{C \cap B} E(Y|\mathcal{G}) dP = \int_C I_B E(Y|\mathcal{G}) dP \\ &= \int_C ZE(Y|\mathcal{G}) dP, \end{aligned}$$

and $ZE(Y|\mathcal{G})$ is \mathcal{G} -measurable. Thus the result holds for indicators.

Now let Z be simple, say

$$Z = \sum_{j=1}^n z_j I_{B_j} \quad \text{with } B_j \in \mathcal{G}.$$

By 6.5.2(a),

$$E(YZ|\mathcal{G}) = \sum_{j=1}^n z_j E(YI_{B_j}|\mathcal{G}) = \sum_{j=1}^n z_j I_{B_j} E(Y|\mathcal{G}) = ZE(Y|\mathcal{G}).$$

If Z is an arbitrary \mathcal{G} -measurable function, let Z_1, Z_2, \dots be simple (and \mathcal{G} -measurable) with $|Z_n| \leq |Z|$ and $Z_n \rightarrow Z$. Now $E(YZ_n|\mathcal{G}) = Z_n E(Y|\mathcal{G})$ by what we have just proved, and $E(YZ_n|\mathcal{G}) \rightarrow E(YZ|\mathcal{G})$ by 6.5.5(a). (The integrability of YZ is used here.) Since Y is integrable, so is $E(Y|\mathcal{G})$; hence $E(Y|\mathcal{G})$ is finite a.e., and consequently $Z_n E(Y|\mathcal{G}) \rightarrow ZE(Y|\mathcal{G})$ a.e. [Note that, for example, $1/n \rightarrow 0$ but $(1/n)(\infty) \rightarrow 0(\infty) = 0$; thus finiteness of $E(Y|\mathcal{G})$ is important.] Therefore $E(YZ|\mathcal{G}) = ZE(Y|\mathcal{G})$.

(a') Let $f = I_B$, $B \in \mathcal{F}'$. Then

$$\begin{aligned} \int_{\{X \in A\}} Y(f \circ X) dP &= \int_{\{X \in A\}} YI_{\{X \in B\}} dP = \int_{\{X \in A \cap B\}} Y dP \\ &= \int_{A \cap B} E(Y|X = x) dP_X(x) = \int_A f(x)E(Y|X = x) dP_X(x). \end{aligned}$$

Thus the result holds when f is an indicator. Passage to simple functions and then to arbitrary measurable functions is carried out just as in (a). ■

6.5.12 Comments. (a) Theorem 6.5.11(a) holds under the weaker assumption that $E(Y)$ and $E(YZ)$ exist [and 6.5.11(b) under the assumption that $E(Y)$ and $E[Y(f \circ X)]$ exist]; see problem 4.

(b) Intuitively, if it is known that $X = x$, then $f \circ X$ may be treated as a constant $f(x)$, so that $E(Y(f \circ X)|X = x)$ should be $f(x)E(Y|X = x)$ as in

6.5.11(b). A similar interpretation may be given to 6.5.11(a) as follows. We can express $E(YZ|\mathcal{G})$ as $E(YZ|\mathcal{F}(X)) = E(YZ|X)$ for an appropriate random object X (see 6.4.4). Since Z is $\mathcal{F}(X)$ -measurable, $Z = f \circ X$ for some $f: (\Omega', \mathcal{F}') \rightarrow (\bar{\Omega}, \mathcal{B})$ by 6.4.2(c). Thus if X is given, Z may be treated as a constant and factored out.

Problems

1. (a) If X is a random object, Y a random variable such that $E(Y)$ exists, and X and Y are independent, show that $E(Y|X) = E(Y)$ a.e. [P], and $E(Y|X = x) = E(Y)$ a.e. [P_X].
 (b) Give an example of integrable random variables X and Y such that $E(Y|X) = E(Y)$, but X and Y are not independent.
2. If Y is an integrable random variable and X and Z are random objects, show that if (X, Y) and Z are independent, then $E(Y|X, Z) = E(Y|X)$.
3. This problem illustrates how to obtain a theorem about $E(Y|X = x)$ from a corresponding theorem about $E(Y|X)$ without writing a separate proof or saying “similarly.” As a typical example, suppose we have proved that if $Y_1 \leq Y_2$ a.e., then $E(Y_1|X) \leq E(Y_2|X)$ a.e. Show that

$$P\{E(Y_1|X) > E(Y_2|X)\} = P_X\{E(Y_1|X = x) > E(Y_2|X = x)\}.$$

Conclude that if $Y_1 \leq Y_2$ a.e., then $E(Y_1|X = x) \leq E(Y_2|X = x)$ a.e. [P_X].

4. Extend Theorem 6.5.11 to the case where $E(Y)$ and $E(YZ)$ exist but are not necessarily finite.
5. Let (Ω, \mathcal{F}, P) be a probability space, and \mathcal{G} a sub σ -field of \mathcal{F} . If $Y \in L^1(\Omega, \mathcal{F}, P)$, then $E(Y|\mathcal{G})$ is also in $L^1(\Omega, \mathcal{F}, P)$, by 6.5.4(a). Thus $A(Y) = E(Y|\mathcal{G})$ defines a linear operator on L^1 , and A may be transferred to the Banach space $L^1(\Omega, \mathcal{F}, P)$.
 - (a) Show that $\|A\| = 1$.
 - (b) Define $\langle X, Y \rangle = \int_{\Omega} XY dP$, $X \in L^1$, $Y \in L^\infty$. Show that A has the “self-adjointness” property $\langle AX, Y \rangle = \langle X, AY \rangle$. [Note that $L^\infty \subset L^1$, and if $Y \in L^\infty$, then $AY \in L^\infty$, so $\langle X, AY \rangle$ is well defined.]

6.6 Regular Conditional Probabilities

We have seen that if B_1, B_2, \dots are disjoint sets in \mathcal{F} , $n = 1, 2, \dots$, then

$$P\left(\bigcup_{n=1}^{\infty} B_n | \mathcal{G}\right) = \sum_{n=1}^{\infty} P(B_n | \mathcal{G}) \quad \text{a.e.}$$

This does not imply that we will be able to choose $P(B|\mathcal{G})(\omega)$, $B \in \mathcal{F}$, $\omega \in \Omega$, so that it is a measure in B for all (or almost all) $\omega \in \Omega$. To clarify the problem,

suppose that for each $B \in \mathcal{F}$, we choose a particular version of the conditional probability $P(B|\mathcal{G})(\omega)$, $\omega \in \Omega$; we now have a number $P(B|\mathcal{G})(\omega)$ for each $B \in \mathcal{F}$ and $\omega \in \Omega$. The difficulty is that for a fixed ω , $P(\cdot|\mathcal{G})(\omega)$ need not be countably additive on \mathcal{F} . For suppose that B_1, B_2, \dots are disjoint sets in \mathcal{F} . Then

$$P\left[\bigcup_{n=1}^{\infty} B_n|\mathcal{G}\right](\omega) = \sum_{n=1}^{\infty} P(B_n|\mathcal{G})(\omega)$$

except for ω belonging to a set $N(B_1, B_2, \dots)$ of probability 0. Thus the set of ω 's for which countable additivity fails is

$$M = \bigcup\{N(B_1, B_2, \dots) : B_1, B_2, \dots \text{ disjoint sets in } \mathcal{F}\}.$$

In general, M is an uncountable union of sets of probability 0, and therefore M need not have probability 0 (or even be in \mathcal{F}). Thus there is no guarantee that we can specify the $P(B|\mathcal{G})$ to be countably additive in B . [Similarly, there is no guarantee that $P(B|X=x)$ will be countably additive in B for P_X -almost all $x \in \Omega'$.]

We are going to establish conditions under which the countable additivity requirement can be met.

6.6.1 Definition. Let Y be a random variable on (Ω, \mathcal{F}, P) , \mathcal{G} a sub σ -field of \mathcal{F} . The function $F = F(\omega, y)$, $\omega \in \Omega$, $y \in R$, is called a *regular conditional distribution function* for Y given \mathcal{G} iff the following two conditions are satisfied.

1. For each ω , $F(\omega, \cdot)$ is a proper distribution function of R , that is, increasing and right-continuous, with $F(\omega, \infty) = 1$, $F(\omega, -\infty) = 0$.
2. For each y , $F(\omega, y) = P\{Y \leq y|\mathcal{G}\}(\omega)$ for almost every ω .

The key step in the development is the result that a regular conditional distribution function for a given random variable always exists.

6.6.2 Theorem. If Y is a random variable on (Ω, \mathcal{F}, P) , \mathcal{G} a sub σ -field of \mathcal{F} , there is always a regular conditional distribution function for Y given \mathcal{G} .

PROOF. Select a version of $F_r(\omega) = P\{Y \leq r|\mathcal{G}\}(\omega)$ for each fixed rational r . If r_1, r_2, \dots is an enumeration of the rationals, let

$$A_{ij} = \{\omega : F_{r_j}(\omega) < F_{r_i}(\omega)\}, \quad A = \bigcup\{A_{ij} : r_i < r_j\}.$$

Then $P(A) = 0$ since $r_i < r_j$ implies $P\{Y \leq r_i|\mathcal{G}\} \leq P\{Y \leq r_j|\mathcal{G}\}$ a.e. by 6.5.1 (b).

Now let

$$B_i = \left\{\omega : \lim_{n \rightarrow \infty} F_{r_i + (1/n)}(\omega) \neq F_{r_i}(\omega)\right\}, \quad B = \bigcup_{i=1}^{\infty} B_i.$$

Since

$$I_{\{Y \leq r_i + (1/n)\}} \downarrow I_{\{Y \leq r_i\}} \quad \text{as } n \rightarrow \infty,$$

6.5.5(a) yields

$$F_{r_i + (1/n)}(\omega) \rightarrow F_{r_i}(\omega) \quad \text{a.e.};$$

hence $P(B) = 0$.

If $C = \{\omega: \lim_{n \rightarrow \infty} F_n(\omega) \neq 1\}$, then $P(C) = 0$ since $\{Y \leq n\} \uparrow \Omega$, so that $P\{Y \leq n | \mathcal{G}\} \rightarrow 1$ a.e. Similarly, $D = \{\omega: \lim_{n \rightarrow -\infty} F_n(\omega) \neq 0\}$ has probability 0.

Define

$$F(\omega, y) = \begin{cases} \lim_{r \rightarrow y^+} F_r(\omega) & \text{if } \omega \notin A \cup B \cup C \cup D \\ \text{any proper distribution function } G(y) & \text{if } \omega \in A \cup B \cup C \cup D. \end{cases}$$

Then F is well defined, for if $\omega \notin A$, then $F_r(\omega)$ is monotone in r , so that $\lim_{r \rightarrow y^+} F_r(\omega)$ exists. Note also that if $\omega \notin A \cup B$, then $\lim_{r \rightarrow y^+} F_r(\omega) = F_y(\omega)$ if y is rational, so that in this case, $F(\omega, y) = F_y(\omega)$. Similarly, if $\omega \notin A \cup C \cup D$, then $\lim_{r \rightarrow \infty} F_r(\omega) = 1$, $\lim_{r \rightarrow -\infty} F_r(\omega) = 0$.

We show that F is a regular conditional distribution function for Y given \mathcal{G} . Fix $\omega \notin A \cup B \cup C \cup D$; then $F(\omega, \cdot)$ is clearly increasing. If $y < y' \leq r$, then $F(\omega, y) \leq F(\omega, y') \leq F(\omega, r) = F_r(\omega) \rightarrow F(\omega, y)$ as $r \rightarrow y$. Thus $F(\omega, \cdot)$ is right-continuous. If $r \leq y$, then $F(\omega, y) \geq F(\omega, r) = F_r(\omega) \rightarrow 1$ as $r \rightarrow \infty$; hence $F(\omega, y) \rightarrow 1$ as $y \rightarrow \infty$; similarly, $F(\omega, y) \rightarrow 0$ as $y \rightarrow -\infty$. Thus the first requirement is satisfied.

Now $P\{Y \leq r | \mathcal{G}\}(\omega) = F_r(\omega) = F(\omega, r)$ by construction of F . As $r \downarrow y$, $F(\omega, r) \rightarrow F(\omega, y)$ for all ω by right-continuity, and

$$P\{Y \leq r | \mathcal{G}\} \rightarrow P\{Y \leq y | \mathcal{G}\} \quad \text{a.e.}$$

by 6.5.5(a). Thus $P\{Y \leq y | \mathcal{G}\}(\omega) = F(\omega, y)$ for almost every ω (y fixed), establishing the second requirement. ■

We are going to show that if Y is a random variable, $P\{Y \in B | \mathcal{G}\}$ can be chosen so as to be countably additive in B . This will follow from 6.6.2 and the fact that a distribution function determines a unique Lebesgue–Stieltjes measure.

6.6.3 Definition. Let $Y: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ be a random object, and \mathcal{G} a sub σ -field of \mathcal{F} . The function $Q: \Omega \times \mathcal{F}' \rightarrow [0, 1]$ is called a *regular conditional probability* for Y given \mathcal{G} iff

- (1) $Q(\omega, B)$ is a probability measure in B for each fixed $\omega \in \Omega$, and
- (2) for each fixed $B \in \mathcal{F}'$, $Q(\omega, B) = P\{Y \in B | \mathcal{G}\}(\omega)$ a.e.

If Y is a random variable, so that $\Omega' = R$, $\mathcal{F}' = \mathcal{B}(R)$, a regular conditional probability for Y given \mathcal{G} always exists.

6.6.4 Theorem. Let Y be a random variable on (Ω, \mathcal{F}, P) , \mathcal{G} a sub σ -field of \mathcal{F} . There exists a regular conditional probability for Y given \mathcal{G} .

PROOF. Let F be a regular conditional distribution function for Y given \mathcal{G} . Define

$$Q(\omega, B) = \int_{y \in B} dF(\omega, y).$$

Thus for each ω , $Q(\omega, \cdot)$ is the Lebesgue–Stieltjes measure corresponding to $F(\omega, \cdot)$; hence Q is a probability measure in B if ω is fixed.

Now let $\mathcal{C} = \{B \in \mathcal{B}(R) : Q(\omega, B) = P\{Y \in B | \mathcal{G}\}(\omega) \text{ a.e.}\}$. Then \mathcal{C} contains all intervals $(-\infty, y]$ since $F(\omega, y) = P\{Y \leq y | \mathcal{G}\}(\omega)$ a.e. If $A, B \in \mathcal{C}$, $A \subset B$, then $B - A \in \mathcal{C}$, and it follows that \mathcal{C} contains all intervals $(a, b]$, hence all finite disjoint unions of right-semiclosed intervals. By the monotone class theorem, $\mathcal{C} = \mathcal{B}(R)$. Thus Q is a regular conditional probability for Y given \mathcal{G} . ■

We now extend this result to objects Y more general than random variables.

6.6.5 Theorem. Let $Y: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ be a random object, and \mathcal{G} a sub σ -field of \mathcal{F} . Suppose there is a map $\Psi: (\Omega', \mathcal{F}') \rightarrow (R, \mathcal{B}(R))$ such that Ψ is one-to-one, $E = \Psi(\Omega')$ is a Borel subset of R , and Ψ^{-1} is measurable, that is, $\Psi^{-1}: (E, \mathcal{B}(E)) \rightarrow (\Omega', \mathcal{F}')$. Then there is a regular conditional probability for Y given \mathcal{G} .

PROOF. Let $Q_0 = Q_0(\omega, B)$, $B \in \mathcal{B}(R)$, $\omega \in \Omega$, be a regular conditional probability for the random variable $\Psi(Y)$ given \mathcal{G} . Define $Q(\omega, A) = Q_0(\omega, \Psi(A))$, $A \in \mathcal{F}'$; since Ψ^{-1} is measurable, $\Psi(A) \in \mathcal{B}(E) \subset \mathcal{B}(R)$, and Q is well-defined. Now Q is a probability measure in A for ω fixed, and if A is fixed, then

$$Q(\omega, A) = P\{\Psi(Y) \in \Psi(A) | \mathcal{G}\}(\omega) = P\{Y \in A | \mathcal{G}\}(\omega) \quad \text{a.e.} \blacksquare$$

A map Ψ of the type described in 6.6.5 is called a *Borel equivalence* of Ω' and E . For our purposes, the most important result concerning Borel equivalences is the following theorem, whose proof is given in Section 4.4, Problem 8.

6.6.6 Theorem. Let Ω' be a complete separable metric space, \mathcal{F}' the Borel subsets of Ω' , that is, the smallest σ -field containing the open subsets of Ω' . Then Ω' is Borel equivalent to a Borel subset E of $[0, 1]$.

Thus there is a regular conditional probability given \mathcal{G} for any random object with values in a complete separable metric space Ω' . In particular, we may take $\Omega' = R^n$, $\mathcal{F}' = \mathcal{B}(R^n)$, so there is a regular conditional probability given \mathcal{G} for any random vector. Similarly, we may take $\Omega' = R^\infty$, $\mathcal{F}' = \mathcal{B}(R^\infty) = [\mathcal{B}(R)]^\infty$ (see 4.4.2), and conclude that there is a regular conditional probability given \mathcal{G} for any random sequence.

Note also that if Ω itself is a complete separable metric space, and \mathcal{F} is the smallest σ -field containing the open sets, we may take Y as the identity map on Ω . By Theorems 6.6.5 and 6.6.6, we can find a function $Q = Q(\omega, B)$ that is a probability measure on \mathcal{F} for each ω , and $Q(\omega, B) = P(B | \mathcal{G})(\omega)$ a.e. for each fixed $B \in \mathcal{F}$.

A similar analysis, with only notational changes, shows that if $Y: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ is a random object with values in a complete separable metric space Ω' , with \mathcal{F}' generated by the open sets, and X is an arbitrary random object, we can find a function $Q = Q(x, B)$ that is a probability measure on \mathcal{F}' for each fixed x , and $Q(x, B) = P\{Y \in B | X = x\}$ a.e. $[P_x]$ for each fixed $B \in \mathcal{F}'$.

Problems

1. Let $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ and $Y: (\Omega, \mathcal{F}) \rightarrow (\Omega'', \mathcal{F}'')$ be random objects on (Ω, \mathcal{F}, P) , and assume that $P_x(B) = P\{Y \in B | X = x\}$ can be chosen so as to be a probability measure in B for each fixed x . If $g: (\Omega' \times \Omega'', \mathcal{F}' \times \mathcal{F}'') \rightarrow (\bar{R}, \mathcal{B})$ and $E[g(X, Y)]$ exists, show that
 - (a) $E[g(X, Y) | X = x] = \int_{\Omega''} g(x, y) dP_x(y)$ a.e. $[P_x]$.
 In particular, if $C \in \mathcal{F}' \times \mathcal{F}''$, then
 - (b) $P\{(X, Y) \in C | X = x\} = P_x(C(x))$ a.e. $[P_x]$.
 Conclude that
 - (c) $P\{(X, Y) \in C\} = \int_{\Omega'} P_x(C(x)) dP_X(x)$.
2. Let Y_1, \dots, Y_n be independent random variables, each uniformly distributed between 0 and 1. Let Z_k be the product $Y_1 \cdots Y_k$, $1 \leq k \leq n$. Show that, given $Z_1 = z_1, \dots, Z_k = z_k$, Z_{k+1} is uniformly distributed between 0 and z_k , that is,

$$P\{Z_{k+1} \in B | Z_1 = z_1, \dots, Z_k = z_k\} = \int_B g_k(z) dz,$$

where $g_k(z) = 1/z_k$, $0 \leq z \leq z_k$; $g_k(z) = 0$ elsewhere. (This is another way of looking at Example 6.2.2.)

3. The following result is preparatory to the next problem; it is adapted from Halmos (1950).
 - (a) Let E be a Lebesgue measurable subset of R with $0 < \mu(E) < \infty$, where μ is Lebesgue measure. If $0 < \delta < 1$, show that there is an open interval I such that $\mu(E \cap I) \geq \delta \mu(I)$.

- (b) With E as in part (a), let $D(E) = \{x - y: x, y \in E\}$. Show that $D(E)$ includes a neighborhood of 0. [This holds also if $\mu(E) = \infty$ since $0 < \mu(E \cap [-n, n]) < \infty$ for some n .]
- (c) Let ζ be an irrational number, and let $A = \{m + n\zeta: m, n \text{ integers}\}$. Show that A is dense in R . Equivalently, the set of numbers $n\zeta$, n an integer, reduced modulo 1, is dense in $[0, 1]$. If $[0, 1]$ is identified with the unit circle under the correspondence $\theta \rightarrow e^{i2\pi\theta}$, the problem is as follows. If $\alpha/2\pi$ is irrational, the set $\{e^{inx}: n \text{ an integer}\}$ is dense in the circle. In fact more can be proved. Let z_0 be any point on the circle and let $z_n = e^{inx}z_0 = z_0$ rotated by $n\alpha$, $n = 1, 2, \dots$. If z is an arbitrary point on the circle and $\varepsilon > 0$, $|z_n - z| < \varepsilon$ for infinitely many values of n .
- (d) If ζ is irrational, let $B = \{m + n\zeta: n \text{ an integer, } m \text{ an even integer}\}$, $C = \{m + n\zeta: n \text{ an integer, } m \text{ an odd integer}\}$. Show that B and C are dense in R .
- (e) Define an equivalence relation on R by $x \sim y$ iff $x - y \in A$, where A is as defined in (c). Form a set E_0 by selecting one point from each distinct equivalence class. Show that E_0 is not a Lebesgue measurable set. [If F is a Borel set with $F \subset E_0$, show that $\mu(F) = 0$. Then show that R is a disjoint union of the sets $E_0 + a$, $a \in A$. Use the translation-invariance of Lebesgue measure to show that E_0 is not Lebesgue measurable.]
- (f) Let $M = \{x + y: x \in E_0, y \in B\}$, $M' = \{x + y: x \in E_0, y \in C\}$. Show that $R = M \cup M'$, and any Borel subset of M or of M' has Lebesgue measure 0.
- (g) Let E be an arbitrary Lebesgue measurable subset of R . Show that if F is a Borel subset of $E \cap M$, then $\mu(F) = 0$, and if G is a Borel set such that $E \cap M \subset G \subset E$, then $\mu(E - G) = 0$.
4. Let H be a subset of $[0, 1]$ with inner Lebesgue measure 0 [$\sup\{\mu(B): B \text{ a Borel subset of } H\} = 0$, $\mu = \text{Lebesgue measure}$] and outer Lebesgue measure 1 [$\inf\{\mu(B): B \text{ a Borel overset of } H\} = 1$]. To construct such a set, take $E = [0, 1]$, $H = E \cap M$ in Problem 3(g).
- Let $\Omega = [0, 1]$, and let \mathcal{F} consist of all sets $(B_1 \cap H) \cup (B_2 \cap H^c)$, where B_1, B_2 are Borel subsets of $[0, 1]$. Define

$$P[(B_1 \cap H) \cup (B_2 \cap H^c)] = \frac{1}{2}(\mu(B_1) + \mu(B_2)).$$

Thus $P = \mu$ on $\mathcal{B}[0, 1]$ [if $B \in \mathcal{B}[0, 1]$, $B = (B \cap H) \cup (B \cap H^c)$]. Take $Y(\omega) = \omega$, $\omega \in \Omega$.

- (a) Show that P is well-defined, that is, if $(B_1 \cap H) \cup (B_2 \cap H^c) = (B_1' \cap H) \cup (B_2' \cap H^c)$, then $\mu(B_1) = \mu(B_1')$, $\mu(B_2) = \mu(B_2')$.
- (b) Suppose that Q is a regular conditional probability for Y given $\mathcal{G} = \mathcal{B}[0, 1]$. Show that $Q(\omega, H) = Q(\omega, H^c) = \frac{1}{2}$ a.e.

- (c) If $B \in \mathcal{G}$, show that $Q(\omega, B) = I_B(\omega)$ a.e.
- (d) Show that $Q(\omega, \{\omega\}) = 1$ for almost every ω , and thus arrive at a contradiction.

Conclude that there is no regular conditional probability for Y given \mathcal{G} .

6.7 References

Books that discuss martingale theory must also treat conditional probability and expectation. Thus for references on the material of this chapter, see Section 7.10.

7

Strong Laws of Large Numbers and Martingale Theory

7.1 Introduction

At the end of Chapter 5, we indicated that the physical fact of convergence of the relative frequency of heads in coin tossing is best expressed as a statement about almost everywhere convergence of S_n/n , where S_n is a sum of independent random variables X_1, X_2, \dots, X_n ; we attached the name “strong law of large numbers” to such a result. Now a “strong law of large numbers” in the most general sense is any statement about the almost everywhere convergence of a sequence of random variables, and this is the main subject matter of this chapter. A large class of convergence theorems will be developed with the aid of martingale theory, but before going into this it will be useful to consider the classical approach.

First we prove some results from real analysis that will be needed.

7.1.1 Lemma. Let $A = [a_{ij}]$ be an infinite matrix of real numbers; assume that $a_{nj} \rightarrow 0$ as $n \rightarrow \infty$ for each fixed j , and that for some nonnegative real number c , $\sum_{j=1}^{\infty} |a_{nj}| \leq c$ for all n . If $\{x_n\}$ is a bounded sequence of real numbers, define

$$y_n = \sum_{j=1}^{\infty} a_{nj} x_j, \quad n = 1, 2, \dots.$$

Then:

- (a) If $x_n \rightarrow 0$, then $y_n \rightarrow 0$.
- (b) If $\sum_{j=1}^{\infty} a_{nj} \rightarrow 1$ and $x_n \rightarrow x$ (x real), then $y_n \rightarrow x$.

PROOF. (a) We may write

$$|y_n| \leq \sum_{j=1}^N |a_{nj}| |x_j| + \sum_{j=N+1}^{\infty} |a_{nj}| |x_j|. \quad (1)$$

Given $\varepsilon > 0$, choose N so that $|x_j| \leq \varepsilon/c$ for $j > N$; the second term on the right-hand side of (1) is at most $c(\varepsilon/c) = \varepsilon$. Since the first term approaches 0 as $n \rightarrow \infty$ for any fixed N , it follows that $y_n \rightarrow 0$.

- (b) By (a), $\sum_{j=1}^{\infty} a_{nj}(x_j - x) \rightarrow 0$, and the result follows. ■

7.1.2 Toeplitz Lemma. Let $\{a_n\}$ be a sequence of nonnegative real numbers, and let $b_n = \sum_{j=1}^n a_j$; assume $b_n > 0$ for all n , and $b_n \rightarrow \infty$ as $n \rightarrow \infty$. If $\{x_n\}$ is a sequence of real numbers converging to the real number x , then

$$\frac{1}{b_n} \sum_{j=1}^n a_j x_j \rightarrow x.$$

PROOF. Form an infinite matrix A whose n th row is

$$\left(\begin{array}{cccccc} \frac{a_1}{b_n} & \frac{a_2}{b_n} & \dots & \frac{a_n}{b_n} & 0 & 0 & \dots \end{array} \right)$$

and apply 7.1.1(b). ■

7.1.3 Kronecker Lemma. Let $\{b_n\}$ be an increasing sequence of positive real numbers with $b_n \rightarrow \infty$, and let $\{x_n\}$ be a sequence of real numbers with $\sum_{n=1}^{\infty} x_n = x$ (finite). Then

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. If $s_n = \sum_{j=1}^n x_j$ (with $s_0 = 0$),

$$\begin{aligned} \sum_{j=1}^n b_j x_j &= \sum_{j=1}^n b_j(s_j - s_{j-1}) \\ &= b_n s_n - b_0 s_0 - \sum_{j=1}^{n-1} s_{j-1}(b_j - b_{j-1}), \end{aligned}$$

taking $b_0 = 0$. (This is the “summation by parts” formula; it is proved by brute force.) Thus

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j = s_n - \frac{1}{b_n} \sum_{j=1}^{n-1} a_j s_{j-1},$$

where $a_j = b_j - b_{j-1} \geq 0$. Since $s_n \rightarrow x$ as $n \rightarrow \infty$, and $s_{j-1} \rightarrow x$ as $j \rightarrow \infty$,

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j \rightarrow 0 \quad \text{by 7.1.2. } \blacksquare$$

Now if S_n is a random variable with finite expectation, Chebyshev's inequality implies that

$$P\{|S_n - E(S_n)| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} E[(S_n - E(S_n))^2] = \frac{\text{Var } S_n}{\varepsilon^2}.$$

If in fact S_n is a sum of independent random variables, this result can be strengthened considerably, as follows.

7.1.4 Kolmogorov's Inequality. Let X_1, \dots, X_n be independent random variables with finite expectation, and let $S_j = X_1 + \dots + X_j$, $j = 1, \dots, n$. Then for any $\varepsilon > 0$,

$$P\left\{\max_{1 \leq j \leq n} |S_j - E(S_j)| \geq \varepsilon\right\} \leq \frac{\text{Var } S_n}{\varepsilon^2}.$$

PROOF. We may assume without loss of generality that $E(S_j) \equiv 0$. Let

$$A_k = \{|S_j| < \varepsilon, j = 1, \dots, k-1, |S_k| \geq \varepsilon\},$$

$$A = \left\{\max_{1 \leq j \leq n} |S_j| \geq \varepsilon\right\};$$

A is the disjoint union of the A_k , $k = 1, \dots, n$. Now

$$\text{Var } S_n = \int_{\Omega} S_n^2 dP \geq \int_A S_n^2 dP = \sum_{k=1}^n \int_{A_k} S_n^2 dP. \quad (1)$$

But $S_n = S_k + Y_k$, where $Y_k = X_{k+1} + \dots + X_n$; hence

$$\int_{A_k} S_n^2 dP = \int_{A_k} S_k^2 dP + 2 \int_{A_k} S_k Y_k dP + \int_{A_k} Y_k^2 dP. \quad (2)$$

The second term on the right-hand side of (2) is $2E[I_{A_k} S_k Y_k]$, which is 0 since $I_{A_k} S_k$ and Y_k , being functions of independent random variables, are independent [see 5.8.2(d) and 5.10.8]. Since the third term of the right-hand side of (2) is nonnegative, we have

$$\int_{A_k} S_n^2 dP \geq \int_{A_k} S_k^2 dP \geq \varepsilon^2 P(A_k) \quad \text{by definition of } A_k.$$

By (1),

$$\text{Var } S_n \geq \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A). \blacksquare$$

We shall use the Borel–Cantelli lemma (2.2.4) quite often: If A_1, A_2, \dots are events such that $\sum_n P(A_n) < \infty$, then $\limsup_n A_n$ has probability 0. There is a partial converse which will also be needed.

7.1.5 Second Borel–Cantelli Lemma. Let (Ω, \mathcal{F}, P) be a probability space, and let A_1, A_2, \dots be independent events in \mathcal{F} . If $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(\limsup_n A_n) = 1$.

PROOF.

$$P\left(\limsup_n A_n\right) = P\left(\bigcap_n \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} P\left(\bigcup_{k=n}^m A_k\right).$$

Now

$$\begin{aligned} P\left(\bigcup_{k=n}^m A_k\right)^c &= P\left(\bigcap_{k=n}^m A_k^c\right) \\ &= \prod_{k=n}^m P(A_k^c) \quad \text{by independence} \\ &\leq \prod_{k=n}^m \exp[-P(A_k)] \quad \text{since } P(A_k^c) = 1 - P(A_k) \leq \exp[-P(A_k)] \\ &\rightarrow 0 \quad \text{as } m \rightarrow \infty \quad \text{since } \sum P(A_k) = \infty. \blacksquare \end{aligned}$$

Problems

1. (*Extension of the second Borel–Cantelli lemma*). Let A_1, A_2, \dots be events in a given probability space such that $\sum_{n=1}^{\infty} P(A_n) = \infty$ and

$$\liminf_{n \rightarrow \infty} \frac{\sum_{j,k=1}^n P(A_j \cap A_k)}{\left(\sum_{j=1}^n P(A_j)\right)^2} = 1.$$

- (a) Show that the lim inf condition above is satisfied if the A_n are *pairwise independent* (A_j and A_k are independent whenever $j \neq k$) and $\sum_{n=1}^{\infty} P(A_n) = \infty$.
- (b) Use Chebyshev's inequality to show that if $I_n = I_{A_n}$, then

$$\liminf_{n \rightarrow \infty} P\left\{ \left| \sum_{k=1}^n I_k - \sum_{k=1}^n P(A_k) \right| > \frac{1}{2} \sum_{k=1}^n P(A_k) \right\} = 0.$$

- (c) Conclude from (b) that there is a sequence of integers $n_1 < n_2 < \dots$ such that with probability 1,

$$\sum_{k=1}^{n_j} I_k \geq \frac{1}{2} \sum_{k=1}^{n_j} P(A_k) \quad \text{for sufficiently large } j.$$

- (d) Show that $P(\limsup_n A_n) = 1$.
 2. Let θ be uniformly distributed on $[0, 2\pi]$ and define $X_k = \sin k\theta$ ($k = 1, 2, \dots$). Show that

$$\frac{X_1 + \dots + X_n}{n} \rightarrow 0 \quad \text{a.e.}$$

7.2 Convergence Theorems

We are now in a position to establish several basic results on convergence of sequences of random variables. We start with an example that motivated some of the early work in this subject, the problem of *random signs*. Let a_1, a_2, \dots be a fixed sequence of real numbers, and let an unbiased coin be tossed independently over and over again. If the n th toss results in heads, we write down the number $+a_n$, if tails, the number $-a_n$. In this way we generate a series such as $a_1 - a_2 - a_3 + a_4 + \dots$, where the signs are chosen at random. Will the series converge?

The general question suggested by the random signs problem involves the convergence of a series of independent random variables. The following result gives considerable information.

7.2.1 Theorem. Let X_1, X_2, \dots be independent random variables with finite expectation. If $\sum_{n=1}^{\infty} \text{Var } X_n < \infty$, then $\sum_{n=1}^{\infty} [X_n - E(X_n)]$ converges a.e. [All random variables are assumed to be defined on a fixed probability space (Ω, \mathcal{F}, P) , and “almost everywhere” refers to the probability measure P . Also, throughout this chapter, convergence will always mean to a *finite* limit.]

PROOF. We may assume that $E(X_n) \equiv 0$. Let $S_n = X_1 + \dots + X_n$. Then S_n converges iff $S_j - S_k \rightarrow 0$ as $j, k \rightarrow \infty$, and this happens a.e. iff for each $\varepsilon > 0$,

$$P\left[\bigcup_{j,k \geq n} \{|S_j - S_k| \geq \varepsilon\}\right] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

(see Section 2.5, Problem 4). Equivalently, we must prove that for each $\varepsilon > 0$,

$$P\left[\bigcup_{k=1}^{\infty} \{|S_{m+k} - S_m| \geq \varepsilon\}\right] \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

But

$$\begin{aligned}
 P\left[\bigcup_{k=1}^{\infty} \{|S_{m+k} - S_m| \geq \varepsilon\}\right] &= \lim_{n \rightarrow \infty} P\left[\bigcup_{k=1}^n \{|S_{m+k} - S_m| \geq \varepsilon\}\right] \\
 &= \lim_{n \rightarrow \infty} P\left\{\max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq \varepsilon\right\} \\
 &\leq \frac{1}{\varepsilon^2} \text{Var}(S_{m+n} - S_m) \quad \text{by 7.1.4} \\
 &= \frac{1}{\varepsilon^2} \sum_{j=1}^n \text{Var}(X_{m+j}) \quad \text{by 5.10.11} \\
 &\rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{since } \sum_{n=1}^{\infty} \text{Var } X_n < \infty. \blacksquare
 \end{aligned}$$

In the random signs problem we have $X_n = a_n Y_n$, where the Y_n are independent, taking values $+1$ and -1 with equal probability. It follows that if $\sum_{n=1}^{\infty} a_n^2 < \infty$, the series $\sum_{n=1}^{\infty} X_n$ converges a.e. After we prove Theorem 7.8.7, we shall see that the condition $\sum_{n=1}^{\infty} a_n^2 < \infty$ is necessary as well as sufficient for a.e. convergence of the series.

If X_1, X_2, \dots are independent random variables, we proved in Chapter 5 that under appropriate conditions, $(S_n - E(S_n))/n$ converges to 0 in probability (the weak law of large numbers). We now consider almost everywhere convergence.

7.2.2 Kolmogorov Strong Law of Large Numbers. Let X_1, X_2, \dots be independent random variables, each with finite mean and variance, and let $\{b_n\}$ be an increasing sequence of positive real numbers with $b_n \rightarrow \infty$. If

$$\sum_{n=1}^{\infty} \frac{\text{Var } X_n}{b_n^2} < \infty,$$

then (with $S_n = X_1 + \dots + X_n$)

$$\frac{S_n - E(S_n)}{b_n} \rightarrow 0 \quad \text{a.e.}$$

PROOF.

$$\sum_{n=1}^{\infty} \text{Var}\left(\frac{X_n - E(X_n)}{b_n}\right) = \sum_{n=1}^{\infty} \frac{\text{Var } X_n}{b_n^2} < \infty \quad \text{by hypothesis.}$$

By 7.2.1, $\sum_{n=1}^{\infty} (X_n - E(X_n))/b_n$ converges a.e. But

$$\frac{S_n - E(S_n)}{b_n} = \frac{1}{b_n} \sum_{k=1}^n b_k \left(\frac{X_k - E(X_k)}{b_k} \right),$$

and this approaches zero a.e. by 7.1.3. \blacksquare

In particular, if the X_n are independent random variables, each with finite mean m and finite variance σ^2 , then $S_n/n \rightarrow m$ a.e. (take $b_n = n$ in 7.2.2).

Another special case: If the X_n are independent and the fourth central moments are uniformly bounded, that is, for some finite M we have $E[(X_n - E(X_n))^4] \leq M$ for all n , then $(S_n - E(S_n))/n \rightarrow 0$ a.e. For by the Cauchy-Schwarz inequality,

$$\text{Var } X_n = E[(X_n - E(X_n))^2 \cdot 1] \leq (E[(X_n - E(X_n))^4])^{1/2} \leq M^{1/2},$$

and therefore 7.2.2 applies with $b_n = n$. This result, due to Cantelli, may in fact be proved without much machinery from measure theory; see Ash (1970, p. 206).

If the X_n are independent and all have the same distribution, in other words, for each Borel set $B \subset R$, $P\{X_n \in B\}$ is the same for all n , a version of the strong law of large numbers may be proved under a hypothesis on the mean of the X_n but no assumptions about higher moments. We first indicate some terminology that will be used in the remainder of the book.

7.2.3 Definition. If the random variables X_n all have the same distribution, they will be called *identically distributed*. The phrase “independent and identically distributed” will be abbreviated *iid*.

We need one preliminary result.

7.2.4 Lemma. If Y is a nonnegative random variable,

$$\sum_{n=1}^{\infty} P\{Y \geq n\} \leq E(Y) \leq 1 + \sum_{n=1}^{\infty} P\{Y \geq n\}.$$

PROOF.

$$\begin{aligned} \sum_{n=1}^{\infty} P\{Y \geq n\} &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{k \leq Y < k+1\} = \sum_{k=1}^{\infty} \sum_{n=1}^k P\{k \leq Y < k+1\} \\ &= \sum_{k=1}^{\infty} k P\{k \leq Y < k+1\} = \sum_{k=0}^{\infty} \int_{\{k \leq Y < k+1\}} k dP \\ &\leq \sum_{k=0}^{\infty} \int_{\{k \leq Y < k+1\}} Y dP = E(Y) \leq \sum_{k=0}^{\infty} (k+1) P\{k \leq Y < k+1\} \\ &= \sum_{n=1}^{\infty} P\{Y \geq n\} + \sum_{k=0}^{\infty} P\{k \leq Y < k+1\} = \sum_{n=1}^{\infty} P\{Y \geq n\} + 1. \blacksquare \end{aligned}$$

7.2.5 Strong Law of Large Numbers, iid Case. If X_1, X_2, \dots are iid random variables with finite expectation m , and $S_n = X_1 + \dots + X_n$, then $S_n/n \rightarrow m$ a.e.

PROOF. Since all X_n have the same distribution,

$$\sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \sum_{n=1}^{\infty} P\{|X_1| \geq n\};$$

thus

$$\sum_{n=1}^{\infty} P\{|X_n| \geq n\} \leq E(|X_1|) < \infty \quad \text{by 7.2.4.}$$

By the Borel–Cantelli lemma, $P\{|X_n| \geq n\} = 0$. Thus if we define $Y_n = X_n$ if $|X_n| < n$; $Y_n = 0$ if $|X_n| \geq n$, then except on a set of probability 0, $Y_n = X_n$ for sufficiently large n . Thus assuming (without loss of generality) that $m = 0$, it suffices to show that

$$\frac{1}{n} \sum_{j=1}^n Y_j \rightarrow 0 \quad \text{a.e.}$$

Now

$$\begin{aligned} E(Y_n) &= E(X_n I_{\{|X_n| < n\}}) \\ &= E(X_1 I_{\{|X_1| < n\}}) \\ &\quad \text{by the iid hypothesis} \\ &\rightarrow E(X_1) = 0 \quad \text{as } n \rightarrow \infty \\ &\quad \text{by the dominated convergence theorem.} \end{aligned}$$

Consequently,

$$\frac{1}{n} \sum_{j=1}^n E(Y_j) \rightarrow 0,$$

and therefore it is sufficient to show that

$$\frac{1}{n} \sum_{j=1}^n [Y_j - E(Y_j)] \rightarrow 0 \quad \text{a.e.}$$

If we can show that

$$\sum_{n=1}^{\infty} \left[\frac{Y_n - E(Y_n)}{n} \right]$$

converges a.e., then the Kronecker lemma 7.1.3 with $b_n = n$ and

$$x_n = \frac{Y_n - E(Y_n)}{n},$$

yields

$$\frac{1}{n} \sum_{j=1}^n [Y_j - E(Y_j)] \rightarrow 0 \quad \text{a.e.},$$

as desired. Now the Y_n are functions of the independent random variables X_n , and hence are independent, so by 7.2.1, it suffices to show that

$$V = \sum_{n=1}^{\infty} \text{Var} \left(\frac{Y_n}{n} \right) < \infty.$$

(Note that $|Y_n| < n$, so $\text{Var } Y_n$ is finite, although nothing is known about $\text{Var } X_n$.) But

$$\begin{aligned} V &= \sum_{n=1}^{\infty} \frac{1}{n^2} \text{Var } Y_n \\ &\leq \sum_{n=1}^{\infty} \frac{1}{n^2} E(Y_n^2) \quad \text{since } \text{Var } Y_n = E(Y_n^2) - [E(Y_n)]^2 \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} E(X_1^2 I_{\{|X_1| < n\}}) \quad \text{by the iid hypothesis} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{m=1}^n E(X_1^2 I_{\{m-1 \leq |X_1| < m\}}) \\ &= \sum_{m=1}^{\infty} E(X_1^2 I_{\{m-1 \leq |X_1| < m\}}) \sum_{n=m}^{\infty} \frac{1}{n^2}. \end{aligned}$$

By comparing $\sum(1/n^2)$ with $\int(1/x^2) dx$, we find that $\sum_{n=m}^{\infty}(1/n^2) \leq K/m$ for some fixed positive constant K . Thus

$$V \leq K \sum_{m=1}^{\infty} \frac{1}{m} E(X_1^2 I_{\{m-1 \leq |X_1| < m\}}).$$

If $m-1 \leq |X_1| < m$, then $X_1^2 = |X_1| |X_1| \leq m |X_1|$; hence

$$V \leq K \sum_{m=1}^{\infty} E(|X_1| I_{\{m-1 \leq |X_1| < m\}}) = KE(|X_1|) < \infty. \blacksquare$$

If $E(X_i)$ exists but is not necessarily finite in Theorem 7.2.5, the result still holds. To see this, first assume that the X_i are nonnegative, with infinite expectation. If $M > 0$ and $S_n' = \sum_{i=1}^n X_i I_{\{X_i \leq M\}}$, then, almost everywhere,

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n} \geq \liminf_{n \rightarrow \infty} \frac{S_n'}{n} = E(X_1 I_{\{X_1 \leq M\}}) \rightarrow E(X_1) = \infty \quad \text{as } M \rightarrow \infty.$$

Therefore $n^{-1} S_n \rightarrow \infty$ a.e. The general case is handled by splitting the random variables X_i into positive and negative parts.

We conclude this section with a remarkable result about independent random variables. If X_1, X_2, \dots are independent, a question which might

occur is: What is the probability that $\sum_{n=1}^{\infty} X_n$ converges? It might be expected that examples exist with any number between 0 and 1 as the probability of convergence, but in fact, the probability must be 0 or 1. Many other events defined in terms of independent random variables have this "zero-one" property, as we shall see.

7.2.6 Definitions and Comments. Let X_1, X_2, \dots be a sequence of random variables, and let $\mathcal{F}_n = \mathcal{F}(X_n, X_{n+1}, \dots)$, $n = 1, 2, \dots$; \mathcal{F}_n may be thought of as the σ -field of events involving X_n, X_{n+1}, \dots . The σ -field $\mathcal{F}_{\infty} = \bigcap_{n=1}^{\infty} \mathcal{F}_n$ is called the *tail σ -field* of the X_n , sets in \mathcal{F}_{∞} are called *tail events* and \mathcal{F}_{∞} -measurable functions, that is, functions $f: (\Omega, \mathcal{F}_{\infty}) \rightarrow (\bar{R}, \mathcal{B}(\bar{R}))$ are called *tail functions* (relative to the X_n). Intuitively, a tail event is one whose occurrence or nonoccurrence is not affected by changing the values of finitely many of the X_i , and a tail function is one whose value is not affected by such a change. Thus $\{\lim_{n \rightarrow \infty} X_n \text{ exists}\}$, $\{\sum_{n=1}^{\infty} X_n \text{ converges}\}$, and $\{X_n < 1 \text{ for infinitely many } n\}$ are tail events, and $\limsup_{n \rightarrow \infty} X_n$ and $\liminf_{n \rightarrow \infty} X_n$ are tail functions. [Example of a formal proof: $\{\sum_{n=1}^{\infty} X_n \text{ converges}\} = \{\sum_{k=n}^{\infty} X_k \text{ converges}\} \in \mathcal{F}_n$ for each n ; hence the event belongs to \mathcal{F}_{∞} . Similarly,

$$\left\{ \limsup_{n \rightarrow \infty} X_n < c \right\} = \left\{ \limsup_{\substack{k \rightarrow \infty \\ k \geq n}} X_k < c \right\} \in \mathcal{F}_n \quad \text{for all } n;$$

hence $\limsup_{n \rightarrow \infty} X_n$ is \mathcal{F}_{∞} -measurable.]

7.2.7 Kolmogorov Zero–One Law. All tail events relative to a sequence of independent random variables have probability 0 or 1, and all tail functions are constant almost everywhere.

PROOF. Fix $A \in \mathcal{F}_{\infty}$; the idea is to show that A is independent of itself, so that $P(A \cap A) = P(A)P(A)$, and consequently $P(A) = 0$ or 1. Since $\mathcal{F}_{\infty} \subset \mathcal{F}_1$, A is of the form $\{(X_1, X_2, \dots) \in A'\}$ for some $A' \in \mathcal{B}(\bar{R})^{\circ}$. Let \mathcal{C} be the class of sets $C' \in \mathcal{B}(\bar{R})^{\circ}$ such that A and C are independent, where $C = \{(X_1, X_2, \dots) \in C'\}$. If C' is a measurable cylinder, then C is of the form $\{(X_1, \dots, X_n) \in B_n\}$; since $A \in \mathcal{F}_{n+1}$, A can be written in the form $\{(X_{n+1}, X_{n+2}, \dots) \in A_{n+1}\}$, and it follows that A and C are independent. Thus \mathcal{C} contains all measurable cylinders. But if $C'_n \in \mathcal{C}$, $C'_n \uparrow C'$ (or $C'_n \downarrow C'$), and $P(A \cap C_n) = P(A)P(C_n)$, $n = 1, 2, \dots$, then $C_n \uparrow C$ (or $C_n \downarrow C$); hence $P(A \cap C) = P(A)P(C)$. Therefore \mathcal{C} is a monotone class containing the measurable cylinders; hence \mathcal{C} contains all sets in $\mathcal{B}(\bar{R})^{\circ}$, in particular, A' . But then A is independent of itself.

Finally, if f is a tail function, then for each $c \in \bar{R}$, $\{\omega : f(\omega) < c\}$ is a tail event, and hence has probability 0 or 1. If $k = \sup\{c \in \bar{R} : P\{f < c\} = 0\}$, then $f = k$ a.e. ■

If the X_n are identically distributed as well as independent, a wider class of events will be shown to have probability 0 or 1.

7.2.8 Definitions and Comments. Let X_1, X_2, \dots be a sequence of random variables, and define the σ -fields \mathcal{F}_n as in 7.2.6. Let $A \in \mathcal{F}_1$, so that $A = \{(X_1, X_2, \dots) \in A'\}$, $A' \in \mathcal{B}(R)^\infty$. The event A is said to be *symmetric* iff the occurrence or nonoccurrence of A is not affected by a permutation of finitely many of the X_i . Formally, if $T: \{1, 2, \dots\} \rightarrow \{1, 2, \dots\}$ is a permutation of finitely many coordinates, we require that $A = \{X_{T(1)}, X_{T(2)}, \dots \in A'\}$.

Any tail event A is symmetric, for if T permutes the first n coordinates, we may write A in the form $\{(X_{n+1}, X_{n+2}, \dots) \in A_{n+1}\}$ since $A \in \mathcal{F}_{n+1}$. Thus

$$A = \{(X_1, X_2, \dots) \in R^n \times A_{n+1}\} = \{(X_{T(1)}, X_{T(2)}, \dots) \in R^n \times A_{n+1}\}$$

since $T(k) = k$, $k > n$.

There are, however, symmetric events that are not tail events, for example, $\{X_n = 0 \text{ for all } n\}$ and $\{\lim_{n \rightarrow \infty} (X_1 + \dots + X_n) \text{ exists and is less than } c\}$.

If $B = \{(X_1, X_2, \dots) \in B'\} \in \mathcal{F}_1$, not necessarily symmetric, and T permutes finitely many coordinates, we denote by $X(T)$ the sequence $(X_{T(1)}, X_{T(2)}, \dots)$ and we denote by $B(T)$ the event $\{X(T) \in B'\}$. Thus B is symmetric iff $B = B(T)$ for all T .

7.2.9 Hewitt–Savage Zero–One Law. Let X_1, X_2, \dots be iid random variables. If A is a symmetric set in $\mathcal{F}(X_1, X_2, \dots)$, then $P(A) = 0$ or 1.

PROOF. Let $A = \{(X_1, X_2, \dots) \in A'\}$ so that $P(A) = P_X(A')$, $X = (X_1, X_2, \dots)$. Find measurable cylinders C_k' such that $P_X(A' \Delta C_k') \rightarrow 0$ as $k \rightarrow \infty$ (see 1.3.11), and let $C_k = \{X \in C_k'\}$; say $C_k = \{(X_1, \dots, X_{n_k}) \in B_k\}$. Let T_k interchange $(1, \dots, n_k)$ and $(n_k + 1, \dots, 2n_k)$.

Since the X_n are iid, X and $X(T_k)$ have the same distribution; therefore

$$\begin{aligned} P(A \Delta C_k) &= P_X(A' \Delta C_k') = P_{X(T_k)}(A' \Delta C_k') \\ &= P[\{X(T_k) \in A'\} \Delta \{X(T_k) \in C_k'\}] \\ &= P[\{X \in A'\} \Delta \{X(T_k) \in C_k'\}] \quad \text{since } A \text{ is symmetric} \\ &= P(A \Delta C_k(T_k)). \end{aligned}$$

Thus $P(A \Delta C_k)$ and $P(A \Delta C_k(T_k))$ approach 0, and therefore so does $P(A \Delta [C_k \cap C_k(T_k)])$. It follows that $P(C_k)$, $P(C_k(T_k))$ and $P[C_k \cap C_k(T_k)]$ all approach $P(A)$. But

$$\begin{aligned} P[C_k \cap C_k(T_k)] &= P[\{(X_1, \dots, X_{n_k}) \in B_k, (X_{n_k+1}, \dots, X_{2n_k}) \in B_k\}] \\ &= P(C_k)P(C_k(T_k)). \end{aligned}$$

Let $k \rightarrow \infty$ to obtain $P(A) = P(A)P(A)$. ■

Problems

1. Let X_1, X_2, \dots be iid random variables. If

$$\frac{1}{n} \sum_{k=1}^n X_k$$

converges a.e. to a finite limit, show that $E(X_1)$ is finite and the limit equals $E(X_1)$ a.e.

2. The following result generalizes Lemma 7.2.4, which was used in the proof of the strong law of large numbers.

Let F be an increasing, right-continuous function from the set of non-negative real numbers to itself, with $F(0) = 0$. If Y is an arbitrary non-negative random variable, show that

$$E[F(Y)] = \int_0^\infty P\{Y \geq \lambda\} dF(\lambda).$$

In particular, if $F(\lambda) = \lambda$ we obtain

$$E(Y) = \int_0^\infty P\{Y \geq \lambda\} d\lambda.$$

We obtain 7.2.4 by expressing the integral from 0 to ∞ as a sum of integrals from n to $n + 1$. [Let $h(\omega, \lambda) = 1$ if $Y(\omega) \geq \lambda$, and 0 if $Y(\omega) < \lambda$, and use Fubini's theorem. Note also that $\{Y \geq \lambda\}$ can be replaced by $\{Y > \lambda\}$ in the above formulas; the same proof applies.]

3. Give an example to show that the Hewitt–Savage zero–one law may fail when the X_n are independent but not identically distributed.
4. Let X_1, X_2, \dots be iid random variables, and let $g : (R^\infty, \mathcal{B}^\infty) \rightarrow (R^\infty, \mathcal{B}^\infty)$, where $\mathcal{B} = \mathcal{B}(R)$. If g is symmetric, in other words, $g(x_{T(1)}, x_{T(2)}, \dots) = g(x_1, x_2, \dots)$ whenever T is a permutation of finitely many coordinates, show that $g(X_1, X_2, \dots)$ is constant a.e.
5. Let X_1, X_2, \dots be independent random variables, with $P\{X_n = 1\} = p_n$, $P\{X_n = 0\} = 1 - p_n$. Show that

$$X_n \xrightarrow{P} 0 \quad \text{iff} \quad \lim_{n \rightarrow \infty} p_n = 0$$

and

$$X_n \xrightarrow{\text{a.e.}} 0 \quad \text{iff} \quad \sum_{n=1}^{\infty} p_n < \infty.$$

6. Let X_1, X_2, \dots be nonnegative random variables, with X_n having density $\lambda_n \exp(-\lambda_n x)$, $x \geq 0$ ($\lambda_n > 0$).
- (a) If $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$, show that $\sum_{n=1}^{\infty} X_n < \infty$ a.e.

- (b) If the X_n are independent and $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$, show that

$$\sum_{n=1}^{\infty} X_n = \infty \quad \text{a.e.}$$

[In (b), consider $\exp(-\sum_{j=1}^n X_j)$.]

7. Let X_1, X_2, \dots be iid random variables, with $P\{X_n = i\} = 1/r$, $i = 0, 1, \dots, r-1$, and define $X = \sum_{n=1}^{\infty} r^{-n} X_n$. Thus X is the number in $[0, 1]$ with r -adic expansion $.X_1 X_2 \dots$.
- (a) Show that X is uniformly distributed, in other words, P_X is Lebesgue measure.
 - (b) Show that for almost every $x \in [0, 1]$ (Lebesgue measure), the following condition holds:

For every $r = 2, 3, \dots$ and all $i = 0, 1, \dots, r-1$, the relative frequency of i in the first n digits of the r -adic expansion of x converges to $1/r$ as $n \rightarrow \infty$.

- (c) If $x \in [0, 1]$, let $R_n(x) = 2x_n - 1$, where x_n is the n th digit of the binary expansion of x (to avoid ambiguity, eliminate expansions with only a finite number of zeros). The R_n are called the *Rademacher functions*. Use part (a) to show that $\int_0^1 R_n(x) dx = 0$, and

$$\int_0^1 R_n(x) R_m(x) dx = \begin{cases} 0, & n \neq m \\ 1, & n = m. \end{cases}$$

7.3 Martingales

Probability theory has its roots in games of chance, and it is often profitable to interpret results in terms of a gambling situation. For example, if X_1, X_2, \dots is a sequence of random variables, we may think of X_n as our total winnings after n trials in a succession of games. Having survived the first n trials, our expected fortune after trial $n+1$ is $E(X_{n+1} | X_1, \dots, X_n)$. If this equals X_n , the game is “fair” since the expected gain on trial $n+1$ is $E(X_{n+1} - X_n | X_1, \dots, X_n) = X_n - X_n = 0$. If $E(X_{n+1} | X_1, \dots, X_n) \geq X_n$, the game is “favorable,” and if $E(X_{n+1} | X_1, \dots, X_n) \leq X_n$, the game is “unfavorable.”

We are going to study sequences of this type; the results to be obtained will have significance outside the casino as well as inside.

7.3.1 Definition. Let (Ω, \mathcal{F}, P) be a probability space, $\{X_1, X_2, \dots\}$ a sequence of integrable random variables on (Ω, \mathcal{F}, P) , and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ an increasing sequence of sub σ -fields of \mathcal{F} ; X_n is assumed \mathcal{F}_n -measurable, that is, $X_n : (\Omega, \mathcal{F}_n) \rightarrow (R, \mathcal{B}(R))$. The sequence $\{X_n\}$ is said to be a *martingale*

relative to the \mathcal{F}_n (alternatively, we say that $\{X_n, \mathcal{F}_n\}$ is a martingale) iff for all $n = 1, 2, \dots$, $E(X_{n+1}|\mathcal{F}_n) = X_n$ a.e., a *submartingale* iff $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ a.e., a *supermartingale* iff $E(X_{n+1}|\mathcal{F}_n) \leq X_n$ a.e. (In statements involving conditional expectations, the “a.e.” is always understood and will usually be omitted.)

7.3.2 Comments.

(a) If $\{X_n, \mathcal{F}_n\}$ is a martingale, then

$$E(X_{n+k}|\mathcal{F}_n) = X_n, \quad n, k = 1, 2, \dots$$

(with corresponding statements for sub- and supermartingales). For

$$\begin{aligned} E(X_{n+2}|\mathcal{F}_n) &= E[E(X_{n+2}|\mathcal{F}_{n+1})|\mathcal{F}_n] \quad \text{by 6.5.10(a)} \\ &= E(X_{n+1}|\mathcal{F}_n) \\ &= X_n. \end{aligned}$$

The general statement follows by induction.

(b) If $\{X_n, \mathcal{F}_n\}$ is a martingale, then

$$E(X_{n+1}|X_1, \dots, X_n) = X_n, \quad n = 1, 2, \dots$$

Thus $\{X_n\}$ is automatically a martingale relative to the standard σ -fields $\mathcal{F}(X_1, \dots, X_n)$ (with corresponding statements for sub- and supermartingales).

For $\mathcal{F}_1 \subset \dots \subset \mathcal{F}_n$, and thus X_1, \dots, X_n are all \mathcal{F}_n -measurable. Since $\mathcal{F}(X_1, \dots, X_n)$ is the smallest σ -field making X_1, \dots, X_n measurable [see 6.4.2(b)], we have $\mathcal{F}(X_1, \dots, X_n) \subset \mathcal{F}_n$. If, in the defining relation

$$E(X_{n+1}|\mathcal{F}_n) = X_n,$$

we take conditional expectations with respect to $\mathcal{F}(X_1, \dots, X_n)$, we obtain the desired result by 6.5.10(a) and 6.5.11(a).

If we say that $\{X_n\}$ is a martingale (or sub-, supermartingale) without mentioning the σ -fields \mathcal{F}_n , we shall always mean $\mathcal{F}_n = \mathcal{F}(X_1, \dots, X_n)$, so that $E(X_{n+1}|X_1, \dots, X_n) = X_n$.

(c) $\{X_n, \mathcal{F}_n\}$ is a martingale iff

$$\int_A X_n dP = \int_A X_{n+1} dP \quad \text{for all } A \in \mathcal{F}_n, \quad n = 1, 2, \dots$$

This follows since the condition $E(X_{n+1}|\mathcal{F}_n) = X_n$ a.e. [P] is equivalent to

$$\int_A E(X_{n+1}|\mathcal{F}_n) dP = \int_A X_n dP \quad \text{for all } A \in \mathcal{F}_n$$

(see 1.6.11); also

$$\int_A E(X_{n+1}|\mathcal{F}_n) dP = \int_A X_{n+1} dP$$

by definition of conditional expectation.

Similarly, $\{X_n, \mathcal{F}_n\}$ is a submartingale iff

$$\int_A X_n dP \leq \int_A X_{n+1} dP \quad \text{for all } A \in \mathcal{F}_n,$$

and a supermartingale iff

$$\int_A X_n dP \geq \int_A X_{n+1} dP \quad \text{for all } A \in \mathcal{F}_n.$$

In particular, $E(X_n)$ is constant in a martingale, increases in a submartingale, and decreases in a supermartingale.

(d) The defining condition for a martingale relative to the σ -fields $\mathcal{F}(X_1, \dots, X_n)$ is equivalent to

$$E(X_{n+1} | X_1 = x_1, \dots, X_n = x_n) = x_n \quad \text{a.e. } [P_{(X_1, \dots, X_n)}]$$

with similar statements for sub- and supermartingales.

For if $A \in \mathcal{F}(X_1, \dots, X_n)$, then A is of the form $\{(X_1, \dots, X_n) \in B\}$, $B \in \mathcal{B}(R^n)$ (see 6.4.1). If $X = (X_1, \dots, X_n)$, then

$$\begin{aligned} \int_A X_{n+1} dP &= \int_{\{X \in B\}} X_{n+1} dP \\ &= \int_B E(X_{n+1} | X_1 = x_1, \dots, X_n = x_n) dP_X \quad \text{by 6.3.3} \end{aligned}$$

and

$$\begin{aligned} \int_A X_n dP &= \int_B E(X_n | X_1 = x_1, \dots, X_n = x_n) dP_X \\ &= \int_B x_n dP_X \quad \text{by 6.5.11(a').} \end{aligned}$$

The result now follows from (c).

(e) A finite sequence $\{X_k, \mathcal{F}_k, k = 1, \dots, n\}$ is called a martingale iff $E(X_{k+1} | \mathcal{F}_k) = X_k, k = 1, 2, \dots, n - 1$; finite sub- and supermartingale sequences are defined similarly.

(f) If $\{X_n, \mathcal{F}_n\}$ and $\{Y_n, \mathcal{F}_n\}$ are submartingales, so is $\{\max(X_n, Y_n), \mathcal{F}_n\}$.

For $E(\max(X_{n+1}, Y_{n+1}) | \mathcal{F}_n) \geq E(X_{n+1} | \mathcal{F}_n) \geq X_n$, and similarly

$$E(\max(X_{n+1}, Y_{n+1}) | \mathcal{F}_n) \geq Y_n.$$

The same approach shows that if $\{X_n, \mathcal{F}_n\}$ and $\{Y_n, \mathcal{F}_n\}$ are supermartingales, so is $\{\min(X_n, Y_n), \mathcal{F}_n\}$.

7.3.3 Examples. If $X_n \equiv X$, then $\{X_n\}$ is a martingale; if $X_1 \leq X_2 \leq \dots$, then $\{X_n\}$ is a submartingale; if $X_1 \geq X_2 \geq \dots$, then $\{X_n\}$ is a supermartingale (assuming all random variables integrable).

We give some more substantial examples.

(a) Let Y_1, Y_2, \dots be independent random variables with zero mean, and set $X_n = \sum_{k=1}^n Y_k$, $\mathcal{F}_n = \mathcal{F}(Y_1, \dots, Y_n)$. Then $\{X_n, \mathcal{F}_n\}$ is a martingale. For

$$\begin{aligned} E(X_{n+1} | \mathcal{F}_n) &= E(X_n + Y_{n+1} | Y_1, \dots, Y_n) \\ &= X_n + E(Y_{n+1} | Y_1, \dots, Y_n) \\ &\quad \text{since } X_n \text{ is } \mathcal{F}_n\text{-measurable} \\ &= X_n + E(Y_{n+1}) \\ &\quad \text{by independence (Problem 1, Section 6.5)} \\ &= X_n \\ &\quad \text{since } E(Y_j) \equiv 0. \end{aligned}$$

(b) Let Y_1, Y_2, \dots be independent random variables with $E(Y_j) = a_j \neq 0$, and set $X_n = \prod_{j=1}^n (Y_j/a_j)$, $\mathcal{F}_n = \mathcal{F}(Y_1, \dots, Y_n)$. Then $\{X_n, \mathcal{F}_n\}$ is a martingale. For

$$\begin{aligned} E(X_{n+1} | \mathcal{F}_n) &= E\left[\left(\frac{X_n Y_{n+1}}{a_{n+1}}\right) \middle| Y_1, \dots, Y_n\right] \\ &= X_n E\left(\frac{Y_{n+1}}{a_{n+1}}\right) \\ &\quad \text{by 6.5.11(a) and Problem 1, Section 6.5} \\ &= X_n. \end{aligned}$$

(c) Let Y be an integrable random variable on (Ω, \mathcal{F}, P) .

If $\{\mathcal{F}_n\}$ is an increasing sequence of sub σ -fields of \mathcal{F} , and $X_n = E(Y | \mathcal{F}_n)$, then $\{X_n, \mathcal{F}_n\}$ is a martingale. For

$$\begin{aligned} E(X_{n+1} | \mathcal{F}_n) &= E[E(Y | \mathcal{F}_{n+1}) | \mathcal{F}_n] \\ &= E(Y | \mathcal{F}_n) \quad \text{since } \mathcal{F}_n \subset \mathcal{F}_{n+1} \\ &= X_n. \end{aligned}$$

If $\{\mathcal{F}_n\}$ is a decreasing sequence of sub σ -fields of \mathcal{F} , and $X_n = E(Y | \mathcal{F}_n)$, then $\{X_n, \mathcal{F}_n\}$ is a *reverse martingale*, that is, $E(X_n | \mathcal{F}_{n+1}) = X_{n+1}$. [Similarly, if $\{\mathcal{F}_n\}$ is a decreasing sequence of sub σ -fields of \mathcal{F} , $E(X_n | \mathcal{F}_{n+1}) \geq X_{n+1}$ defines a reverse submartingale, $E(X_n | \mathcal{F}_{n+1}) \leq X_{n+1}$ defines a reverse supermartingale; in all cases it is required that X_n be \mathcal{F}_n -measurable for all n .] For

$$\begin{aligned} E(X_n | \mathcal{F}_{n+1}) &= E[E(Y | \mathcal{F}_n) | \mathcal{F}_{n+1}] \\ &= E(Y | \mathcal{F}_{n+1}) \quad \text{since } \mathcal{F}_{n+1} \subset \mathcal{F}_n \\ &= X_{n+1}. \end{aligned}$$

Note that as in 7.3.2(a), $E(X_n | \mathcal{F}_{n+k}) = X_{n+k}$, $n, k = 1, 2, \dots$

(d) (Branching Processes) We define a Markov chain (see Section 5.11) with state space $S = \{0, 1, 2, \dots\}$. The state at time n , denoted by X_n , is to represent the number of “offspring” after n “generations.” We take $X_0 = 1$, and, if $X_n = k$, X_{n+1} is the sum of k independent, identically distributed, non-negative integer valued random variables, say Y_1, \dots, Y_k , where $P\{Y_i = r\} = p_r$, $r = 0, 1, 2, \dots$. Thus p_r is the probability that a given “being” will produce exactly r “offspring.” (Formally, we take $p_{kj} = P\{Y_1 + \dots + Y_k = j\}$, $k = 1, 2, \dots, j = 0, 1, \dots; p_{00} = 1$.)

Let $m = E(Y_i) = \sum_{r=0}^{\infty} rp_r$. If m is finite and greater than 0, then $\{X_n/m^n\}$ is a martingale [relative to

$$\mathcal{F}_n = \mathcal{F}(X_0, \dots, X_n) = \mathcal{F}\left(\frac{X_1}{m}, \frac{X_2}{m^2}, \dots, \frac{X_n}{m^n}\right).$$

For

$$\begin{aligned} E\left(\frac{X_{n+1}}{m^{n+1}} \middle| X_0 = i_0, \dots, X_n = i_n\right) &= \sum_{j=0}^{\infty} p_{i_n j} \frac{j}{m^{n+1}} \\ &\quad (\text{see 6.3.5(a), Eq. (2) and the definition of a Markov chain}) \\ &= \frac{1}{m^{n+1}} \sum_{j=0}^{\infty} j P\{Y_1 + \dots + Y_{i_n} = j\} \\ &= \frac{1}{m^{n+1}} E(Y_1 + \dots + Y_{i_n}) \\ &= \frac{i_n m}{m^{n+1}} = \frac{i_n}{m^n}. \end{aligned}$$

The result now follows from 7.3.2(d).

(e) Consider the branching process of part (d). Let $g(s) = \sum_j p_j s^j$, $s \geq 0$. If $g(r) = r$, then $\{r^{X_n}\}$ is a martingale relative to the σ -fields $\mathcal{F}(X_0, \dots, X_n)$. For as in (d),

$$\begin{aligned} E(r^{X_{n+1}} | X_0 = i_0, \dots, X_n = i_n) &= \sum_{j=0}^{\infty} p_{i_n j} r^j \\ &= \sum_{j=0}^{\infty} r^j P\{Y_1 + \dots + Y_{i_n} = j\} \\ &= E[\exp_r(Y_1 + \dots + Y_{i_n})] \\ &= [E(r^{Y_1})]^{i_n} = [g(r)]^{i_n} = r^{i_n}. \end{aligned}$$

The examples we have given thus far are all of martingales. We now show how to construct a large class of submartingales. (Note also that if $\{X_n, \mathcal{F}_n\}$ is a submartingale, then $\{-X_n, \mathcal{F}_n\}$ is a supermartingale.) First we need some basic convexity theorems.

7.3.4 Line of Support Theorem. Let $g : I \rightarrow R$, where I is an open interval of reals, bounded or unbounded. Assume g is convex, that is,

$$g(ax + (1 - a)y) \leq ag(x) + (1 - a)g(y)$$

for all $x, y \in I$ and all $a \in [0, 1]$. Then there are sequences $\{a_n\}$ and $\{b_n\}$ of real numbers such that for all $y \in I$ we have $g(y) = \sup_n(a_n y + b_n)$.

PROOF. In the course of the proof, we develop many of the basic properties of convex functions of one variable.

Let g be a convex function from I to R , I an open interval of reals. If $0 < h_1 < h_2$, then by convexity,

$$g(x \pm h_1) \leq \frac{(h_2 - h_1)}{h_2} g(x) + \frac{h_1}{h_2} g(x \pm h_2);$$

hence

$$\frac{1}{h_1} [g(x + h_1) - g(x)] \leq \frac{1}{h_2} [g(x + h_2) - g(x)] \quad (1)$$

and

$$-\frac{1}{h_2} [g(x - h_2) - g(x)] \leq -\frac{1}{h_1} [g(x - h_1) - g(x)]. \quad (2)$$

Also, if $h, h' > 0$,

$$g(x) \leq \frac{h'}{h + h'} g(x - h) + \frac{h}{h + h'} g(x + h'),$$

so

$$\frac{g(x - h) - g(x)}{-h} \leq \frac{g(x + h') - g(x)}{h'}. \quad (3)$$

By (1) and (2), the right and left derivatives g_+' and g_-' exist on I ; by (3) they are finite. [Note that if $x \in I$, we have $x - h, x + h' \in I$ for small $h, h' > 0$ since I is open; thus, in fact, the difference quotients $[g(x + h') - g(x)]/h'$, which decrease as $h' \downarrow 0$ by (1), are bounded below by a finite constant.] Furthermore, by (1) and (2),

$$g_+'(x) = \inf_{y > x} \frac{g(y) - g(x)}{y - x}, \quad g_-'(x) = \sup_{y < x} \frac{g(x) - g(y)}{x - y} \quad (4)$$

and by (3),

$$g_-'(x) \leq g_+'(x). \quad (5)$$

If $x_1 < x_2$, then

$$g_+'(x_1) \leq \frac{g(x_2) - g(x_1)}{x_2 - x_1} = \frac{g(x_1) - g(x_2)}{x_1 - x_2} \leq g_-'(x_2); \quad (6)$$

hence g_+' and g_-' are *increasing* on I . The existence of right and left derivatives implies that g has right and left limits at each point; thus g is *continuous*; for if not, g_+' or g_-' would be infinite at some point. If $y \geq x$, then $g(y) \geq g(x) + (y - x)g_+'(x)$, and if $y < x$, then $g(y) \geq g(x) + (y - x)g_-'(x)$ [by (4)]. We conclude that if $g_-'(x) \leq a_x \leq g_+'(x)$, then

$$g(y) \geq g(x) + (y - x)a_x \quad \text{for all } y \in I. \quad (7)$$

The function L_x given by $L_x(y) = g(x) + (y - x)a_x$, $y \in I$, is called a *line of support* for g at the point x . It is immediate that $g(x) = \sup\{L_s(x) : s \in I\}$, but we are trying to prove that g is the sup of *countably* many lines of support. If $x \in I$, let r approach x through a sequence of rational numbers. Then

$$L_r(x) = g(x) + (x - r)a_r.$$

But $g_-'(r) \leq a_r \leq g_+'(r)$, and the g' , being increasing functions, are bounded on any finite closed subinterval of I , so that the a_r form a bounded sequence. Consequently, $L_r(x) \rightarrow g(x)$. By (7), $L_r(y) \leq g(y)$ for all $y \in I$ and all r ; hence $g(x) = \sup\{L_s(x) : s \in I, s \text{ rational}\}$, that is, $g(x) = \sup\{g(s) + (x - s)a_s : s \in I, s \text{ rational}\}$. The proof is complete. ■

If I is not open, the theorem is false: consider $g(x) = 0$, $0 \leq x < 1$, $g(1) = 1$.

Now if X is a random variable with finite mean, we have seen that $E(X^2) \geq [E(X)]^2$ (5.10.6). This is a special case of the following general convexity theorem.

7.3.5 Jensen's Inequality. Let g be a convex function from I to R , where I is an open interval of reals, bounded or unbounded. Let X be a random variable on (Ω, \mathcal{F}, P) , with $X(\omega) \in I$ for all ω . Assume $E(X)$ to be finite. If \mathcal{H} is a sub σ -field of \mathcal{F} , then $E[g(X) | \mathcal{H}] \geq g[E(X | \mathcal{H})]$ a.e. In particular, $E[g(X)] \geq g[E(X)]$.

PROOF. First note that $E(X | \mathcal{H}) \in I$ a.e. For if, say, a is real and $X > a$, then $E(X | \mathcal{H}) > a$ a.e. because

$$0 \geq \int_{\{E(X | \mathcal{H}) \leq a\}} E(X - a | \mathcal{H}) dP = \int_{\{E(X | \mathcal{H}) \leq a\}} (X - a) dP \geq 0;$$

hence $X = a$ a.e. on $\{E(X | \mathcal{H}) \leq a\}$, which implies that $P\{E(X | \mathcal{H}) \leq a\} = 0$. Thus $g[E(X | \mathcal{H})]$ is well-defined.

By 7.3.4 we may write $g(y) = \sup_n(a_n y + b_n)$, $y \in I$, so $g(X) \geq a_n X + b_n$ for all n . Therefore $E[g(X)|\mathcal{H}] \geq a_n E(X|\mathcal{H}) + b_n$ a.e. Take the sup over n to finish the proof. ■

The proof of 7.3.5 shows that the hypothesis that $E(X)$ is finite may be dropped if it is known that $E(X)$ and $E[g(X)]$ exist, and $E(X|\mathcal{H}) \in I$ a.e.

We now show how to construct submartingales from martingales (or from other submartingales).

7.3.6 Theorem. (a) Let $\{X_n, \mathcal{F}_n\}$ be a submartingale, g a convex, increasing function from R to R . If $g(X_n)$ is integrable for all n , then $\{g(X_n), \mathcal{F}_n\}$ is a submartingale. Thus, for example, if $\{X_n\}$ is a submartingale, so is $\{X_n^+\}$.

(b) Let $\{X_n, \mathcal{F}_n\}$ be a martingale, g a convex function from R to R . If $g(X_n)$ is integrable for all n , then $\{g(X_n), \mathcal{F}_n\}$ is a submartingale. Thus if $r \geq 1$, $\{X_n\}$ is a martingale and $|X_n|^r$ is integrable for all n , then $\{|X_n|^r\}$ is a submartingale.

PROOF. We have $E[g(X_{n+1})|\mathcal{F}_n] \geq g[E(X_{n+1}|\mathcal{F}_n)]$ by Jensen's inequality. In (a), $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ by the submartingale property; hence $g[E(X_{n+1}|\mathcal{F}_n)] \geq g(X_n)$ since g is increasing. In (b), $E(X_{n+1}|\mathcal{F}_n) = X_n$ by the martingale property, so $g[E(X_{n+1}|\mathcal{F}_n)] = g(X_n)$. The result follows. ■

Problems

- Let $X_n = \sum_{k=1}^n Y_k$, where the Y_k are independent, with $P\{Y_k = 1\} = p$, $P\{Y_k = -1\} = q$ ($p, q > 0$, $p + q = 1$). Show that $\{(q/p)^{X_n}\}$ is a martingale relative to the σ -fields $\mathcal{F}(X_1, \dots, X_n)$ [$= \mathcal{F}(Y_1, \dots, Y_n)$].
- Consider a Markov chain whose state space is the integers, and assume that p_{ij} depends only on the difference between j and i : $p_{ij} = q_{j-i}$, where $q_k \geq 0$, $\sum_k q_k = 1$.
 - Show that if X_n is the state at time n , X_n may be written as $X_0 + Y_1 + \dots + Y_n$, where X_0, Y_1, \dots, Y_n are independent and the Y_i all have the same distribution, namely, $P\{Y_i = k\} = q_k$, k an integer.
 - If $\sum_j q_j r^j = 1$ (where the series is assumed to converge absolutely), show that $\{r^{X_n}\}$ is a martingale relative to the σ -fields $\mathcal{F}(X_0, \dots, X_n)$.
 - Assume $q_0 < 1$. If $r > 1$ in (b), show that $X_n \rightarrow -\infty$ a.e.; hence the states are transient.
- Let λ be a countably additive set function on the σ -field \mathcal{F} , and let \mathcal{F}_n be generated by the sets A_{n1}, A_{n2}, \dots , assumed to form a partition of Ω ,

with $P(A_{nj}) > 0$ for all j . Assume that the $(n+1)$ st partition refines the n th, so that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

Define

$$X_n(\omega) = \frac{\lambda(A_{nj})}{P(A_{nj})} \quad \text{if } \omega \in A_{nj}, \quad n, j = 1, 2, \dots$$

Show that $\{X_n, \mathcal{F}_n\}$ is a martingale.

4. Define a sequence of random variables as follows. Let X_1 be uniformly distributed between 0 and 1. Given that $X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}$, let X_n be uniformly distributed between 0 and x_{n-1} . Show that $\{X_n\}$ is a supermartingale and $E(X_n) = 2^{-n}$. Conclude that $X_n \rightarrow 0$ a.e.
5. Let X_1, X_2, \dots be real-valued Borel measurable functions on (Ω, \mathcal{F}) . Assume that under the probability measure P on \mathcal{F} , (X_1, \dots, X_n) has density p_n , and under the probability measure Q on \mathcal{F} , (X_1, \dots, X_n) has density q_n . Define

$$Y_n = \begin{cases} \frac{q_n(X_1, \dots, X_n)}{p_n(X_1, \dots, X_n)} & \text{if the denominator is greater than 0,} \\ 0 & \text{otherwise.} \end{cases}$$

Show that if $\mathcal{F}_n = \mathcal{F}(X_1, \dots, X_n)$, $\{Y_n, \mathcal{F}_n\}$ is a supermartingale on (Ω, \mathcal{F}, P) and $0 \leq E(Y_n) \leq 1$ for all n .

6. Let $\{X_n, \mathcal{F}_n, n \geq 0\}$ be a supermartingale, and define

$$\begin{aligned} Y_0 &= X_0, \\ Y_1 &= Y_0 + (X_1 - E(X_1 | \mathcal{F}_0)), \\ &\vdots \\ Y_n &= Y_{n-1} + (X_n - E(X_n | \mathcal{F}_{n-1})). \end{aligned}$$

Also define

$$\begin{aligned} A_0 &= 0, \\ A_1 &= X_0 - E(X_1 | \mathcal{F}_0), \\ A_2 &= A_1 + (X_1 - E(X_2 | \mathcal{F}_1)), \\ &\vdots \\ A_n &= A_{n-1} + (X_{n-1} - E(X_n | \mathcal{F}_{n-1})). \end{aligned}$$

- (a) Show that $X_n = Y_n - A_n$.
- (b) Show that $\{Y_n, \mathcal{F}_n\}$ is a martingale.
- (c) Show that for a.e. ω , $A_n(\omega)$ increases with n .

Thus a supermartingale can be expressed as the difference between a martingale and an increasing sequence. (Similarly, a submartingale can be expressed as the sum of a martingale and an increasing sequence.)

7.4 Martingale Convergence Theorems

Under rather mild conditions, sub- and supermartingales converge almost everywhere. This result has very many ramifications in probability theory.

We first prove a theorem which has an interesting gambling interpretation.

7.4.1 Optimal Skipping Theorem (Halmos). Let $\{X_n, \mathcal{F}_n\}$ be a submartingale. Let $\varepsilon_1, \varepsilon_2, \dots$ be random variables defined by

$$\varepsilon_k = \begin{cases} 1 & \text{if } (X_1, \dots, X_k) \in B_k, \\ 0 & \text{if } (X_1, \dots, X_k) \notin B_k, \end{cases}$$

where the B_k are arbitrary sets in $\mathcal{B}(R^n)$. Set

$$\begin{aligned} Y_1 &= X_1, \\ Y_2 &= X_1 + \varepsilon_1(X_2 - X_1) \\ &\vdots \\ Y_n &= X_1 + \varepsilon_1(X_2 - X_1) + \dots + \varepsilon_{n-1}(X_n - X_{n-1}). \end{aligned}$$

Then $\{Y_n, \mathcal{F}_n\}$ is also a submartingale and $E(Y_n) \leq E(X_n)$ for all n . If $\{X_n, \mathcal{F}_n\}$ is a martingale, so is $\{Y_n, \mathcal{F}_n\}$ and $E(Y_n) = E(X_n)$ for all n .

Interpretation. Let X_n be the gambler's fortune after n trials; then Y_n is our fortune if we follow an optional skipping strategy. After observing X_1, \dots, X_k , we may choose to bet with the gambler at trial $k+1$ [in this case $\varepsilon_k = \varepsilon_k(X_1, \dots, X_k) = 1$] or we may pass ($\varepsilon_k = 0$). Our gain on trial $k+1$ is $\varepsilon_k(X_{k+1} - X_k)$. The theorem states that whatever strategy we employ, if the game is initially "fair" (a martingale) or "favorable" (a submartingale), it remains fair (or favorable), and no strategy of this type can increase the expected winning.

PROOF.

$$\begin{aligned} E(Y_{n+1} | \mathcal{F}_n) &= E(Y_n + \varepsilon_n(X_{n+1} - X_n) | \mathcal{F}_n) \\ &= Y_n + \varepsilon_n E[(X_{n+1} - X_n) | \mathcal{F}_n] \end{aligned}$$

since ε_n is a Borel measurable function of X_1, \dots, X_n , and hence is

$$\mathcal{F}(X_1, \dots, X_n) \subset \mathcal{F}_n\text{-measurable.}$$

Therefore

$$\begin{aligned} E(Y_{n+1} | \mathcal{F}_n) &= Y_n + \varepsilon_n(X_n - X_n) = Y_n && \text{in the martingale case} \\ &\geq Y_n + \varepsilon_n(X_n - X_n) = Y_n && \text{in the submartingale case.} \end{aligned}$$

Since $Y_1 = X_1$, we have $E(X_1) = E(Y_1)$. Having shown $E(X_k - Y_k) \geq 0$ ($=0$ in the martingale case),

$$X_{k+1} - Y_{k+1} = X_{k+1} - Y_k - \varepsilon_k(X_{k+1} - X_k) = (1 - \varepsilon_k)(X_{k+1} - X_k) + X_k - Y_k.$$

Thus

$$\begin{aligned} E(X_{k+1} - Y_{k+1} | \mathcal{F}_k) &= (1 - \varepsilon_k)E(X_{k+1} - X_k | \mathcal{F}_k) + E(X_k - Y_k | \mathcal{F}_k) \\ &\geq E(X_k - Y_k | \mathcal{F}_k) = X_k - Y_k, \end{aligned}$$

with equality in the martingale case. Take expectations and use $E[E(X | \mathcal{G})] = E(X)$ to obtain

$$E(X_{k+1} - Y_{k+1}) \geq E(X_k - Y_k) \geq 0,$$

with equality in the martingale case. ■

The key step in the development is the following result, due in its original form to Doob.

7.4.2 Upcrossing Theorem. Let $\{X_k, \mathcal{F}_k, k = 1, 2, \dots, n\}$ be a submartingale. If a and b are real numbers, with $a < b$, let U_{ab} be the number of upcrossings of (a, b) by X_1, \dots, X_n , defined as follows.

Let $T_1 = T_1(\omega)$ be the first integer in $\{1, 2, \dots, n\}$ such that $X_{T_1} \leq a$, T_2 be the first integer greater than T_1 such that $X_{T_2} \geq b$, T_3 be the first integer greater than T_2 such that $X_{T_3} \leq a$, T_4 be the first integer greater than T_3 such that $X_{T_4} \geq b$, and so on. (Set $T_i = \infty$ if the condition cannot be satisfied.) If N is the number of finite T_i , define $U_{ab} = N/2$ if N is even, and $(N - 1)/2$ if N is odd. Then

$$E(U_{ab}) \leq \frac{1}{b - a} E[(X_n - a)^+].$$

PROOF. First assume $a = 0$, and all $X_j \geq 0$. Define the T_i as above ($X_{T_i} \leq a$ is now equivalent to $X_{T_i} = 0$). Let $\varepsilon_j = 0$ for $j < T_1$; $\varepsilon_j = 1$ for $T_1 \leq j < T_2$; $\varepsilon_j = 0$ for $T_2 \leq j < T_3$; $\varepsilon_j = 1$ for $T_3 \leq j < T_4$, and so on (see Fig. 7.1).

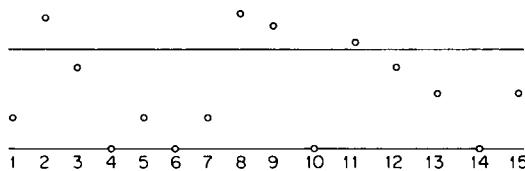


Figure 7.1.

In Figure 7.1, we have (with $n = 15$) $T_1 = 4$, $T_2 = 8$, $T_3 = 10$, $T_4 = 11$, $T_5 = 14$, $T_n = \infty$, $n > 5$, $U_{ab} = 2$;

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 0, \quad \varepsilon_4 = \varepsilon_5 = \varepsilon_6 = \varepsilon_7 = 1, \quad \varepsilon_8 = \varepsilon_9 = 0,$$

$$\varepsilon_{10} = 1, \quad \varepsilon_{11} = \varepsilon_{12} = \varepsilon_{13} = 0, \quad \varepsilon_{14} = 1;$$

$$X_1 + \varepsilon_1(X_2 - X_1) + \cdots + \varepsilon_{14}(X_{15} - X_{14}) \\ = X_1 + X_8 - X_4 + X_{11} - X_{10} + X_{15} - X_{14}.$$

Note that Y_n , as defined in 7.4.1, is the total increase during upcrossings, plus possibly a “partial upcrossing” at the end, plus a contribution due to X_1 (necessarily nonnegative). Thus $Y_n \geq bU$. But the ε_j can be expressed in terms of X_1, \dots, X_j , so the optional skipping theorem applies; hence $\{Y_k, \mathcal{F}_k, k = 1, 2, \dots, n\}$ is a submartingale, and $E(Y_n) \leq E(X_n)$. Thus

$$E(U_{ab}) \leq \frac{1}{b} E(Y_n) \leq \frac{1}{b} E(X_n),$$

as asserted.

In general, $\{(X_k - a)^+, \mathcal{F}_k, k = 1, 2, \dots, n\}$ is a submartingale by 7.3.6(a), and the number of upcrossings of (a, b) by $\{X_j\}$ is the same as the number of upcrossings of $(0, b - a)$ by $\{(X_j - a)^+\}$ (note that $X_j \leq a$, $X_j - a \leq 0$, and $(X_j - a)^+ \leq 0$ are equivalent, as are $X_j \geq b$, $X_j - a \geq b - a$, and $(X_j - a)^+ \geq b - a$). The result follows from the above argument. ■

We now prove the main convergence theorem.

7.4.3 Submartingale Convergence Theorem. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a submartingale. If $\sup_n E(X_n^+) < \infty$, there is an integrable random variable X_∞ such that $X_n \rightarrow X_\infty$ almost everywhere.

PROOF

$P\{\omega: X_n(\omega) \text{ does not converge to a finite or infinite limit}\}$

$$= P\left[\bigcup_{\substack{a < b \\ a, b \text{ rational}}} \{\omega: \liminf_{n \rightarrow \infty} X_n(\omega) < a < b < \limsup_{n \rightarrow \infty} X_n(\omega)\}\right].$$

If for some $a < b$, $P\{\liminf_{n \rightarrow \infty} X_n < a < b < \limsup_{n \rightarrow \infty} X_n\} > 0$, then $\{X_n\}$ has an infinite number of upcrossings of (a, b) on a set of positive probability; hence $E(U_{ab}) = \infty$. But U_{ab} is the limit of the monotone sequence $U_{ab; n} =$ the number of upcrossings of (a, b) by X_1, \dots, X_n , so that $E(U_{ab; n}) \rightarrow E(U_{ab})$. But by 7.4.2,

$$E(U_{ab; n}) \leq (b - a)^{-1} E[(X_n - a)^+] \leq (b - a)^{-1} [\sup_n E(X_n^+) + a^-] < \infty,$$

a contradiction. Thus X_n converges to a limit X_∞ a.e. Now $|X_n| = X_n^+ + X_n^- = 2X_n^+ - X_n$, and $E(X_n) \geq E(X_1)$ by the submartingale property.

Therefore

$$E(|X_n|) \leq 2 \sup_n E(X_n^+) - E(X_1) < \infty.$$

By Fatou's lemma,

$$E(|X_\infty|) \leq \liminf_{n \rightarrow \infty} E(|X_n|) < \infty;$$

hence X_∞ is integrable, and therefore finite a.e. By changing X_∞ on a set of measure 0, if necessary, we may take X_∞ as a random variable (rather than an extended random variable). The theorem is proved. ■

7.4.4 Corollary. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a reverse submartingale (the \mathcal{F}_n decrease as n increases, and $E(X_n | \mathcal{F}_{n+1}) \geq X_{n+1}$ a.e.). If $\inf_n E(X_n) > -\infty$, there is an integrable random variable X_∞ such that $X_n \rightarrow X_\infty$ a.e. [Note that the hypothesis is satisfied for any reverse martingale since $E(X_n)$ is constant.]

PROOF. Proceed as in 7.4.3, but instead let $U_{ab;n}$ be the number of upcrossings of (a, b) by $\{X_n, X_{n-1}, \dots, X_1\}$, which is a submartingale because

$$E(X_k | X_{k+1}, \dots, X_n) = E[E(X_k | \mathcal{F}_{k+1}) | X_{k+1}, \dots, X_n] \geq X_{k+1}.$$

We obtain $E(U_{ab;n}) \leq (b-a)^{-1} E[(X_1 - a)^+] < \infty$, and thus $X_n \rightarrow X_\infty$ a.e. as before.

Now $|X_n| = 2X_n^+ - X_n$ and $E(X_n) \geq \inf_n E(X_n) > -\infty$. Also $\{X_n^+, \dots, X_1^+\}$ is a submartingale by 7.3.6(a), so $E(X_n^+) \leq E(X_1^+)$. Thus $E(|X_n|) \leq 2E(X_1^+) - \inf_n E(X_n) < \infty$, so X_∞ is integrable by Fatou's lemma as before. ■

7.4.5 Comments. (a) In 7.4.3 and 7.4.4, the proofs show that $\{X_n\}$ must be L^1 bounded, that is, $\sup_n E(|X_n|) < \infty$. Thus for a submartingale, $\sup_n E(X_n^+) < \infty$ is equivalent to L^1 boundedness, and implies convergence. However, a submartingale may converge without being L^1 bounded (see Problems 1 and 2).

(b) Results analogous to 7.4.3 and 7.4.4 hold for supermartingales: If $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ is a supermartingale and $\sup_n E(X_n^-) < \infty$, then there is an integrable random variable X_∞ such that $X_n \rightarrow X_\infty$ a.e. In particular, a nonnegative supermartingale converges a.e. If $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ is a reverse supermartingale and $\sup_n E(X_n) < \infty$, there is an integrable random variable X_∞ such that $X_n \rightarrow X_\infty$ a.e. The first statement follows from 7.4.3 since $\{-X_n, \mathcal{F}_n\}$ is a submartingale and $\sup_n E[(-X_n)^+] = \sup_n E(X_n^-)$. The second follows from 7.4.4 since $\{-X_n, \mathcal{F}_n\}$ is a reverse submartingale and $\inf_n E(-X_n) = -\sup_n E(X_n)$.

Problems

1. Consider the following Markov chain. Take $X_1 = 0$. If $X_n = 0$ (regardless of X_k , $k < n$), then:

$$\begin{aligned} X_{n+1} &= a_{n+1} && \text{with probability } p_{n+1} \\ &= -a_{n+1} && \text{with probability } p_{n+1} \\ &= 0 && \text{with probability } 1 - 2p_{n+1}, \end{aligned}$$

where $0 < p_{n+1} < \frac{1}{2}$ and the a_n are distinct and greater than 0. If $X_n \neq 0$, take $X_{n+1} = X_n$ (thus if $X_n \neq 0$, we have $X_j = X_n$ for all $j \geq n$).

- (a) Show that $\{X_n\}$ is a martingale, and X_n converges everywhere.
 (b) If $\sum_{k=1}^{\infty} p_k < \infty$ and $\sum_{k=1}^{\infty} a_k p_k = \infty$, show that $\sup_n E(|X_n|) = \infty$.

2. (Problem by W. F. Stout) Consider the following Markov chain. Take $X_0 = 0$, and let

$$\begin{aligned} P\{X_{n+1} = n+1 | X_n = n\} &= p_{n+1}, \\ P\{X_{n+1} = -(n+1) | X_n = n\} &= 1 - p_{n+1}, \\ P\{X_{n+1} = -k | X_n = -k\} &= 1, \quad (n = 0, 1, \dots, \quad k = 1, 2, \dots). \end{aligned}$$

- (a) Show that if $p_{n+1} = (2n+1)/(2n+2)$ for all n , then $\{X_n\}$ is a martingale.
 (b) If the p_n are chosen as in (a), show that the martingale converges a.e. to a finite limit, although $E(|X_n|) \rightarrow \infty$.

(Note: In Problems 1 and 2, the Markov chain has nonstationary transition probabilities, in other words, the probability of moving from state i at time n to state j at time $n+1$ depends on n . However, the basic construction of 5.11.2 carries over.)

3. (Kemeny, Snell, and Knapp, 1966) Let $\{X_n\}$ be a Markov chain with state space $S =$ the set of rationals in $(0, 1)$, and the following transition probabilities:

Let $0 < b \leq a < 1$, a, b rational. If $x \in S$ and $X_n = x$, then $X_{n+1} = bx$ with probability $1 - x$, and $X_{n+1} = bx + 1 - a$ with probability x .

- (a) If $a = b$, show that $\{X_n\}$ is a martingale, and $X_n \rightarrow X_{\infty}$ a.e., where $X_{\infty} = 0$ or 1; also $P\{X_{\infty} = 1\} = E(X_0)$.

- (b) If $b < a$, show that $\{X_n\}$ is a supermartingale and $X_n \rightarrow 0$ a.e.

4. (Polya urn scheme) An urn contains white and black balls; one ball is drawn, and then replaced by two of the same color, and the process is repeated. Thus if the urn contains c white and $r - c$ black balls, and a white ball is drawn, the fraction of white balls in the urn before the next drawing is $(c+1)/(r+1)$.

If X_n is the fraction of white balls in the urn must before the n th drawing, show that $\{X_n\}$ is a martingale, and X_n converges a.e. to a limit X_∞ , where $E(X_\infty) = E(X_1)$.

5. Martingales may be defined on a measure space $(\Omega, \mathcal{F}, \mu)$ if μ is finite; simply replace μ by the probability measure $P = \mu/[\mu(\Omega)]$. If $\mu(\Omega) = \infty$ we can use 7.3.2(c) in the definition: $\int_A X_n d\mu = \int_A X_{n+1} d\mu$ for all $A \in \mathcal{F}_n$ [of course, the X_n are still required to be \mathcal{F}_n -measurable and integrable]. Sub- and supermartingales may be defined similarly.

Show that an L^1 bounded submartingale converges a.e. [μ] to an integrable limit.

7.5 Uniform Integrability

We now introduce a concept that has important application to martingale theory, and in fact to integration theory in general.

7.5.1 Definitions and Comments. Let f_1, f_2, \dots be real- or complex-valued Borel measurable functions on the measure space $(\Omega, \mathcal{F}, \mu)$, μ finite. The f_n are said to be *uniformly integrable* iff

$$\int_{\{|f_n| \geq c\}} |f_n| d\mu \rightarrow 0 \quad \text{as } c \rightarrow \infty$$

uniformly in n .

It is immediate that if the f_n are uniformly integrable, each f_n is integrable. Also, if $|f_n| \leq g$ for all n , where g is integrable, in particular, if the f_n are uniformly bounded, then the f_n are uniformly integrable.

Furthermore, if the f_n are uniformly integrable, then $\sup_n \int_\Omega |f_n| d\mu < \infty$, because if $\varepsilon > 0$,

$$\int_\Omega |f_n| d\mu = \int_{\{|f_n| \geq c\}} |f_n| d\mu + \int_{\{|f_n| < c\}} |f_n| d\mu \leq \varepsilon + c\mu(\Omega)$$

for large n .

One basic application of uniform integrability is the following extension of Fatou's lemma and the dominated convergence theorem.

7.5.2 Theorem. Let f_1, f_2, \dots be real-valued and uniformly integrable.

$$\begin{aligned} \text{(a)} \quad \int_\Omega \left(\liminf_n f_n \right) d\mu &\leq \liminf_n \int_\Omega f_n d\mu \\ &\leq \limsup_n \int_\Omega f_n d\mu \leq \int_\Omega \left(\limsup_n f_n \right) d\mu. \end{aligned}$$

(b) If $f_n \rightarrow f$ a.e. or in measure, then f is integrable and

$$\int_{\Omega} f_n d\mu \rightarrow \int_{\Omega} f d\mu.$$

PROOF. (a) We have

$$\int_{\Omega} f_n d\mu = \int_{\{f_n < -c\}} f_n d\mu + \int_{\{f_n \geq -c\}} f_n d\mu, \quad c > 0.$$

By uniform integrability, c may be chosen so large that $|\int_{\{f_n < -c\}} f_n d\mu| < \varepsilon$ for all n , where $\varepsilon > 0$ is preassigned. Since $f_n I_{\{f_n \geq -c\}} \geq -c$, which is integrable since μ is finite, Fatou's lemma yields

$$\liminf_n \int_{\{f_n \geq -c\}} f_n d\mu \geq \int_{\Omega} \liminf_n (f_n I_{\{f_n \geq -c\}}) d\mu.$$

Since $f_n I_{\{f_n \geq -c\}} \geq f_n$, this integral is in turn greater than or equal to $\int_{\Omega} (\liminf_n f_n) d\mu$. Thus

$$\liminf_n \int_{\Omega} f_n d\mu \geq \int_{\Omega} \left(\liminf_n f_n \right) d\mu - \varepsilon,$$

proving the “ \liminf ” part. The “ \limsup ” part is done by a symmetrical argument.

(b) This is immediate from (a) if $f_n \rightarrow f$ a.e., so assume $f_n \rightarrow f$ in measure. By 2.5.3, there is a subsequence $f_{n_k} \rightarrow f$ a.e.; hence by (a) applied to the f_{n_k} , f is integrable and $\int_{\Omega} f_{n_k} d\mu \rightarrow \int_{\Omega} f d\mu$. If $\int_{\Omega} f_n d\mu$ does not converge to $\int_{\Omega} f d\mu$, then for some $\varepsilon > 0$ we have $|\int_{\Omega} f_n d\mu - \int_{\Omega} f d\mu| \geq \varepsilon$ for infinitely many n , and for convenience in notation we may assume this holds for all n . But then we find a subsequence $f_{m_j} \rightarrow f$ a.e., and as above, $\int_{\Omega} f_{m_j} d\mu \rightarrow \int_{\Omega} f d\mu$, a contradiction. ■

We now establish a useful criterion for uniform integrability.

7.5.3 Theorem. The complex-valued Borel measurable functions f_n are uniformly integrable iff the integrals $\int_{\Omega} |f_n| d\mu$ are uniformly bounded and also uniformly continuous, that is, $\int_{\mathcal{A}} |f_n| d\mu \rightarrow 0$ as $\mu(\mathcal{A}) \rightarrow 0$, uniformly in n .

PROOF. Assume the integrals are uniformly bounded and uniformly continuous. Then

$$\mu\{|f_n| \geq c\} \leq \frac{1}{c} \int_{\Omega} |f_n| d\mu$$

by Chebyshev's inequality, and this approaches 0 as $c \rightarrow \infty$, uniformly in n , by the uniform boundedness. Thus $\int_{\{|f_n| \geq c\}} |f_n| d\mu \rightarrow 0$ as $c \rightarrow \infty$, uniformly in n , by the uniform continuity.

Conversely, assume uniform integrability. We have

$$\begin{aligned} \int_A |f_n| d\mu &= \int_{A \cap \{|f_n| \geq c\}} |f_n| d\mu + \int_{A \cap \{|f_n| < c\}} |f_n| d\mu \\ &\leq \int_{\{|f_n| \geq c\}} |f_n| d\mu + c\mu(A). \end{aligned} \quad (1)$$

Choose c so that $\int_{\{|f_n| \geq c\}} |f_n| d\mu < \varepsilon/2$ for all n ; if $\mu(A) < \varepsilon/2c$, then by (1), $\int_A |f_n| d\mu < (\varepsilon/2) + (\varepsilon/2) = \varepsilon$ for all n , proving uniform continuity. Uniform boundedness was verified in 7.5.1. ■

We have seen in 2.5.1 that L^p convergence implies convergence in measure. The converse holds under an additional hypothesis of uniform integrability.

7.5.4 Theorem. Let μ be a finite measure on (Ω, \mathcal{F}) , and let $0 < p < \infty$. If $f_n \xrightarrow{\mu} f$ and the $|f_n|^p$ are uniformly integrable, then $f_n \xrightarrow{L^p} f$.

PROOF. First assume that the $|f_n - f|^p$ are uniformly integrable. By 2.5.3, there is a subsequence $\{f_{n_k}\}$ converging to f a.e. and in measure. By 7.5.2(b), $\int_{\Omega} |f_{n_k} - f|^p d\mu \rightarrow 0$ as $k \rightarrow \infty$. The same argument shows that any subsequence of $\{f_n\}$ has a subsequence converging to f in L^p . Hence $f_n \xrightarrow{L^p} f$, for if not, there would be an $\varepsilon > 0$ and a subsequence $\{f_{n_i}\}$ such that $\int_{\Omega} |f_{n_i} - f|^p d\mu \geq \varepsilon$ for all i .

Now assume the $|f_n|^p$ to be uniformly integrable. Then $|f_n - f|^p \leq |f_n|^p + |f|^p$ if $p \leq 1$, and $|f_n - f|^p \leq 2^{p-1}(|f_n|^p + |f|^p)$ if $p \geq 1$. (See 2.4.6 and the end of 2.4.12.) As above, we have a subsequence $f_{n_k} \rightarrow f$ a.e. By 7.5.2(b), $|f|^p$ is integrable, and it follows that the integrals $\int_{\Omega} |f_n - f|^p d\mu$ are uniformly bounded and uniformly continuous. By 7.5.3, the $|f_n - f|^p$ are uniformly integrable, and the previous argument applies. ■

7.5.5 Corollary. In 7.5.2(b), $f_n \xrightarrow{L^1} f$, that is, $\int_{\Omega} |f_n - f| d\mu \rightarrow 0$.

PROOF. The $|f_n|$ are uniformly integrable by hypothesis, and $f_n \xrightarrow{\mu} f$ either by hypothesis or by 2.5.5. ■

Problems

- If $0 < p < \infty$, $f, f_1, f_2, \dots \in L^p$, and $f_n \xrightarrow{L^p} f$, show that the $|f_n|^p$ are uniformly integrable.
- Give an example of a uniformly integrable sequence of functions f_n on a measure space $(\Omega, \mathcal{F}, \mu)$, μ finite, such that the $|f_n|$ cannot be bounded above by an integrable function.

7.6 Uniform Integrability and Martingale Theory

The application of the uniform integrability concept yields many additional facts about martingales. In particular, we shall find that a martingale is uniformly integrable iff it can be represented as $X_n = E(Y | \mathcal{F}_n)$ for some fixed Y . Also, we shall see that uniform integrability is related to the problem of adjoining a random variable X_∞ (a so-called “last element”) to a martingale so that the martingale property is preserved, in other words, $E(X_\infty | \mathcal{F}_n) = X_n$.

7.6.1 Lemma. Let Y be an integrable random variable on (Ω, \mathcal{F}, P) and let \mathcal{G}_i , $i \in I$, be arbitrary sub σ -fields of \mathcal{F} . Then the random variables $X_i = E(Y | \mathcal{G}_i)$, $i \in I$, are uniformly integrable, that is,

$$\int_{\{|X_i| \geq c\}} |X_i| dP \rightarrow 0 \quad \text{as } c \rightarrow \infty$$

uniformly in i .

PROOF. Since $|E(Y | \mathcal{G}_i)| \leq E(|Y| | \mathcal{G}_i)$,

$$\int_{\{|X_i| \geq c\}} |X_i| dP \leq \int_{\{|X_i| \geq c\}} E(|Y| | \mathcal{G}_i) dP = \int_{\{|X_i| \geq c\}} |Y| dP$$

since $\{|X_i| \geq c\} \in \mathcal{G}_i$ (remember X_i is \mathcal{G}_i -measurable). But by Chebyshev’s inequality,

$$P(|X_i| \geq c) \leq c^{-1} E(|X_i|) \leq c^{-1} E[E(|Y| | \mathcal{G}_i)] = c^{-1} E(|Y|) \rightarrow 0 \quad \text{as } c \rightarrow \infty$$

uniformly in i . ■

The following result, due to Lévy, was historically the first of the martingale convergence theorems.

7.6.2 Theorem. Let $\{\mathcal{F}_n\}$ be an increasing sequence of sub σ -fields of \mathcal{F} , and let \mathcal{F}_∞ be the σ -field generated by $\bigcup_{n=1}^\infty \mathcal{F}_n$. If Y is integrable and $X_n = E(Y | \mathcal{F}_n)$, $n = 1, 2, \dots$, then $X_n \rightarrow E(Y | \mathcal{F}_\infty)$ a.e. and in L^1 .

PROOF. $\{X_n, \mathcal{F}_n\}$ is a martingale by 7.3.3(c), and is uniformly integrable by 7.6.1. Since $E(|X_n|) \leq E(|Y|) < \infty$, X_n converges a.e. to an integrable random variable X_∞ , by 7.4.3; L^1 convergence follows from 7.5.5. It remains to show that $X_\infty = E(Y | \mathcal{F}_\infty)$ a.e.

If $A \in \mathcal{F}_n$, then

$$\int_A Y dP = \int_A E(Y|\mathcal{F}_n) dP = \int_A X_n dP \rightarrow \int_A X_\infty dP$$

by L^1 convergence. Thus $\int_A Y dP = \int_A X_\infty dP$ for all A in the field $\bigcup_{n=1}^\infty \mathcal{F}_n$, and hence for all $A \in \mathcal{F}_\infty$ (monotone class theorem). Since X_n is $\mathcal{F}_n \subset \mathcal{F}_\infty$ -measurable, X_∞ is \mathcal{F}_∞ -measurable, and therefore $X_\infty = E(Y|\mathcal{F}_\infty)$ a.e. by definition of conditional expectation. ■

A result similar to 7.6.2 holds if the σ -fields form a decreasing sequence.

7.6.3 Theorem. Let $\{\mathcal{F}_n\}$ be a decreasing sequence of sub σ -fields of \mathcal{F} , and let $\mathcal{F}_\infty = \bigcap_{n=1}^\infty \mathcal{F}_n$. If Y is integrable and $X_n = E(Y|\mathcal{F}_n)$, $n = 1, 2, \dots$, then $X_n \rightarrow E(Y|\mathcal{F}_\infty)$ a.e. and in L^1 .

PROOF. Just as in 7.6.2 (using 7.4.4 instead of 7.4.3), $X_n \rightarrow X_\infty$ a.e. and in L^1 , so we must show that $X_\infty = E(Y|\mathcal{F}_\infty)$ a.e. If $A \in \mathcal{F}_\infty \subset \mathcal{F}_n$, then

$$\int_A Y dP = \int_A E(Y|\mathcal{F}_n) dP = \int_A X_n dP \rightarrow \int_A X_\infty dP.$$

Since X_n is $\mathcal{F}_n \subset \mathcal{F}_k$ -measurable for $n \geq k$, X_∞ is \mathcal{F}_k -measurable for all k ; hence X_∞ is \mathcal{F}_∞ -measurable. ■

7.6.4 Comments. Let $Z_i : (\Omega, \mathcal{F}) \rightarrow (\Omega'_i, \mathcal{F}'_i)$ be a random object ($i = 1, 2, \dots$). In 7.6.2, if $\mathcal{F}_n = \mathcal{F}(Z_1, \dots, Z_n)$, then $\mathcal{F}_\infty = \mathcal{F}(Z_1, Z_2, \dots)$ (see Problem 1). Thus $E(Y|Z_1, \dots, Z_n) \rightarrow E(Y|Z_1, Z_2, \dots)$ a.e. and in L^1 . In 7.6.3, $\mathcal{F}_n = \mathcal{F}(Z_n, Z_{n+1}, \dots)$ is a common special case. Then \mathcal{F}_∞ is the tail σ -field of the Z_n (see 7.2.6).

We now show that uniform integrability of a submartingale implies a.e. and L^1 convergence, and also implies that a last element can be attached.

7.6.5 Theorem. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a uniformly integrable submartingale. Then $\sup_n E(X_n^+) < \infty$, and X_n converges to a limit X_∞ a.e. and in L^1 . Furthermore, if \mathcal{F}_∞ is the σ -field generated by $\bigcup_{n=1}^\infty \mathcal{F}_n$, then $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ is a submartingale. If $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ is a uniformly integrable martingale, so is $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$. [If $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ is a (sub- or super-) martingale, where \mathcal{F}_∞ is any σ -field including all the \mathcal{F}_n , X_∞ is said to be a *last element*.]

PROOF. By 7.5.3, $\sup_n E(|X_n|) < \infty$, so by 7.4.3, $X_n \rightarrow X_\infty$ a.e. By 7.5.5, $X_n \xrightarrow{L^1} X_\infty$.

Now if $A \in \mathcal{F}_n$ and $k \geq n$, then by 7.3.2(c), $\int_A X_n dP \leq \int_A X_k dP$. Let $k \rightarrow \infty$; the L^1 -convergence yields $\int_A X_n dP \leq \int_A X_\infty dP$. Thus [7.3.2(c) again] $X_n \leq E(X_\infty | \mathcal{F}_n)$; hence $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ is a submartingale. The last statement follows from the above argument with “ \leq ” replaced by “ $=$.” ■

Theorem 7.6.5 indicates that uniform integrability of a submartingale implies some rather strong conclusions. In fact, a uniformly integrable martingale must assume a very special form.

7.6.6 Theorem. $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ is a uniformly integrable martingale iff there is an integrable random variable Y such that $X_n = E(Y | \mathcal{F}_n)$, $n = 1, 2, \dots$; in this case, $X_n \rightarrow E(Y | \mathcal{F}_\infty)$ a.e. and in L^1 , where \mathcal{F}_∞ is the σ -field generated by $\bigcup_{n=1}^\infty \mathcal{F}_n$.

PROOF. The “if” part follows from 7.6.1 and 7.6.2. The “only if” part follows from 7.6.5 with $Y = X_\infty$. ■

In 7.6.6, if we require that Y be \mathcal{F}_∞ -measurable, then Y is unique (up to sets of measure 0). For if $X_n = E(Y | \mathcal{F}_n)$, then $E(Y | \mathcal{F}_n) = E(X_\infty | \mathcal{F}_n)$, $n = 1, 2, \dots$, so $\int_A Y dP = \int_A X_\infty dP$ for all $A \in \bigcup_{n=1}^\infty \mathcal{F}_n$, and hence for all $A \in \mathcal{F}_\infty$ (monotone class theorem). Thus $Y = X_\infty$ a.e. by 1.6.11.

A sub- or supermartingale with a last element need not be uniformly integrable, but there are partial results in this direction. If $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ is a martingale with a last element, then $X_n = E(X_\infty | \mathcal{F}_n)$, so the X_n are uniformly integrable. We can say a bit more.

7.6.7 Theorem. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ be a nonnegative submartingale with a last element. Then the X_n are uniformly integrable.

PROOF.

$$\int_{\{X_n \geq c\}} X_n dP \leq \int_{\{X_n \geq c\}} X_\infty dP,$$

and

$$P\{X_n \geq c\} \leq \frac{E(X_n)}{c} \leq \frac{E(X_\infty)}{c} \rightarrow 0 \quad \text{as } c \rightarrow \infty$$

uniformly in n . ■

We now give an example of a supermartingale $\{X_n, \mathcal{F}_n\}$ with a last element that is not uniformly integrable. (Also, $\{-X_n, \mathcal{F}_n\}$ will be a submartingale with a last element that is not uniformly integrable.)

7.6.8 Example. Let Y_1, Y_2, \dots be independent, with $P\{Y_j = 1\} = p$, $P\{Y_j = 0\} = 1 - p$, $0 < p < 1$. Let $X_n = p^{-n} \prod_{j=1}^n Y_j$, $\mathcal{F}_n = \mathcal{F}(Y_1, \dots, Y_n)$. Then $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ is a martingale, and hence a supermartingale [see 7.3.3(b)]. Since all $X_n \geq 0$, we have $E(0|\mathcal{F}_n) = 0 \leq X_n$, so 0 is a last element when the sequence is regarded as a supermartingale, but not when it is regarded as a martingale. Note also that any negative constant can be taken as a last element, so the last element is not unique, and there need not be convergence to the last element. But the X_n are not uniformly integrable; for $P\{Y_j = 1 \text{ for all } j\} = \lim_{n \rightarrow \infty} p^n = 0$; hence (a.e.) $X_n = 0$ eventually, so $X_n \rightarrow 0$ a.e. If the X_n were uniformly integrable, then $X_n \xrightarrow{L^1} 0$ by 7.5.5; hence $E(X_n) \rightarrow 0$. But $E(X_n) = 1$ for all n , a contradiction. If we regard the sequence as a martingale, there can be no last element. For if X_∞ is a last element, then $X_n = E(X_\infty|\mathcal{F}_n)$ for all n ; hence by 7.6.1, the X_n are uniformly integrable.

Since (a.e.) $X_n = 0$ eventually, we have an example of a “fair” game in which the gambler is almost certain to be wiped out. Thus the term “locally fair” is perhaps more appropriate than “fair.”

Note that a sub- or supermartingale with a last element converges a.e. [In the submartingale case, for example, $\sup_n E(X_n^+) \leq E(X_\infty^+) < \infty$.] But the limit need not coincide with the last element.

We now look at the problem of L^p convergence.

7.6.9 Lemma. Let $X_i, i \in I$, be integrable random variables on (Ω, \mathcal{F}, P) . If $h: [0, \infty) \rightarrow [0, \infty)$ is Borel measurable, $h(t)/t \rightarrow \infty$ as $t \rightarrow \infty$, and $\sup_{i \in I} E[h(|X_i|)] < \infty$, then the X_i are uniformly integrable. For example, take $h(t) = t^p$, $p > 1$; thus if $E(|X_n|^p) \leq M \leq \infty$ for all $n = 1, 2, \dots$, the X_n are uniformly integrable.

PROOF. Given $\varepsilon > 0$, let $M = \sup_{i \in I} E[h(|X_i|)]$ and set $a = M/\varepsilon$. There is a positive number c such that $h(t)/t \geq a$ for $t \geq c$. Thus

$$\int_{\{|X_i| \geq c\}} |X_i| dP \leq \int_{\{|X_i| \geq c\}} \frac{h(|X_i|)}{a} dP \leq \int_{\Omega} \frac{h(|X_i|)}{a} dP \leq \frac{M}{a} = \varepsilon. \blacksquare$$

7.6.10 Theorem. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a martingale or a nonnegative submartingale with $E[|X_n|^p] \leq M < \infty$ for all n , where $p > 1$. Then X_n converges to a limit X_∞ a.e. and in L^p .

PROOF. By 7.6.9, the X_n are uniformly integrable, so by 7.6.5, X_n converges a.e. to a limit X_∞ , and X_∞ is a last element.

Now $\{|X_n|^p, \mathcal{F}_n, n = 1, 2, \dots, \infty\}$ is a nonnegative submartingale. [In the case in which $\{X_n, \mathcal{F}_n\}$ is a martingale, use 7.3.6(b) with $g(x) = |x|^p$; in the nonnegative submartingale case, use 7.3.6(a) with $g(x) = x^p, x \geq 0; g(x) = 0, x < 0$.] By 7.6.7, the $|X_n|^p$ are uniformly integrable, and by 7.5.4, $X_n \rightarrow X_\infty$ in L^p . ■

Problems

1. Let $Z_n : (\Omega, \mathcal{F}) \rightarrow (\Omega'_n, \mathcal{F}'_n)$, $n = 1, 2, \dots$, be random objects. If $\mathcal{F}_n = \mathcal{F}(Z_1, \dots, Z_n)$, $n = 1, 2, \dots$, show that the σ -field generated by $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ is $\mathcal{F}(Z_1, Z_2, \dots)$.
2. Let $\{X_n, \mathcal{F}_n\}$ be a nonnegative supermartingale, so that X_n converges a.e. to an integrable random variable X_∞ .
 - (a) Show that $E(X_n) \rightarrow E(X_\infty)$ iff the X_n are uniformly integrable. (This holds for any sequence of nonnegative integrable random variables converging a.e. to an integrable limit; the supermartingale property is not involved. Note also that we have $E[|X_n - X_\infty|] \rightarrow 0$ since we have a.e. convergence and uniform integrability.)
 - (b) Show that X_∞ is a last element.
 - (c) If $E(X_n) \rightarrow 0$, show that $X_n \rightarrow 0$ a.e.

7.7 Optional Sampling Theorems

Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a martingale, with X_n interpreted as a gambler's total capital after n plays of a game of chance. Suppose that after each trial, the gambler decides either to quit or to keep playing. If T is the time at which he quits, what can be said about his final capital X_T ?

First of all, the random variable T must have the property that if we observe X_1, \dots, X_n , we can come to a definite decision as to whether or not $T = n$. A nonnegative random variable of this type is called a *stopping time*.

7.7.1 Definition. Let $\{\mathcal{F}_n, n = 0, 1, \dots\}$ be an increasing sequence of sub σ -fields of \mathcal{F} . A *stopping time* for the \mathcal{F}_n is a map $T : \Omega \rightarrow \{0, 1, \dots, \infty\}$ such that $\{T \leq n\} \in \mathcal{F}_n$ for each nonnegative integer n . Since $\{T = n\} = \{T \leq n\} - \{T \leq n-1\}$ and $\{T \leq n\} = \bigcup_{k=0}^n \{T = k\}$, the definition is equivalent to the requirement that $\{T = n\} \in \mathcal{F}_n$ for all $n = 0, 1, \dots$. If $\{X_n, n = 0, 1, \dots\}$ is a sequence of random variables, a stopping time for $\{X_n\}$ is, by definition, a stopping time relative to the σ -fields $\mathcal{F}_n = \mathcal{F}(X_0, \dots, X_n)$. (The above

definitions are modified in the obvious way if the index n starts from 1 rather than 0.)

By far the most important example of a stopping time is the *hitting time* of a set. If $\{X_n\}$ is a sequence of random variables and $B \in \mathcal{B}(R)$, let $T(\omega) = \min\{n: X_n(\omega) \in B\}$ if $X_n(\omega) \in B$ for some n ; $T(\omega) = \infty$ if $X_n(\omega)$ is never in B . T is a stopping time since $\{T \leq n\} = \bigcup_{k \leq n} \{X_k \in B\} \in \mathcal{F}(X_k, k \leq n)$.

If T is a stopping time for $\{X_n\}$, an event A is said to be "prior to T " iff, whenever $T = n$, we can tell by examination of the X_k , $k \leq n$, whether or not A has occurred. The formal definition is as follows.

7.7.2 Definition. Let T be a stopping time for the σ -fields \mathcal{F}_n , $n = 0, 1, \dots$, and let A belong to \mathcal{F} . The set A is said to be *prior to T* iff $A \cap \{T \leq n\} \in \mathcal{F}_n$ for all $n = 0, 1, \dots$ [Equivalently, as in 7.7.1, $A \cap \{T = n\} \in \mathcal{F}_n$ for all $n = 0, 1, \dots$] The collection of all sets prior to T will be denoted by \mathcal{F}_T ; it follows quickly that \mathcal{F}_T is a σ -field.

The σ -field \mathcal{F}_T arises very often in the following way. Let T be a finite stopping time for $\{X_n\}$, and define X_T in the natural way; if $T(\omega) = n$, let $X_T(\omega) = X_n(\omega)$. If $B \in \mathcal{B}(R)$, then $\{X_T \in B\} \in \mathcal{F}_T$, in other words, X_T is \mathcal{F}_T -measurable. (Since $\mathcal{F}_T \subset \mathcal{F}$ by definition, it follows in particular that X_T is a random variable.) To see this, write

$$\{X_T \in B\} \cap \{T \leq n\} = \bigcup_{k=0}^n [\{X_k \in B\} \cap \{T = k\}].$$

Since $\{X_k \in B\} \cap \{T = k\} \in \mathcal{F}(X_0, \dots, X_n)$ for $k \leq n$, we have

$$\{X_T \in B\} \cap \{T \leq n\} \in \mathcal{F}(X_0, \dots, X_n).$$

Also, since T is finite, we have $\bigcup_{n=0}^{\infty} \{T \leq n\} = \Omega$, so that $\{X_T \in B\} \in \mathcal{F}$, as desired.

Now in the gambling situation described at the beginning of the section, a basic quantity of interest is $E(X_T)$, the average accumulation at the quitting time. For example, if $E(X_T)$ turns out to be the same as $E(X_1)$ [$= E(X_n)$ for all n by the martingale property], the gambler's strategy does not offer any improvement over the procedure of stopping at a fixed time. Now in comparing X_1 and X_T we are considering two stopping times S and T ($S \equiv 1$) with $S \leq T$, and looking at X_S versus X_T . More generally, if $T_1 \leq T_2 \leq \dots$ form an increasing sequence of finite stopping times, we may examine the sequence X_{T_1}, X_{T_2}, \dots . If the sequence forms a martingale, then $E(X_{T_n}) = E(X_{T_1})$ for all n , and if $T_1 \equiv 1$, then $E(X_{T_1}) = E(X_1)$.

Thus if we sample the gambler's fortune at random times T_1, T_2, \dots , the basic question is whether the martingale (or submartingale) property is preserved. The following result gives useful sufficient conditions.

7.7.3 Optional Sampling Theorem. Let $\{X_1, X_2, \dots\}$ be a submartingale, and let T_1, T_2, \dots be an increasing sequence of finite stopping times for $\{X_n\}$, with $Y_n = X_{T_n}, n = 1, 2, \dots$. If

- (A) $E(|Y_n|) < \infty$ for all n , and
- (B) $\liminf_{k \rightarrow \infty} \int_{\{T_n > k\}} |X_k| dP = 0$ for all n ,

then $\{Y_n\}$ is a submartingale relative to the σ -fields \mathcal{F}_{T_n} . If $\{X_n\}$ is a martingale, so is $\{Y_n\}$.

PROOF. We follow Breiman (1968). First note that the \mathcal{F}_{T_n} increases with n ; for if $A \in \mathcal{F}_{T_n}$, then

$$A \cap \{T_{n+1} \leq k\} = \bigcup_{i=1}^k [A \cap \{T_n = i\}] \cap \{T_{n+1} \leq k\}.$$

But $A \cap \{T_n = i\} \in \mathcal{F}(X_1, \dots, X_i) \subset \mathcal{F}(X_1, \dots, X_k)$, and $\{T_{n+1} \leq k\} \in \mathcal{F}(X_1, \dots, X_k)$; hence $A \in \mathcal{F}_{T_{n+1}}$.

Now if $A \in \mathcal{F}_{T_n}$, we must show that $\int_A Y_{n+1} dP \geq \int_A Y_n dP$ (with equality in the martingale case). Since $A = \bigcup_j [A \cap \{T_n = j\}]$, it suffices to replace A by $D_j = A \cap \{T_n = j\}$, which belongs to $\mathcal{F}(X_1, \dots, X_j)$. Now if $k > j$, we note that $T_n = j$ implies $T_{n+1} \geq j$, so that

$$\int_{D_j} Y_{n+1} dP = \sum_{i=j}^k \int_{D_j \cap \{T_{n+1} = i\}} Y_{n+1} dP + \int_{D_j \cap \{T_{n+1} > k\}} Y_{n+1} dP.$$

Thus

$$\begin{aligned} \int_{D_j} Y_{n+1} dP &= \sum_{i=j}^k \int_{D_j \cap \{T_{n+1} = i\}} X_i dP + \int_{D_j \cap \{T_{n+1} > k\}} X_k dP \\ &\quad - \int_{D_j \cap \{T_{n+1} > k\}} (X_k - Y_{n+1}) dP. \end{aligned} \tag{1}$$

Now combine the $i = k$ term in (1) with the $\int X_k dP$ term to obtain

$$\begin{aligned} \int_{D_j \cap \{T_{n+1} = k\}} X_k dP + \int_{D_j \cap \{T_{n+1} > k\}} X_k dP &= \int_{D_j \cap \{T_{n+1} \geq k\}} X_k dP \\ &\geq \int_{D_j \cap \{T_{n+1} \geq k\}} X_{k-1} dP \end{aligned}$$

since $\{T_{n+1} \geq k\} = \{T_{n+1} \leq k-1\}^c \in \mathcal{F}(X_1, \dots, X_{k-1})$ and $D_j \in \mathcal{F}(X_1, \dots, X_j) \subset \mathcal{F}(X_1, \dots, X_{k-1})$. But

$$\int_{D_j \cap \{T_{n+1} \geq k\}} X_{k-1} dP = \int_{D_j \cap \{T_{n+1} > k-1\}} X_{k-1} dP,$$

so this term may be combined with the $i = k-1$ term of (1) to obtain

$$\int_{D_j \cap \{T_{n+1} > k-2\}} X_{k-2} dP.$$

Proceeding inductively, we find

$$\int_{D_j} Y_{n+1} dP \geq \int_{D_j \cap \{T_{n+1} \geq j\}} X_j dP - \int_{D_j \cap \{T_{n+1} > k\}} (X_k - Y_{n+1}) dP.$$

Now by hypothesis, $\int_{D_j \cap \{T_{n+1} > k\}} X_k dP \rightarrow 0$ as $k \rightarrow \infty$ through an appropriate subsequence, and $\int_{D_j \cap \{T_{n+1} > k\}} Y_{n+1} dP \rightarrow 0$ as $k \rightarrow \infty$ since $\{T_{n+1} > k\}$ decreases to the empty set. Finally, $D_j \cap \{T_{n+1} \geq j\} = D_j$ since $D_j \subset \{T_n = j\}$, and $X_j = Y_n$ on D_j . Thus

$$\int_{D_j} Y_{n+1} dP \geq \int_{D_j} Y_n dP$$

as desired. In the martingale case, all inequalities in the proof become equalities. ■

Here are some common special cases.

7.7.4 Theorem. Let $\{X_n\}$ be a submartingale, $\{T_n\}$ an increasing sequence of finite stopping times for $\{X_n\}$. Conditions (A) and (B) of 7.7.3 hold in each of the following situations.

- (a) Each T_n is bounded, that is, for each n there is a positive constant K_n such that $T_n \leq K_n$ a.e.
- (b) $E(\sup_n |X_n|) < \infty$ (in particular, this holds if the X_n are uniformly bounded).

PROOF. (a) $\int_{\Omega} |X_{T_n}| dP = \sum_{i \leq K_n} \int_{\{T_n = i\}} |X_i| dP \leq \sum_{i \leq K_n} E(|X_i|) < \infty$,

proving (A). Condition (B) holds because $\{T_n > k\}$ is empty for sufficiently large k .

(b) If $Z = \sup_n |X_n|$, then $\int_{\Omega} |X_{T_n}| dP \leq E(Z) < \infty$, proving (A). To prove (B), write

$$\int_{\{T_n > k\}} |X_k| dP \leq \int_{\{T_n > k\}} Z dP \rightarrow 0 \quad \text{as } k \rightarrow \infty$$

because $\{T_n > k\} \downarrow \emptyset$ and Z is integrable. ■

Theorem 7.7.4(b) admits the following generalization.

7.7.5 Theorem. If $\{T_n\}$ is an increasing sequence of finite stopping times for a submartingale $\{X_n\}$ having a last element X_∞ , then $\{X_{T_n}\}$ is a submartingale relative to the σ -fields \mathcal{F}_{T_n} ; if $\{X_n\}$ is a martingale, so is $\{X_{T_n}\}$. In particular, this holds if the X_n are uniformly integrable (see 7.6.5).

PROOF. *Case 1:* $X_n \leq 0$ for all n , and $X_\infty = 0$. For any fixed n , let $S_k = T_n \wedge k = \min(T_n, k)$, $k = 1, 2, \dots$; it is easily checked that S_k is a stopping time for the X_n . Now $X_{S_k} \rightarrow Y_n = X_{T_n}$ as $k \rightarrow \infty$; hence by Fatou's lemma,

$$\limsup_{k \rightarrow \infty} \int_{\Omega} X_{S_k} dP \leq \int_{\Omega} Y_n dP \leq 0.$$

But by 7.7.4(a), $\{X_{S_k}\}$ is a submartingale; hence

$$\int_{\Omega} X_{S_k} dP \geq \int_{\Omega} X_{S_1} dP = \int_{\Omega} X_1 dP,$$

which is finite. Therefore Y_n is integrable.

Again by 7.7.4(a), $\{X_{T_n \wedge k}, \mathcal{F}_{T_n \wedge k}, n = 1, 2, \dots\}$ is a submartingale (k fixed). Thus

$$\int_A X_{T_n \wedge k} dP \leq \int_A X_{T_{n+1} \wedge k} dP$$

if $A \in \mathcal{F}_{T_n \wedge k}$. But if $A \in \mathcal{F}_{T_n}$, then $A \cap \{T_n \leq k\} \in \mathcal{F}_{T_n \wedge k}$, for

$$A \cap \{T_n \leq k\} \cap \{T_n \wedge k \leq i\} = \begin{cases} A \cap \{T_n \leq i\} & \text{for } i \leq k, \\ A \cap \{T_n \leq k\} & \text{for } i > k. \end{cases}$$

Thus

$$\int_{A \cap \{T_n \leq k\}} X_{T_n \wedge k} dP \leq \int_{A \cap \{T_n \leq k\}} X_{T_{n+1} \wedge k} dP, \quad A \in \mathcal{F}_{T_n}.$$

But on $\{T_n \leq k\}$, $T_n \wedge k = T_n$; also, $\{T_{n+1} \leq k\} \subset \{T_n \leq k\}$ and $X_{T_{n+1} \wedge k} \leq 0$; hence

$$\int_{A \cap \{T_n \leq k\}} X_{T_n} dP \leq \int_{A \cap \{T_{n+1} \leq k\}} X_{T_{n+1}} dP.$$

Let $k \rightarrow \infty$ to obtain the desired submartingale property.

Case 2: $X_n = E(X_\infty | \mathcal{F}_n)$, $n = 1, 2, \dots$; thus we have a uniformly integrable martingale by 7.6.6. In this case, the $X_{T_n \wedge k}$, $k = 1, 2, \dots$, are uniformly integrable by 7.6.1, and $X_{T_n \wedge k} \rightarrow X_{T_n}$ as $k \rightarrow \infty$. Thus X_{T_n} is integrable by 7.5.2(b), so that condition (A) of 7.7.3 is satisfied.

Condition (B) follows from the uniform integrability since by 7.5.3,

$$\int_A |X_k| dP \rightarrow 0 \quad \text{as } P(A) \rightarrow 0,$$

uniformly in k , and $P\{T_n > k\} \rightarrow 0$ as $k \rightarrow \infty$.

General Case: Write $X_n = X_n' + X_n''$, where $X_n' = X_n - E(X_\infty | \mathcal{F}_n)$, $X_n'' = E(X_\infty | \mathcal{F}_n)$. The X_n' fall into case 1 and the X_n'' into case 2, and the result follows. Note that if $\{X_n\}$ is a martingale with last element X_∞ , we must have $X_n = E(X_\infty | \mathcal{F}_n)$ so that $\{Y_n\}$ is a martingale by the analysis of case 2. ■

The case of a finite sequence occurs often enough to be stated separately.

7.7.6 Theorem. Let $\{X_1, \dots, X_n\}$ be a submartingale, and let T_1, T_2, \dots be an increasing sequence of stopping times for $\{X_i, 1 \leq i \leq n\}$. [In other words, the T_i take values in $\{1, \dots, n\}$ and $\{T_i \leq k\} \in \mathcal{F}(X_1, \dots, X_k)$, $k = 1, \dots, n$. The σ -fields \mathcal{F}_{T_i} are defined as before:

$$\mathcal{F}_{T_i} = \{A \in \mathcal{F} : A \cap \{T_i \leq k\} \in \mathcal{F}(X_1, \dots, X_k), k = 1, \dots, n\}.$$

Then the X_{T_i} form a submartingale relative to the σ -fields \mathcal{F}_{T_i} , a martingale if $\{X_i\}$ is a martingale.

PROOF. Extend the given sequence to an infinite sequence $X_1, \dots, X_n, X_n, X_n, \dots$, and apply 7.7.4(a). ■

To conclude this section we give an example of a situation where the optional sampling theorem does not apply. Consider the problem of fair coin tossing, that is, let Y_1, Y_2, \dots be independent random variables, each taking on values ± 1 with equal probability. If $X_n = Y_1 + \dots + Y_n$, the X_n form a martingale by 7.3.3(a). Now with probability 1, $X_n = 1$ for some n . [This is a standard random walk result; for a proof, see Ash (1970, p. 185).] If T is the time that 1 is reached (the hitting time for $\{1\}$), and $S \equiv 1$, then S and T are (a.e.) finite stopping times, but $\{X_S, X_T\}$ is not a martingale. For if this were the case, we would have $E(X_T) = E(X_S) = E(X_1) = 0$. But $X_T \equiv 1$; hence $E(X_T) = 1$, a contradiction. In addition, we obtain from 7.7.5 the result that the X_n are not uniformly integrable.

Problems

- Let $\{X_1, \dots, X_n\}$ be a submartingale, and let T be a stopping time for $\{X_i, 1 \leq i \leq n\}$. Show that

$$E(|X_T|) \leq 2E(X_n^+) - E(X_1).$$

The corresponding result for supermartingales, which may be obtained by replacing X_i by $-X_i$, is

$$E(|X_T|) \leq 2E(X_n^-) + E(X_1).$$

- Let $\{X_1, X_2, \dots\}$ be a submartingale, and T a finite stopping time for $\{X_n\}$. Show that

$$E(|X_T|) \leq 2 \sup_n E(X_n^+) - E(X_1).$$

As in Problem 1, the analogous result for supermartingales is

$$E(|X_T|) \leq 2 \sup_n E(X_n^-) + E(X_1).$$

3. (*Sub- and supermartingale inequalities*) (a) Let $\{X_1, \dots, X_n\}$ be a submartingale. If $\lambda \geq 0$, show that

$$\lambda P\left\{\max_{1 \leq i \leq n} X_i \geq \lambda\right\} \leq \int_{\left\{\max_{1 \leq i \leq n} X_i \geq \lambda\right\}} X_n dP \leq E(X_n^+).$$

- (b) Let $\{X_1, \dots, X_n\}$ be a supermartingale. If $\lambda \geq 0$, show that

$$\begin{aligned} \lambda P\left\{\max_{1 \leq i \leq n} X_i \geq \lambda\right\} &\leq E(X_1) - \int_{\left\{\max_{1 \leq i \leq n} X_i < \lambda\right\}} X_n dP \\ &\leq E(X_1) + E(X_n^-). \end{aligned}$$

(Apply 7.7.6, with $T = \min\{i: X_i \geq \lambda\}$; $T = n$ if all $X_i < \lambda$.)

- (c) If $\{X_1, X_2, \dots\}$ is a submartingale and $\lambda \geq 0$, show that

$$\lambda P\{\sup_n X_n > \lambda\} \leq \sup_n E(X_n^+);$$

if $\{X_1, X_2, \dots\}$ is a supermartingale and $\lambda \geq 0$, show that

$$\lambda P\{\sup_n X_n > \lambda\} \leq E(X_1) + \sup_n E(X_n^-).$$

4. Use Problem 3 to give an alternative proof of Kolmogorov's inequality 7.1.4.
5. (*Wald's theorem on the sum of a random number of random variables*) Let Y_1, Y_2, \dots be independent, identically distributed random variables with finite mean m , and let $X_n = \sum_{k=1}^n Y_k$. If T is a finite stopping time for $\{X_n\}$, establish the following:
- (a) If all $Y_j \geq 0$, then $E(X_T) = mE(T)$.
 - (b) If $E(T) < \infty$, then $E(|X_T|) < \infty$ and $E(X_T) = mE(T)$.
- [Let $T_n = T \wedge n$ and apply 7.7.4(a) to $\{X_n - nm\}$ to prove (a); use (a) to prove (b).]
- (c) If T is a positive integer-valued random variable that is independent of (Y_1, Y_2, \dots) , but not necessarily a stopping time, show that the results (a) and (b) still hold.
6. [*Alternative proof of the upcrossing theorem* (Meyer, 1966)] Let

$$\{X_k, \mathcal{F}_k, k = 1, \dots, n\}$$

be a nonnegative submartingale, and U the number of upcrossings of $(0, b)$ by X_1, \dots, X_n . Define the stopping times T_i as in 7.4.2, and for convenience set $X_\infty = X_n$.

- (a) Show that $X_n = \sum_{k=1}^n (X_{T_k} - X_{T_{k-1}})$ (take $X_{T_0} = 0$).
 (b) Show that $E(X_n) \geq bE(U)$; the general upcrossing theorem is then obtained just as in 7.4.2. $\left[\text{Since } E(X_{T_k} - X_{T_{k-1}}) \geq 0 \text{ for all } k, \text{ and } X_{T_k} - X_{T_{k-1}} \geq b \text{ if } k \text{ is even and } T_k < \infty, \right]$

$$E(X_n) \geq \sum_{\substack{k=1 \\ k \text{ even}}}^n E(X_{T_k} - X_{T_{k-1}}) \geq bE(U).$$

7.8 Applications of Martingale Theory

Martingale ideas provide fresh insights and simplifications for many problems in probability; in this section we consider some important examples. First, we use the martingale convergence theorem to provide a short proof of the strong law of large numbers for iid random variables (see 7.2.5). We need two preliminary facts.

7.8.1 Lemma. If X_1, \dots, X_n are independent, identically distributed random variables with finite expectation, and $S_n = \sum_{k=1}^n X_k$, then

$$E(X_k | S_n) = \frac{S_n}{n} \quad \text{a.e.,} \quad k = 1, \dots, n.$$

Intuitively, given $S_n = X_1 + \dots + X_n$, the average contribution of each X_k is the same, and hence must be S_n/n .

PROOF. If $B \in \mathcal{B}(R)$, then

$$\begin{aligned} \int_{\{S_n \in B\}} X_k dP &= E[X_k I_{\{S_n \in B\}}] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k I_B(x_1 + \dots + x_n) dF(x_1) \cdots dF(x_n), \end{aligned}$$

where F is the distribution function of the X_i . By Fubini's theorem, this is independent of k ; hence

$$\int_{\{S_n \in B\}} X_k dP = \frac{1}{n} \int_{\{S_n \in B\}} \sum_{k=1}^n X_k dP = \int_{\{S_n \in B\}} \frac{S_n}{n} dP. \quad \blacksquare$$

7.8.2 Lemma. If X_1, X_2, \dots are random variables and $S_n = \sum_{k=1}^n X_k$, then $\mathcal{F}(S_n, S_{n+1}, S_{n+2}, \dots) = \mathcal{F}(S_n, X_{n+1}, X_{n+2}, \dots)$.

PROOF. Since $X_{n+k} = S_{n+k} - S_{n+k-1}$, $S_n, X_{n+1}, X_{n+2}, \dots$ are each $\mathcal{F}(S_n, S_{n+1}, S_{n+2}, \dots)$ -measurable; therefore $\mathcal{F}(S_n, X_{n+1}, X_{n+2}, \dots) \subset \mathcal{F}(S_n, S_{n+1}, S_{n+2}, \dots)$. Similarly, $S_n, S_{n+1}, S_{n+2}, \dots$ are each $\mathcal{F}(S_n, X_{n+1}, X_{n+2}, \dots)$ -measurable; hence $\mathcal{F}(S_n, S_{n+1}, S_{n+2}, \dots) \subset \mathcal{F}(S_n, X_{n+1}, X_{n+2}, \dots)$. ■

7.8.3 Strong Law of Large Numbers, iid Case. If X_1, X_2, \dots are iid random variables with finite expectation m , and $S_n = X_1 + \dots + X_n$, then $S_n/n \rightarrow m$ a.e. and in L^1 .

PROOF. (X_1, \dots, X_n) and $(X_{n+1}, X_{n+2}, \dots)$ are independent [Problem 1(a), Section 5.11]; hence (X_1, S_n) and $(X_{n+1}, X_{n+2}, \dots)$, being functions of independent random objects, are independent by 5.8.2(d). Therefore

$$\begin{aligned} E(X_1 | S_n) &= E(X_1 | S_n, X_{n+1}, X_{n+2}, \dots) && \text{by Problem 2, Section 6.5} \\ &= E(X_1 | S_n, S_{n+1}, S_{n+2}, \dots) && \text{by 7.8.2.} \end{aligned}$$

Thus by 7.8.1,

$$E(X_1 | S_n, S_{n+1}, \dots) = \frac{S_n}{n} \quad \text{a.e.}$$

But by 7.6.3 and 7.6.4, $E(X_1 | S_n, S_{n+1}, \dots) \rightarrow E(X_1 | \mathcal{G}_\infty)$ a.e. and in L^1 , where \mathcal{G}_∞ is the tail σ -field of the S_n . Thus S_n/n converges a.e. and in L^1 to a finite limit.

Now to show that the limit is in fact m , we may proceed in two ways. One approach is to note that $\lim_{n \rightarrow \infty} (S_n/n)$ is a tail function of the X_n , and hence is a.e. constant by the Kolmogorov zero–one law 7.2.7. Since S_n/n is L^1 -convergent and $E(S_n/n) \equiv m$, the constant must be m .

Alternatively, we may use the Hewitt–Savage zero–one law 7.2.9 to show that each set in \mathcal{G}_∞ has probability 0 or 1. For if $A \in \mathcal{G}_\infty$ and T permutes n coordinates, then $A \in \mathcal{F}(S_n, S_{n+1}, \dots)$; hence A is of the form $\{(S_n, S_{n+1}, \dots) \in A'\}$ for some $A' \in [\mathcal{B}(R)]^\infty$. Since $S_k = X_1 + \dots + X_k = X_{T(1)} + \dots + X_{T(k)} = S_{T(k)}$, $k \geq n$, A is symmetric, and therefore the Hewitt–Savage zero–one law is applicable. Thus \mathcal{G}_∞ is trivial, and it follows that $E(X_1 | \mathcal{G}_\infty) = E(X_1)$ a.e. by 6.5.7(a). ■

Notice that we have obtained L^1 convergence in the strong law of large numbers; this would be more cumbersome to derive using the classical approach of Section 7.2.

We now consider the general problem of convergence of series of independent random variables. The following variation of the martingale convergence

theorem will be proved first. The technique of the proof rather than the result itself will be used in the development, but the theorem does have applications to series of *dependent* random variables. (For example, see Problem 5.)

7.8.4 Theorem. Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a submartingale, and let $Z = \sup_n (X_n - X_{n-1})$ (define $X_0 = 0$). If $E(Z) < \infty$ (for example, if $X_n - X_{n-1}$ is less than a constant for all $n \geq 2$), then X_n converges to a finite limit a.e. on the set $\{\sup_n X_n < \infty\}$. If $\{X_n, \mathcal{F}_n\}$ is a supermartingale and $E[\inf_n (X_n - X_{n-1})] > -\infty$, then X_n converges to a finite limit a.e. on $\{\inf_n X_n > -\infty\}$.

PROOF. Fix $M > 0$, and let $T = \inf\{n: X_n > M\}$; set $T = \infty$ if there is no such n . Define $T_n = T \wedge n$; if $Y_n = X_{T_n}$, $n = 1, 2, \dots$, then $\{Y_n\}$ is a submartingale by 7.7.4(a). (This is sometimes called the *optional stopping theorem* since $Y_n = X_n$ if $n < T$, $Y_n = X_T$ if $n \geq T$; thus $\{Y_n\}$ is the original process stopped at time T .)

Now if $n < T$, then $Y_n = X_n = X_{n-1} + (X_n - X_{n-1}) \leq M + Z$, and if $n \geq T$, then $Y_n = X_{T-1} + (X_T - X_{T-1}) \leq M + Z$. Thus $Y_n \leq M + Z$ in any case, so $\sup_n E(Y_n^+) \leq M + E(Z^+) < \infty$ by hypothesis. By 7.4.3, Y_n converges a.e. to a finite limit. But if $T = \infty$, then $Y_n \equiv X_n$; hence X_n converges a.e. on $\{\sup_n X_n \leq M\}$. Since M is arbitrary, X_n converges a.e. on $\{\sup_n X_n < \infty\}$. The last statement of the theorem is proved by applying the above argument to $-X_n$. ■

We have seen in 7.2.1 that if Y_1, Y_2, \dots are independent random variables with 0 mean, and $\sum_{k=1}^{\infty} E(Y_k^2) < \infty$, then $\sum_{k=1}^{\infty} Y_k$ converges a.e. There is a partial converse to this result, which we prove after one preliminary.

7.8.5 Theorem. If $\{X_1, X_2, \dots\}$ is a martingale and $E(X_n^2) < \infty$ for all n , then the martingale differences $X_1, X_2 - X_1, \dots, X_n - X_{n-1}, \dots$ are orthogonal.

PROOF. If $j < k$, and $\mathcal{F}_j = \mathcal{F}(X_1, \dots, X_j)$,

$$\begin{aligned} E[(X_j - X_{j-1})(X_k - X_{k-1})] &= E[E((X_j - X_{j-1})(X_k - X_{k-1}) | \mathcal{F}_j)] \\ &= E[(X_j - X_{j-1})E(X_k - X_{k-1} | \mathcal{F}_j)] \end{aligned}$$

since $X_j - X_{j-1}$ is \mathcal{F}_j -measurable. But $E(X_k - X_{k-1} | \mathcal{F}_j) = X_j - X_j = 0$ by the martingale property, and the result follows. ■

7.8.6 Theorem. Let Y_1, Y_2, \dots be independent random variables with 0 mean, and assume $E[\sup_k Y_k^2] < \infty$. (For example, this holds if the Y_k are

uniformly bounded.) If $\sum_{k=1}^{\infty} Y_k$ converges a.e., then $\sum_{k=1}^{\infty} E(Y_k^2) < \infty$. (As in Section 7.2, “convergence” of a series means convergence to a finite limit.)

PROOF. The $X_n = \sum_{k=1}^n Y_k$ form a martingale. Choose M such that

$$P\left\{\sup_n |X_n| \leq M\right\} > 0;$$

this is possible since the series converges a.e.

Let $T = \inf\{n: |X_n| > M\}$; $T = \infty$ if there is no such n . If $T_n = T \wedge n$, then $\{X_{T_n}\}$ is a martingale, and just as in 7.8.4, $|X_{T_n}| \leq M + \sup_j |X_j - X_{j-1}| = M + Z$, where $E(Z^2) < \infty$ by hypothesis. It follows that the numbers $E(X_{T_n}^2)$ are uniformly bounded, so by 7.6.10, X_{T_n} converges a.e. and in L^2 . But by 7.8.5, $E(X_{T_n}^2) = \sum_{j=1}^n E[(X_{T_j} - X_{T_{j-1}})^2]$ (take $X_{T_0} = 0$). Since X_{T_n} is L^2 -convergent, $E(X_{T_n}^2)$ approaches a finite limit; hence $\sum_{n=1}^{\infty} E[(X_{T_n} - X_{T_{n-1}})^2] < \infty$.

But $X_{T_n} - X_{T_{n-1}} = Y_n I_{\{T \geq n\}}$, so $\sum_{n=1}^{\infty} E(Y_n^2 I_{\{T \geq n\}}) < \infty$, and consequently

$$\sum_{n=1}^{\infty} E(Y_n^2 I_{\{T \geq n\}} | Y_1, \dots, Y_{n-1}) < \infty \quad \text{a.e.}$$

[To see this, note that if $Z_n \geq 0$ and $\sum_n E(Z_n) < \infty$, then $E(\sum_n Z_n) < \infty$, so $\sum_n Z_n < \infty$ a.e.; set $Z_n = E(Y_n^2 I_{\{T \geq n\}} | Y_1, \dots, Y_{n-1})$.]

Now $I_{\{T \geq n\}} = I_{\{T \leq n-1\}^c}$, which is $\mathcal{F}(Y_1, \dots, Y_{n-1})$ -measurable; hence

$$\sum_{n=1}^{\infty} I_{\{T \geq n\}} E(Y_n^2 | Y_1, \dots, Y_{n-1}) = \sum_{n=1}^{\infty} I_{\{T \geq n\}} E(Y_n^2) < \infty \quad \text{a.e.}$$

Pick an ω where the series converges and where $\sup_n |X_n(\omega)| \leq M$ (see the beginning of the proof). Then $T(\omega) = \infty \geq n$ for all n ; it follows that $\sum_{n=1}^{\infty} E(Y_n^2) < \infty$. ■

We now have a partial solution to the general problem.

7.8.7 Theorem. Let Y_1, Y_2, \dots be independent random variables with 0 mean; assume $E[\sup_k Y_k^2] < \infty$. Then $\sum_{k=1}^{\infty} Y_k$ converges a.e. if and only if $\sum_{k=1}^{\infty} E(Y_k^2) < \infty$.

PROOF. Apply 7.2.1 and 7.8.6. ■

We can now complete the solution to the random signs problem (see the discussion after 7.2.1). If $X_n = a_n Y_n$, where the Y_n are independent with $P\{Y_n = 1\} = P\{Y_n = -1\} = \frac{1}{2}$, and $\sum_n X_n$ converges a.e., then $a_n \rightarrow 0$, so the X_n are uniformly bounded. By 7.8.7, $\sum_n a_n^2 < \infty$.

Incidentally, there is a martingale proof of 7.2.1. If $X_n = \sum_{k=1}^n Y_k$, where the Y_k are independent with 0 mean and $\sum_{k=1}^{\infty} E(Y_k^2) < \infty$, then

$$E(|X_n|) \leq [E(X_n^2)]^{1/2}$$

by the Cauchy-Schwarz inequality

$$= \left[\sum_{k=1}^n E(Y_k^2) \right]^{1/2}$$

since the Y_k are independent with 0 mean; hence orthogonal.

Thus $\sup_n E(|X_n|) < \infty$; hence X_n converges a.e.

This argument also shows that if $\{X_n\}$ is a martingale and $\sum_n E(X_n - X_{n-1})^2 < \infty$ (so that the $X_n - X_{n-1}$ are orthogonal by 7.8.5, but not necessarily independent), then X_n converges a.e.

We now drop the hypothesis of zero mean.

7.8.8 Theorem. Let Y_1, Y_2, \dots be independent, uniformly bounded random variables. Then $\sum_{k=1}^{\infty} Y_k$ converges a.e. iff $\sum_{k=1}^{\infty} \text{Var } Y_k < \infty$ and $\sum_{k=1}^{\infty} E(Y_k)$ converges.

PROOF. "If": By 7.8.7, $\sum_{k=1}^{\infty} (Y_k - E(Y_k))$ converges a.e.; since $\sum_{k=1}^{\infty} E(Y_k)$ converges, we have $\sum_{k=1}^{\infty} Y_k$ convergent a.e.

"Only if": by *symmetrization*. Let $Y_1, Z_1, Y_2, Z_2, \dots$ be independent, where for each j , Z_j has the same distribution as Y_j . Then $E(Y_j - Z_j) = E(Y_j) - E(Z_j) = 0$, $E[(Y_j - Z_j)^2] = \text{Var}(Y_j - Z_j) = \text{Var } Y_j + \text{Var } Z_j = 2 \text{Var } Y_j$. Since $\sum_j Y_j$ converges a.e., so does $\sum_j Z_j$. [To see this, note that $P\{(Y_1, \dots, Y_n) \in B\} = P\{(Z_1, \dots, Z_n) \in B\}$ for all n and all $B \in \mathcal{B}(R^n)$; hence $P_Y = P_Z$.] It follows that $\sum_j (Y_j - Z_j)$ converges a.e., so by 7.8.7, $\sum_j \text{Var } Y_j < \infty$. Again by 7.8.7, $\sum_j (Y_j - E(Y_j))$ converges a.e.; hence $\sum_j E(Y_j)$ converges. ■

Finally, we obtain a general criterion for convergence of series of independent random variables.

7.8.9 Kolmogorov Three Series Theorem. Let Y_1, Y_2, \dots be independent random variables. If $M > 0$, define

$$Y'_j = \begin{cases} Y_j & \text{if } |Y_j| \leq M, \\ 0 & \text{if } |Y_j| > M. \end{cases}$$

(a) If $\sum_j Y_j$ converges a.e., then for any M the three series $\sum_j P\{Y_j \neq Y'_j\}$, $\sum_j E(Y'_j)$, $\sum_j \text{Var } Y'_j$ all converge.

(b) If for some $M > 0$, the three series converge, then $\sum_j Y_j$ converges a.e.

PROOF. (a) By hypothesis, $Y_j \rightarrow 0$ a.e., so eventually $Y_j = Y'_j$. Thus (a.e.) $Y_j \neq Y'_j$ for only finitely many j , that is, $P(\limsup_j \{Y_j \neq Y'_j\}) = 0$. By the

second Borel–Cantelli lemma, $\sum_j P\{Y_j \neq Y'_j\} < \infty$. The other two series converge by 7.8.8.

(b) By 7.8.8, $\sum_j Y'_j$ converges a.e. Since $\sum_j P\{Y_j \neq Y'_j\} < \infty$, we have almost surely, $Y_j = Y'_j$ eventually; hence $\sum_j Y_j$ converges a.e. ■

7.8.10 Branching Processes. As a final example we analyze in detail the branching process of 7.3.3(d). Recall that $X_0 = 1$, and if $X_n = k$, then $X_{n+1} = \sum_{i=1}^k Y_i$, where Y_1, \dots, Y_k are independent and $P\{Y_i = r\} = p_r$, $r = 0, 1, \dots$. We assume that $m = E(Y_i) = \sum_{r=1}^{\infty} rp_r > 0$. This excludes the degenerate case $p_0 = 1$ (in this case $X_n = 0$ for all $n \geq 1$). We also assume that $p_0 + p_1 < 1$. (If $p_0 + p_1 = 1$, then $X_n \leq 1$ for all n . If $p_0 > 0$, then X_n is eventually 0 since $P\{X_n = 1\text{ for all }n\} = \lim_{n \rightarrow \infty} p_1^n = 0$, and if $p_0 = 0$, then $X_n \equiv 1$.)

Case 1: $m < 1$. In this case, almost surely, X_n is 0 eventually; thus *the family name is extinguished with probability 1*.

For $E(X_{n+1} | X_n = k) = kE(Y_1) = km$; hence $E(X_{n+1} | X_n) = mX_n$. It follows that $E(X_{n+1}) = mE(X_n)$. If $m < 1$, then

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n) < \infty,$$

so $X_n \rightarrow 0$ a.e. But the X_n are integer-valued, and thus with probability 1, X_n is ultimately 0.

Case 2: $m > 1$. We show that with probability r , X_n is eventually 0, and with probability $1 - r$, $X_n \rightarrow \infty$, where r is the unique root in $[0, 1)$ of the equation $\sum_{j=0}^{\infty} p_j s^j = s$.

Let $g(s) = \sum_{j=0}^{\infty} p_j s^j$, $0 \leq s \leq 1$. Consider $d[g(s) - s]/ds = g'(s) - 1$. We have $g'(0) - 1 = p_1 - 1 < 0$, $g'(1) - 1 = m - 1 > 0$, and since $p_0 + p_1 < 1$, $g'(s) - 1$ is strictly increasing. Since $g(s) - s = p_0$ when $s = 0$ and 0 when $s = 1$, $g(s) - s$ strictly decreases to a minimum occurring somewhere in $(0, 1)$, and then strictly increases to 0 at $s = 1$. It follows that $g(s) = s$ for exactly one $s \in [0, 1]$, say at $s = r$ (Fig. 7.2).

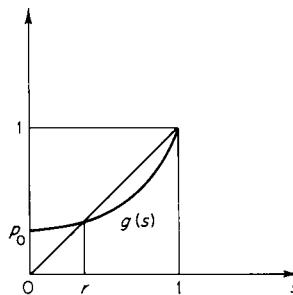


Figure 7.2.

First assume $0 < r < 1$ (hence $p_0 > 0$). By 7.3.3(e), $\{r^{X_n}\}$ is a nonnegative martingale, and hence converges a.e. Since X_n is nonnegative integer-valued, this means that for almost every ω , $X_n(\omega)$ becomes constant (the constant depending on ω) or $X_n(\omega) \rightarrow \infty$.

Now $P\{X_n \text{ eventually constant}\} = \sum_{k=0}^{\infty} P\{X_n = k \text{ for sufficiently large } n\}$. If $k \geq 1$ and $P\{X_n = k \text{ eventually}\} > 0$, then $P\{X_n = k \text{ for all } n \geq N\} > 0$ for some N . But by the Markov property, this probability is $P\{X_N = k\} \cdot \lim_{j \rightarrow \infty} q^j$, where $q = P\{X_{n+1} = k \mid X_n = k\}$. Now $q < 1$ since $P\{X_{n+1} = 0 \mid X_n = k\} = p_0^k > 0$. Thus $\lim_{j \rightarrow \infty} q^j = 0$, a contradiction.

Therefore $X_n \rightarrow X_\infty$ a.e., where $X_\infty = 0$ or ∞ . Since $\{r^{X_n}, n = 0, 1, \dots\}$ is bounded, the dominated convergence theorem gives $E(r^{X_n}) \rightarrow E(r^{X_\infty}) = 1P\{X_\infty = 0\} + 0P\{X_\infty = \infty\} = P\{X_\infty = 0\}$. But by the martingale property, $E(r^{X_n}) = E(r^{X_0}) = r$. Thus with probability r , X_n is eventually 0, and with probability $1 - r$, $X_n \rightarrow \infty$.

If $r = 0$, then $p_0 = 0$; hence $X_{n+1} \geq X_n \geq 1$, so that X_n increases to a limit X . But since the X_n are positive integer-valued,

$$\begin{aligned} P\{X < \infty\} &= P\{X_n \text{ eventually constant}\} \\ &= \sum_{k=1}^{\infty} P\{X_n = k \text{ eventually}\} \\ &= 0 \quad \text{by the same argument as in the case } 0 < r < 1. \end{aligned}$$

(In the present situation, $q = P\{X_{n+1} = k \mid X_n = k\} = P\{Y_1 = \dots = Y_k = 1\} = p_1^k < 1$.) Thus $X_n \rightarrow \infty$ a.e., as desired.

Case 3: $m = 1$. Here, extinction occurs with probability 1. For by 7.3.3(d), $\{X_n\}$ is a nonnegative martingale, and hence converges a.e. to a finite limit. The analysis of case 2 shows that X_n cannot approach a nonzero constant on a set of positive probability, and hence $X_n \rightarrow 0$ a.e.; in other words, with probability 1, X_n is eventually 0.

Problems

- Let X_1, X_2, \dots be arbitrary random variables, with $S_n = X_1 + \dots + X_n$. What is the relation between the tail σ -field of the X_n and the tail σ -field of the S_n ?
- What happens in the branching process (see 7.8.10) if instead of $X_0 = 1$, X_0 is an arbitrary, positive integer-valued random variable?
- (Breiman's realistic gambling model) Let X_n be a gambler's capital after n plays of a game of chance. Assume that the gambler has the option of betting or passing at each trial. If he passes at trial $n + 1$, then $X_{n+1} = X_n$, and if he bets, then we assume that $|X_{n+1} - X_n| \geq b \geq 0$; thus there is a

minimum amount b that can be won or lost on a given trial. We do not spell out the gambler's strategy in detail; we simply assume that his strategy, together with the house rules, determines the distribution of (X_0, X_1, \dots) . It is reasonable to assume that the game is unfavorable or at best fair; thus we take $\{X_n\}$ as a nonnegative supermartingale.

Let T be the time of the last bet, that is, the largest n such that $|X_{n+1} - X_n| \geq b$ ($T = \infty$ if there is no such n). Note that T is *not* a stopping time.

- (a) Show that T is a.e. finite.
 - (b) Show that $E(X_T) \leq E(X_0)$, so no system can increase the expected winning.
 - (c) If the gambler's strategy is always to bet so long as his capital is at least b , show that (a.e.) X_n will eventually be less than b , in other words, *the persistent gambler goes broke with probability 1*.
 - (d) If an unbiased coin is tossed independently over and over again, and we win a dollar for each head and lose a dollar for each tail, our accumulated capital must eventually reach +1 (see the end of Section 7.7). Suppose our strategy is simply to wait until we reach 1 and then quit, thus guaranteeing a profit. Why is this not realistic?
4. Let Y_1, Y_2, \dots be integrable random variables, and $\mathcal{F}_1, \mathcal{F}_2, \dots$ an increasing sequence of sub σ -fields of \mathcal{F} . Assume that Y_k is \mathcal{F}_k -measurable for each k , and define

$$X_n = \sum_{k=1}^n [Y_k - E(Y_k | \mathcal{F}_{k-1})]$$

[take $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so that $E(Y_1 | \mathcal{F}_0) = E(Y_1)$]. Show that $\{X_n, \mathcal{F}_n\}$ is a martingale.

5. (Lévy's extension of the Borel–Cantelli lemma) Let $\{\mathcal{F}_n\}$ be an increasing sequence of σ -fields, and let $A_n \in \mathcal{F}_n$, $n = 1, 2, \dots$. Define

$$q_j = P(A_j | \mathcal{F}_{j-1}), \quad j = 1, 2, \dots$$

[take $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so that $q_1 = P(A_1)$]. Show that (a.e.) infinitely many A_j occur iff $\sum_j q_j = \infty$, that is, $P[(\limsup_j A_j) \Delta \{\sum_j q_j = \infty\}] = 0$. Equivalently, $\sum_j I_{A_j}$ and $\sum_j q_j$ have essentially the same convergence set. [Apply 7.8.4 to $X_n = \sum_{j=1}^n (I_{A_j} - q_j)$.]

7.9 Applications to Markov Chains

In this section we apply martingale theory to the problem of classifying the states of a Markov chain. We must use a few basic properties of Markov chains: the reader who is unfamiliar with this subject may consult Ash (1970,

Chapter 7). In particular, a state i is said to be *recurrent* iff starting at i there will be a return to i with probability 1; otherwise the state is *transient*. If C is a set of states such that every state in C can be reached (in a finite number of steps) from every other state, then all states in C are of the same type, recurrent or transient.

We have the following criterion.

7.9.1 Theorem. Let $[p_{ij}]$ be the transition matrix of a Markov chain such that every state in the state space S is reachable from every other state (sometimes called an irreducible chain). Choose a fixed state, and label it 0 for convenience. The states are transient iff there is a nonconstant bounded $f: S \rightarrow R$ such that $\sum_{j \in S} p_{ij}f(j) = f(i)$ for all $i \neq 0$.

PROOF. Suppose such an f exists. By adding a constant to f we may assume that $f \geq 0$. Assume the initial state is $i \neq 0$, and let $\{X_n\}$ be the corresponding sequence of random variables. Let T be the time at which 0 is reached, and let $Y_n = X_{T \wedge n}$, $n = 0, 1, \dots$; $\{Y_n\}$ can be realized as a Markov chain with the same initial distribution and transition matrix as $\{X_n\}$, except that 0 is now an absorbing state. In other words, the transition matrix for $\{Y_n\}$ is

$$\begin{aligned}\hat{p}_{ij} &= p_{ij} && \text{for all } j \text{ if } i \neq 0, \\ \hat{p}_{00} &= 1.\end{aligned}$$

Thus $\sum_{j \in S} \hat{p}_{ij}f(j) = f(i)$ for all $i \in S$. In matrix form, $\hat{\Pi}f = f$; by induction, $\hat{\Pi}^n f = f$, that is,

$$\sum_{j \in S} \hat{p}_{ij}^{(n)} f(j) = f(i),$$

where $\hat{p}_{ij}^{(n)} = P\{Y_n = j | Y_0 = i\}$. But this says that $E[f(Y_n) | Y_0 = i] = f(i)$.

If the states of the original chain are recurrent, then 0 will be visited with probability 1; hence $Y_n \rightarrow 0$ a.e. By the dominated convergence theorem, $E[f(Y_n) | Y_0 = i] \rightarrow f(0)$. We conclude that $f(i) = f(0)$ for all i , contradicting the hypothesis that f is nonconstant.

Conversely, if the states are transient, we define $f: S \rightarrow R$ as follows. If $i \neq 0$, let $f(i) = f_{i0}$, the probability that, starting from i , 0 will eventually be reached; take $f(0) = 1$. Now in order ultimately to reach 0 from $i \neq 0$, we may either go directly to 0 at step 1, or go to a state $j \neq 0$ and then reach 0 at some time after the first step. It follows that

$$f(i) = \sum_{j \in S} p_{ij}f(j), \quad i \neq 0.$$

(This may be formalized using the Markov property.)

Now f is clearly bounded, and $f_{i0} < 1$ for some $i \neq 0$, otherwise 0 would be a recurrent state. Thus f is nonconstant. ■

Martingale theory enters in deriving the following sufficient condition for recurrence.

7.9.2 Theorem. Let $[p_{ij}]$ be the transition matrix of an irreducible Markov chain whose state space S is the set of nonnegative integers. If there is a function $f: S \rightarrow R$ such that $f(i) \rightarrow \infty$ as $i \rightarrow \infty$, and $\sum_j p_{ij}f(j) \leq f(i)$ for all $i \neq 0$, then the chain is recurrent.

PROOF. Since $f(i) \rightarrow \infty$ as $i \rightarrow \infty$, f is bounded below, so without loss of generality we may assume $f \geq 0$. Let the initial state be $i \neq 0$, and form the process $\{Y_n\}$ as in 7.9.1. Then $\sum_{j \in S} \hat{p}_{ij}f(j) \leq f(i)$ for all i , which implies that $\{f(Y_n)\}$ is a nonnegative supermartingale, and hence converges a.e. to a finite limit. For

$$\begin{aligned} E[f(Y_n) | Y_0 = i_0, \dots, Y_{n-1} = i_{n-1}] &= E[f(Y_n) | Y_{n-1} = i_{n-1}] \\ &\quad \text{by the Markov property} \\ &= \sum_j \hat{p}_{i_{n-1} j} f(j) \\ &\leq f(i_{n-1}). \end{aligned}$$

Note also that

$$E[f(Y_n) | Y_0 = i] = \sum_{j \in S} \hat{p}_{ij}^{(n)} f(j) \leq f(i) < \infty;$$

hence $\{f(Y_n)\}$ is integrable.

Assume the states transient. Then $f_{i0} < 1$ for some $i > 0$. Choose such an i as the initial state. Now $X_n \rightarrow \infty$ a.e. since a finite set of transient states cannot be visited infinitely often [see Ash (1970, p. 223)]. This means that with probability 1, $Y_n \rightarrow 0$ or ∞ . But $P\{Y_n \rightarrow 0\} = f_{i0} < 1$, and hence $P\{Y_n \rightarrow \infty\} > 0$; this implies that $P\{f(Y_n) \rightarrow \infty\} > 0$, a contradiction. ■

The proof of 7.9.2 shows that if $[p_{ij}]$ is the transition matrix of a Markov chain $\{X_n\}$, f is a real-valued function on the state space, and

$$\sum_j p_{ij}f(j) \leq f(i) \quad \text{for all } i,$$

where the series is assumed to converge absolutely, then, with a fixed initial state i , the sequence $\{f(X_n)\}$ is a supermartingale. Similarly, replacement of " \leq " by " $=$ " in this equation yields a martingale, and replacement by " \geq " yields a submartingale.

We now apply 7.9.1 and 7.9.2 to a queueing process.

7.9.3 Example. Assume that customers are to be served at discrete times $t = 0, 1, \dots$, and at most one customer can be served at a given time. Say

there are X_t customers before the completion of service at time t , and in the interval $[t, t + 1)$, Y_t new customers arrive, where $P\{Y_t = k\} = p_k$, $k = 0, 1, \dots$. The number of customers before completion of service at time $t + 1$ is

$$X_{t+1} = (X_t - 1)^+ + Y_t.$$

The queueing process may be represented as a Markov chain whose state space is the set of nonnegative integers and whose transition matrix is

$$\Pi = \begin{bmatrix} 0 & 1 & 2 & 3 & \cdots \\ p_0 & p_1 & p_2 & \cdots & \\ 1 & p_0 & p_1 & p_2 & \cdots \\ 2 & 0 & p_0 & p_1 & p_2 & \cdots \\ 3 & 0 & 0 & p_0 & p_1 & p_2 & \cdots \\ \vdots & & & & & & \end{bmatrix}$$

We assume that $p_0 > 0$ and $p_0 + p_1 < 1$, so that the chain is irreducible. We analyze the behavior of the chain using 7.9.1 and 7.9.2.

The equations $\sum_{j=0}^{\infty} p_{ij}f(j) = f(i)$, $i > 0$, are

$$p_0f(i-1) + p_1f(i) + p_2f(i+1) + \cdots + p_nf(i+n-1) + \cdots = f(i). \quad (1)$$

Let $m = E(Y_t) = \sum_{k=1}^{\infty} kp_k$; we show that the states are transient if $m > 1$, recurrent if $m \leq 1$.

First assume $m > 1$; if $f(i) = r^i$, then (1) becomes

$$p_0r^{i-1} + p_1r^i + p_2r^{i+1} + \cdots + p_nr^{i+n-1} + \cdots = r^i$$

or

$$\sum_{k=0}^{\infty} p_k r^k = r.$$

But this can be satisfied for some $r \in (0, 1)$ (see case 2 of 7.8.10). Thus $\{r^i\}$ is bounded and nonconstant, so by 7.9.1, the states are transient.

Now assume $m \leq 1$, and let $f(i) = i$. Then if $i > 0$,

$$\begin{aligned} \sum_{j=0}^{\infty} p_{ij}f(j) &= p_0(i-1) + p_1i + p_2(i+1) + \cdots \\ &= \sum_{k=i-1}^{\infty} kp_{k-i+1} \\ &= \sum_{k=i-1}^{\infty} (k-i+1)p_{k-i+1} + i-1 \\ &= \sum_{k=0}^{\infty} kp_k + i-1 \leq 1 + i-1 = i = f(i). \end{aligned}$$

By 7.9.2, the states are recurrent.

If i is a recurrent state and μ_i is the average length of time required to return to i when the initial state is i , then i is said to be *recurrent null* if $\mu_i = \infty$, *recurrent positive* if $\mu_i < \infty$. It can be shown that the states are recurrent null if $m = 1$, recurrent positive if $m < 1$ [see Karlin (1966, pp. 74ff.)].

7.10 References

Many classical results on strong laws of large numbers, martingales, and related topics may be found in Loève (1955) and Doob (1953). Another useful reference is Breiman (1968).

Martingales and Markov processes are important in the emerging area of probabilistic potential theory. For an introduction to this subject, see Meyer (1966) and Kemeny, Snell, and Knapp (1966).

8

The Central Limit Theorem

8.1 Introduction

If X_1, X_2, \dots are independent, identically distributed random variables with zero mean, and $S_n = X_1 + \dots + X_n$, the strong law of large numbers states that S_n/n converges a.e. to 0. Thus given $\epsilon > 0$, $|S_n/n|$ will be less than ϵ for large n ; in other words, S_n will eventually be small in comparison with n . The strong law of large numbers gives no information about the distribution of S_n ; the purpose of this chapter is to develop results (called versions of the *central limit theorem*) concerning the approximate distribution of S_n for large n . For example, if the X_n are iid with finite mean m and finite variance σ^2 , then for large n , $(S_n - nm)/\sqrt{n}\sigma$ has, approximately, the normal distribution with mean 0 and variance 1.

There are two basic techniques that will be used. First is the theory of weak convergence. If μ, μ_1, μ_2, \dots are finite measures on $\mathcal{B}(R)$, weak convergence of μ_n to μ means that $\int_R f d\mu_n \rightarrow \int_R f d\mu$ for every bounded continuous $f: R \rightarrow R$. If the corresponding (bounded) distribution functions are F, F_1, F_2, \dots , the equivalent condition is that $F_n(a, b] \rightarrow F(a, b]$ at all continuity points of F . (See Section 4.5 for a discussion of weak convergence; in particular, recall that $+\infty$ and $-\infty$ are, by definition, continuity points of F .) We shall denote weak convergence by $\mu_n \xrightarrow{\text{w}} \mu$ or $F_n \xrightarrow{\text{w}} F$. Also, if B is a Borel subset of R , the terms $F(B)$ and $\mu(B)$ will be synonymous.

Now assume that μ_n is the probability measure induced by a random variable X_n , $n = 0, 1, \dots$ (with $\mu_0 = \mu$, $X_0 = X$). If $\mu_n \xrightarrow{w} \mu$, we say that the sequence $\{X_n\}$ converges in distribution to X , and write $X_n \xrightarrow{d} X$. Since $\int_R f d\mu_n = E[f(X_n)]$, it follows that $X_n \xrightarrow{d} X$ iff $E[f(X_n)] \rightarrow E[f(X)]$ for all bounded continuous $f: R \rightarrow R$.

This in turn implies that if $X_n \xrightarrow{d} X$ and g is a continuous function from R to R , then $g(X_n) \xrightarrow{d} g(X)$. In particular, if $X_n \xrightarrow{d} X$, then $X_n + c \xrightarrow{d} X + c$ for each real number c .

Notice that convergence in distribution is determined completely by the distribution functions, or equivalently by the induced probability measures, of the random variables. In particular, the random variables need not be defined on the same probability space. Note also that since the distribution function of a random variable always has the value 0 at $-\infty$ and the value 1 at $+\infty$, we have $X_n \xrightarrow{d} X$ iff $F_n(x) \rightarrow F(x)$ at all continuity points of F in R .

Now by Theorem 4.5.1, $\mu_n \xrightarrow{w} \mu$ iff $\mu_n(A) \rightarrow \mu(A)$ for each Borel set A whose boundary ∂A has μ -measure 0. Thus $X_n \xrightarrow{d} X$ iff $P\{X_n \in A\} \rightarrow P\{X \in A\}$ for all Borel sets A such that $P\{X \in \partial A\} = 0$. This result justifies the terminology "convergence in distribution," for it says that if $X_n \xrightarrow{d} X$, then X_n and X have approximately the same distribution for large n . Of course it might seem more reasonable to require that $P\{X_n \in A\} \rightarrow P\{X \in A\}$ for all Borel sets A , but actually this is not so. For example, if X_n is uniformly distributed between 0 and $1/n$, that is, X_n has density $f_n(x) = n$, $0 \leq x \leq 1/n$, $f_n(x) = 0$ elsewhere, then for large n , X_n approximates a random variable X that is identically 0. But $P\{X_n = 0\} = 0$ for all n , and $P\{X = 0\} = 1$.

The second technique involves the use of characteristic functions, which we now define.

8.1.1 Definition. Let μ be a finite measure on $\mathcal{B}(R)$. The *characteristic function* of μ is the mapping from R to C given by

$$h(u) = \int_R e^{iux} d\mu(x), \quad u \in R.$$

Thus h is the Fourier transform of μ . If F is a distribution function corresponding to μ , we shall also write $h(u) = \int_R e^{iux} dF(x)$, and call h the characteristic function of F (or of X if X is a random variable with distribution function F).

Characteristic functions are uniquely appropriate in the study of sums of independent random variables, because of the following result.

8.1.2 Theorem. Let X_1, X_2, \dots, X_n be independent random variables, and let $S_n = X_1 + \dots + X_n$. Then the characteristic function of S_n is the product of the characteristic functions of the X_i .

PROOF.

$$\begin{aligned} E(e^{iuS_n}) &= E\left(\prod_{j=1}^n e^{iuX_j}\right) \\ &= \prod_{j=1}^n E(e^{iuX_j}) \quad \text{by independence. } \blacksquare \end{aligned}$$

Theorem 8.1.2 allows us to compute the characteristic function of S_n , knowing only the distribution of the individual X_j 's. In fact, once the characteristic function is known, the distribution function is determined.

8.1.3 Theorem. Let P_1 and P_2 be probability measures (or more generally, finite measures) on $\mathcal{B}(R)$. If $\int_R e^{iux} dP_1(x) = \int_R e^{iux} dP_2(x)$ for all $u \in R$, then $P_1 \equiv P_2$.

PROOF. Let g be a continuous map from R to C , with compact support, that is, for some $T > 0$, $g = 0$ off $[-T, T]$. Since $g(-T) = g(T)$, g may be transferred to the unit circle, in other words, $g(x) = h(\exp i\pi x/T)$, $-T \leq x \leq T$, for some continuous h from $E = \{z \in C : |z| = 1\}$ to C . By the Stone-Weierstrass theorem, there is a sequence of functions of the form

$$h_n(z) = \sum_{k=-m_n}^{m_n} a_k z^k$$

converging uniformly to h . (The algebra of functions $\sum_{k=-m}^m b_k z^k$, $m = 0, 1, \dots$, $z \in E$, separates points, contains the constant functions, and contains the complex conjugate of each of its members since $\bar{z} = 1/z$ on E .) Thus there is a sequence of functions of the form

$$g_n(x) = \sum_{-m_n}^{m_n} a_k \exp\left(\frac{ik\pi x}{T}\right)$$

converging uniformly to g on $[-T, T]$.

Let $M = \max\{|g(x)| : x \in R\} < \infty$. If $\varepsilon > 0$, we may choose T so large that $P_1(R - [-T, T])$ and $P_2(R - [-T, T])$ are less than $\varepsilon/4(M + 1)$ (and $g(x) = 0$ for $|x| \geq T$). Since $g_n \rightarrow g$ uniformly on $[-T, T]$ and the g_n have period $2T$, we may as well assume that $|g_n| \leq M + 1$ for all n .

Let $I = [-T, T]$; by hypothesis,

$$\int_I g_n(x) dP_1(x) + \int_{R-I} g_n(x) dP_1(x) = \int_I g_n(x) dP_2(x) + \int_{R-I} g_n(x) dP_2(x).$$

Let $n \rightarrow \infty$; since

$$\left| \int_{R-I} g_n dP_j \right| \leq (M+1)P_j(R-I) < \frac{\varepsilon}{4},$$

we have $|\int_I g dP_1 - \int_I g dP_2| < \varepsilon/2$. But $|\int_{R-I} g dP_j| < \varepsilon/4$, so $|\int_R g dP_1 - \int_R g dP_2| < \varepsilon$. Since ε is arbitrary, we conclude that $\int_R g dP_1 = \int_R g dP_2$ for all continuous g with compact support.

Now let $(a, b]$ be a right-semiclosed interval. Define $f_n(x) = 0$, $x \leq a$ or $x \geq b + (1/n)$, $f_n(x) = 1$, $a + (1/n) \leq x \leq b$, and interpolate f_n linearly to complete the definition. The f_n are continuous with compact support, and $f_n \rightarrow I_{(a, b]}$. It follows from the dominated convergence theorem that $P_1(a, b] = P_2(a, b]$. Therefore, P_1 and P_2 agree on the field of finite disjoint unions of right-semiclosed intervals; hence $P_1 = P_2$ on all Borel sets. ■

Various procedures involving Fourier or Laplace transforms may be used in actually computing the distribution of a random variable from its characteristic function; see Ash (1970, Chapter 5). There is also an explicit formula for the distribution function in terms of the characteristic function, as follows.

8.1.4 Inversion Formula. If h is the characteristic function of the bounded distribution function F , and $F(a, b] = F(b) - F(a)$, then

$$F(a, b] = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iua} - e^{-iub}}{iu} h(u) du$$

for all points a, b ($a < b$) at which F is continuous. If in addition, h is Lebesgue integrable on $(-\infty, \infty)$, then the function f given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} h(u) du$$

is a density for F , that is, f is nonnegative and $F(x) = \int_{-\infty}^x f(t) dt$ for all x ; furthermore, $F' = f$ everywhere. Thus in this case, f and h are “Fourier transform pairs”:

$$h(u) = \int_{-\infty}^{\infty} e^{iux} f(x) dx,$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} h(u) du.$$

The result is proved in Section 8.7. Note that this gives another proof of Theorem 8.1.3. For the inversion formula shows that h determines F at all

continuity points; but every x is a limit from above of continuity points since F is monotone and thus has only countably many discontinuities. Thus h determines F everywhere.

Characteristic functions have the following basic properties.

8.1.5 Theorem. Let h be the characteristic function of the bounded distribution function F . Then

- (a) $|h(u)| \leq h(0) = F(\infty) - F(\infty)$ for all u ;
- (b) h is continuous on R ;
- (c) $h(-u) = \overline{h(u)}$, the complex conjugate of $h(u)$;
- (d) $h(u)$ is real-valued for all u iff F is symmetric; that is, $\int_B dF(x) = \int_{-B} dF(x)$ for all Borel sets B , where $-B = \{-x: x \in B\}$.
- (e) If $\int_R |x|^r dF(x) < \infty$ for some positive integer r , then the r th derivative of h exists and is continuous on R , and

$$h^{(r)}(u) = \int_R (ix)^r e^{iux} dF(x).$$

PROOF. Part (a) is clear since $|e^{iux}| = 1$, and (b) follows from the dominated convergence theorem (see Problem 1, Section 1.6). Part (c) follows from the fact that the conjugate of e^{iux} is e^{-iux} . To prove (d), assume h to be real-valued and, for now, let F be the distribution function of the random variable X . Now $E(e^{-iux}) = h(-u) = \overline{h(u)} = h(u)$ since $h(u)$ is real, so that $-X$ has characteristic function h . By 8.1.3, X and $-X$ have the same distribution; hence $P\{X \in B\} = P\{X \in -B\}$, and F is therefore symmetric. In general, we may multiply F by a positive constant c such that $c(F(\infty) - F(-\infty)) = 1$. (If $F(\infty) - F(-\infty) = 0$, then $\int_B dF(x) = 0$ for all $B \in \mathcal{B}(R)$ and the result is trivial.) The above argument shows that $\int_B c dF(x) = \int_{-B} c dF(x)$, and hence F is symmetric.

Conversely, assume F symmetric, and let $g: R \rightarrow R$ be a bounded Borel measurable function. If g is odd [$g(-x) = -g(x)$ for all x], then $\int_R g(x) dF(x) = 0$; to prove this, approximate g by odd simple functions. In particular, $\int_R \sin ux dF(x) = 0$, and h is real-valued.

Finally, we prove (e). Since $|(ix)^r e^{iux}| = |x|^r$ and $\int_R |x|^r dF(x) < \infty$, we may differentiate $h(u) = \int_R e^{iux} dF(x)$ r times under the integral sign (see Section 1.6, Problem 3), and the result follows. ■

Now suppose that h is a given function from R to C , and we wish to determine whether or not h is the characteristic function of some bounded distribution function F . There are some practical approaches that often work.

For example, if h fails to satisfy (a), (b), or (c) of 8.1.5, h cannot be a characteristic function. Also, suppose h is continuous and Lebesgue integrable on R , and we compute

$$f(x) = \frac{1}{2\pi} \int_R e^{-iux} h(u) du, \quad x \in R.$$

If f is Lebesgue integrable, then

$$h(u) = \int_R e^{iux} f(x) dx, \quad u \in R.$$

[This is a standard Fourier transform result; see Rudin (1966, p. 186).] Thus if f is everywhere nonnegative, then h is the characteristic function of a measure μ with density f , that is,

$$\mu(B) = \int_B f(x) dx, \quad B \in \mathcal{B}(R).$$

There is a general criterion for deciding whether or not h is a characteristic function. The result is of considerable importance in the development of second order properties of stochastic processes, a subject to be considered in a future volume. However, it is in general not useful when applied to explicit examples.

8.1.6 Bochner's Theorem. If $h: R \rightarrow C$, then h is a characteristic function iff h is continuous at the origin and is *nonnegative definite*, in other words, for all $u_1, \dots, u_n \in R$, $n = 1, 2, \dots$, and all complex numbers a_1, \dots, a_n ,

$$\sum_{j,k=1}^n a_j h(u_j - u_k) \bar{a}_k$$

is real and nonnegative.

PARTIAL PROOF. Assume h is a characteristic function; then h is continuous everywhere by 8.1.5(b). To prove nonnegative definiteness, write

$$\begin{aligned} 0 \leq \int_{-\infty}^{\infty} \left| \sum_{j=1}^n a_j e^{iu_j x} \right|^2 dF(x) &= \sum_{j,k=1}^n \int_{-\infty}^{\infty} a_j e^{iu_j x} \bar{a}_k e^{-iu_k x} dF(x) \\ &= \sum_{j,k=1}^n a_j h(u_j - u_k) \bar{a}_k. \end{aligned}$$

The converse is considerably more difficult, and since the result will not be used in the text, the argument will be omitted. For a complete proof, see Loève (1955). ■

We shall need the result that if X is normally distributed with mean m and variance σ^2 , the characteristic function of X is

$$h(u) = \exp(ium) \exp(-\frac{1}{2}u^2\sigma^2).$$

For the computation, see Ash (1970, p. 163).

Problems

1. Let P_1 and P_2 be probability measures on $\mathcal{B}(R^n)$, and let $\langle u, x \rangle = \sum_{j=1}^n u_j x_j$, where $u = (u_1, \dots, u_n)$, $x = (x_1, \dots, x_n) \in R^n$. If

$$\int_{R^n} \exp(i\langle u, x \rangle) dP_1(x) = \int_{R^n} \exp(i\langle u, x \rangle) dP_2(x) \quad \text{for all } u \in R^n,$$

show that $P_1 \equiv P_2$. (Adjust the proof of Theorem 8.1.3.)

2. The following application of Theorem 8.1.3 is useful in computations involving characteristic functions. Let f and g be nonnegative Borel measurable functions from R to R , and assume that for some fixed real t , $\int_{-\infty}^{\infty} f(x)e^{-tx} dx < \infty$ and $\int_{-\infty}^{\infty} g(x)e^{-tx} dx < \infty$. If

$$\int_{-\infty}^{\infty} f(x)e^{-tx}e^{iux} dx = \int_{-\infty}^{\infty} g(x)e^{-tx}e^{iux} dx \quad \text{for all } u \in R,$$

show that $f = g$ a.e. (Lebesgue measure).

3. If h_1 and h_2 are characteristic functions, show that $h_1 + h_2$ and $\operatorname{Re} h_1$ are also characteristic functions. Is $\operatorname{Im} h_1$ a characteristic function?
4. Let h be the characteristic function of the random variable X .
- If $|h(u)| = 1$ for some $u \neq 0$, show that X has a lattice distribution, that is, with probability 1, X belongs to the set $\{a + nk : n \text{ an integer}\}$ for appropriate a and k ($= 2\pi u^{-1}$). Conversely, if X has a lattice distribution, then $|h(u)| = 1$ for some $u \neq 0$.
 - If $|h(u)| = 1$ at two distinct points u and αu , where α is irrational, show that X is degenerate.
5. Let X be a random variable with $E(|X|^n) < \infty$ for some positive integer n . If h is the characteristic function of X , show that $h^{(k)}(0) = i^k E(X^k)$, $k = 0, \dots, n$, and

$$(a) \quad h(u) = \sum_{k=0}^n \frac{E(X^k)}{k!} (iu)^k + o(u^n),$$

$$(b) \quad \left| h(u) - \sum_{k=0}^{n-1} \frac{E(X^k)}{k!} (iu)^k \right| \leq \frac{|u|^n E(|X|^n)}{n!}.$$

6. (a) If X is a random variable with $E(|X|^r) < \infty$ for all $r > 0$, show that the characteristic function of X is given by

$$h(u) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} (iu)^n$$

within the interval of convergence of the series. This is the *moment-generating property* of characteristic functions.

- (b) Give an example of a random variable X with $E(|X|^r) < \infty$ for all $r > 0$, such that the series

$$\sum_{n=0}^{\infty} \frac{E(X^n)}{n!} (iu)^n$$

converges only at $u = 0$.

7. (a) Let h be the characteristic function of the bounded distribution function F . Define

$$E_r h(u) = h(u + r), \quad r \text{ real.}$$

Show that

$$\left(\frac{E_r - E_{-r}}{2r} \right)^2 h(0) = - \int_{-\infty}^{\infty} \left(\frac{\sin rx}{rx} \right)^2 x^2 dF(x).$$

$$[(E_r - E_{-r})h(0) = h(r) - h(-r); (E_r - E_{-r})^2 h(0) \text{ means}]$$

$$(E_r - E_{-r})(h(r) - h(-r)) = h(2r) - 2h(0) + h(-2r).]$$

- (b) If h'' exists and is finite at the origin, show that $\int_{-\infty}^{\infty} x^2 dF(x) < \infty$. (Use l'Hospital's rule and Fatou's lemma.)
(c) If $h^{(2n)}(0)$ exists and is finite, show that $\int_{-\infty}^{\infty} x^{2n} dF(x) < \infty$ ($n = 1, 2, \dots$).

[It is probably easier to use part (b) and an induction argument rather than to extend part (a).]

8. If X is a random variable, let $N(s) = E(e^{-sx})$, s complex; when $s = -iu$, we obtain the characteristic function of X . If N is analytic at the origin, show that $E(X^k)$ is finite for all $k > 0$, and

$$N(s) = \sum_{k=0}^{\infty} \frac{(-1)^k E(X^k)}{k!} s^k$$

within the circle of convergence of the series. In particular, $N^{(k)}(0) = (-1)^k E(X^k)$.

8.2 The Fundamental Weak Compactness Theorem

The basic connection between weak convergence and characteristic functions is essentially this. Let $\{F_n\}$ be a bounded sequence of distribution functions on R ("bounded" means that for some positive M , $F_n(\infty) - F_n(-\infty) \leq M$ for all n). Let $\{h_n\}$ be the corresponding sequence of characteristic functions. If F is a bounded distribution function with characteristic function h , then weak convergence of F_n to F is equivalent to pointwise convergence of h_n to h . In the course of developing this result, we must consider the following question. If $\{F_n\}$ is a bounded sequence of distribution functions, when will there exist a weakly convergent subsequence? Now any bounded sequence of real numbers has a convergent subsequence, so one might conjecture that any bounded sequence of distribution functions has a weakly convergent subsequence. In fact this is not true, but the following result comes close, in a sense.

8.2.1 Helly's Theorem. Let F_1, F_2, \dots be distribution functions on R . Assume that $F_n(-\infty) = 0$ for all n , and $F_n(\infty) \leq M < \infty$ for all n . Then there is a distribution function F and a subsequence $\{F_{n_k}\}$ such that $F_{n_k}(x) \rightarrow F(x)$ for each $x \in R$ at which F is continuous.

PROOF. Let $D = \{x_1, x_2, \dots\}$ be a countable dense subset of R . Since the sequence $\{F_n(x_1)\}$ is bounded, we can extract a subsequence $\{F_{1j}\}$ of $\{F_n\}$ with $F_{1j}(x_1)$ converging to a limit y_1 as $j \rightarrow \infty$. Since $\{F_{1j}(x_2)\}$ is bounded, there is a subsequence $\{F_{2j}\}$ of $\{F_{1j}\}$ such that $F_{2j}(x_2)$ approaches a limit y_2 . Continuing inductively, we find subsequences $\{F_{mj}\}$ of $\{F_{m-1,j}\}$ with $F_{mj}(x_m) \rightarrow y_m$, $m = 1, 2, \dots$ (of course all $|y_m|$ are bounded by M).

Define $F_D : D \rightarrow R$ by $F_D(x_j) = y_j$, $j = 1, 2, \dots$, and let $F_{n_k} = F_{kk}$, $k = 1, 2, \dots$ (the "diagonal sequence"). Then $F_{n_k}(x) \rightarrow F_D(x)$, $x \in D$. Since F_{n_k} is one of the original F_n , $x < y$ implies $F_{n_k}(x) \leq F_{n_k}(y)$; hence $F_D(x) \leq F_D(y)$. Define

$$F(x) = \inf\{F_D(y) : y \in D, y > x\}.$$

By definition, F is increasing. To prove that F is right-continuous, let $z_n \downarrow x$; then $F(z_n)$ approaches a limit $b \geq F(x)$. If $F(x) < b$, let $y_0 \in D$, $y_0 > x$, with $F_D(y_0) < b$. For large n we have $x < z_n < y_0$, so $F(z_n) \leq F_D(y_0) < b$. Thus $\lim F(z_n) < b$, a contradiction, and therefore $F(z_n) \rightarrow F(x)$.

Now we show that $F_{n_k}(x) \rightarrow F(x)$ at continuity points of F in R . If $x < y \in D$, we have $\limsup_{k \rightarrow \infty} F_{n_k}(x) \leq \limsup_{k \rightarrow \infty} F_{n_k}(y) = F_D(y)$; take the inf over $y \in D$ to obtain $\limsup_{k \rightarrow \infty} F_{n_k}(x) \leq F(x)$.

If $x' < y < x$, $y \in D$, we have $F(x') \leq F_D(y) = \lim_{k \rightarrow \infty} F_{n_k}(y) = \liminf_{k \rightarrow \infty} F_{n_k}(y) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x)$. Let $x' \rightarrow x$ to obtain $F(x^-) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x)$. Thus if $F(x) = F(x^-)$, we have $F_{n_k}(x) \rightarrow F(x)$. ■

8.2.2 Comments. If instead of assuming that $F_n(-\infty) = 0$ and $F_n(\infty) \leq M$ for all n , we assume that $F_n(\infty) - F_n(-\infty) \leq M < \infty$ for all n , 8.2.1 implies that there is a distribution function F and a subsequence $\{F_{n_k}\}$ with $F_{n_k}(a, b] \rightarrow F(a, b]$ for all $a, b \in R$ at which F is continuous. [To see this, consider $G_n(x) = F_n(x) - F_n(-\infty)$.]

If F_1, F_2, \dots are distribution functions on R^k (assumed monotone in each coordinate and 0 at $(-\infty, \dots, -\infty)$), and $F_n(R^k) \leq M < \infty$ for all n , there is a distribution function F and a subsequence $\{F_{n_k}\}$ converging to F at continuity points of F in R^k . The proof is essentially the same as above.

There is no difficulty in constructing a bounded sequence of distribution functions with no weakly convergent subsequence. For example, let F_n be the distribution function of a random variable that is identically n ($F_n(x) = 1$, $x \geq n$; $F_n(x) = 0$, $x < n$). If F_{n_k} converges weakly to F , then $F(a, b]$ must be 0 for all $a, b \in R$, $a < b$; hence $F(\infty) - F(-\infty) = 0$. But $F_n(\infty) - F_n(-\infty) \equiv 1$, a contradiction.

A bounded sequence of distribution functions will always have a weakly convergent subsequence unless, as in the above example, too much mass escapes to infinity. We formalize this idea as follows.

8.2.3 Definition. Let $\mathcal{M} = \{\mu_i, i \in I\}$ be a family of finite measures on the Borel sets of a metric space Ω . We say that \mathcal{M} is *tight* iff for each $\varepsilon > 0$, there is a compact set $K \subset \Omega$ such that $\mu_i(\Omega - K) < \varepsilon$ for all i . (If $\Omega = R^k$, the compact set can be replaced by an interval.) We say that \mathcal{M} is *relatively compact* iff each sequence in \mathcal{M} has a subsequence converging weakly to a finite measure on $\mathcal{B}(\Omega)$.

If $\{F_i, i \in I\}$ is a family of distribution functions, tightness of $\{F_i\}$ means tightness of the associated family of measures, and similarly for relative compactness.

We may now prove the basic result.

8.2.4 Prokhorov's Theorem. Let $\mathcal{M} = \{F_i, i \in I\}$ be a family of distribution functions on R , and assume that $F_i(\infty) - F_i(-\infty) \leq M < \infty$ for all i . Then the family is tight iff it is relatively compact.

PROOF. Assume \mathcal{M} is tight, and let F_1, F_2, \dots be a sequence from \mathcal{M} . By 8.2.1 and 8.2.2, there is a distribution function F and a subsequence $\{F_{n_k}\}$ such that $F_{n_k}(a, b] \rightarrow F(a, b]$ for all $a, b \in R$ at which F is continuous. Given

$\varepsilon > 0$, let a and b be finite continuity points of F such that $F_n(R - (a, b]) < \varepsilon$ for all n , and $F(R - (a, b]) < \varepsilon$. If $x \in R$ and x is a continuity point of F , then

$$F_n(\infty) - F_n(x) = F_n(\infty) - F_n(b) + F_n(b) - F_n(x).$$

But $F_{n_k}(b) - F_{n_k}(x) \rightarrow F(b) - F(x)$, $F_n(\infty) - F_n(b) < \varepsilon$ for all n , and $F(\infty) - F(b) < \varepsilon$. It follows that for sufficiently large k , $F_{n_k}(\infty) - F_{n_k}(x)$ differs from $F(\infty) - F(x)$ by less than 2ε . Therefore $F_{n_k}(\infty) - F_{n_k}(x) \rightarrow F(\infty) - F(x)$, and similarly $F_{n_k}(x) - F_{n_k}(-\infty) \rightarrow F(x) - F(-\infty)$. Thus $F_{n_k} \xrightarrow{w} F$, proving relative compactness.

Now if \mathcal{M} is relatively compact but not tight, then for some $\varepsilon > 0$ we have, for each positive integer n , an $F_n \in \mathcal{M}$ with $F_n(R - (-n, n)) \geq \varepsilon$. If $\{F_{n_k}\}$ is a subsequence converging weakly to F , then since $R - (-n, n)$ is closed,

$$\limsup_k F_{n_k}(R - (-n, n)) \leq F(R - (-n, n)).$$

Thus $F(R - (-n, n)) \geq \varepsilon$ for all n , and if we let $n \rightarrow \infty$, we obtain $0 \geq \varepsilon$, a contradiction. ■

Prokhorov's theorem holds equally well for distribution functions on R^k (with $F_i(R^k) \leq M < \infty$ for all i); the proof is essentially the same as above.

8.2.5 Corollary. Let $\{F_n\}$ be a bounded sequence of distribution functions on R . If $\{F_n\}$ is tight and every weakly convergent subsequence of $\{F_n\}$ converges to the distribution function F , then $F_n \xrightarrow{w} F$.

PROOF. If F_n does not converge weakly to F , then $\int_R f(x) dF_n(x) \not\rightarrow \int_R f(x) dF(x)$ for some bounded continuous f . Hence there is an $\varepsilon > 0$ such that

$$\left| \int_R f(x) dF_n(x) - \int_R f(x) dF(x) \right| \geq \varepsilon$$

for infinitely many n , say for $n \in T$. By 8.2.4, $\{F_n, n \in T\}$ has a subsequence $\{F_{n_k}\}$ converging weakly to a distribution function G , and $G = F$ by hypothesis. But then $\int_R f(x) dF_{n_k}(x) \rightarrow \int_R f(x) dF(x)$, a contradiction. ■

8.2.6 Corollary. Let $\{F_n\}$ be a bounded sequence of distribution functions on R . If $\{F_n\}$ is tight, then $\{F_n\}$ converges weakly iff $\int_R e^{iux} dF_n(x)$ approaches a finite limit as $n \rightarrow \infty$ for each $u \in R$.

PROOF. Assume $\int_R e^{iux} dF_n(x)$ has a finite limit for all u . By 8.2.4, there is a subsequence $\{F_{n_k}\}$ converging weakly to a distribution function F . If F_n does not converge weakly to F , by 8.2.5, there is a subsequence $\{F_{m_k}\}$ converging

weakly to a distribution function $G \neq F$. We know by hypothesis that $\int_R e^{iux} dF_{n_k}(x)$ and $\int_R e^{iux} dF_{m_k}(x)$ have the same limit as $k \rightarrow \infty$. Therefore $\int_R e^{iux} dF(x) = \int_R e^{iux} dG(x)$ for all $u \in R$. By 8.1.3, $F = G$, a contradiction; thus $F_n \xrightarrow{w} F$. The converse follows from the definition of weak convergence. (In this proof, as in 8.2.5, distribution functions that differ by a constant have been identified.) ■

One more result is needed before we can relate weak convergence to convergence of characteristic functions.

8.2.7 Truncation Inequality. Let F be a bounded distribution function on R , with characteristic function h . If $u > 0$, then for some constant $k > 0$.

$$\int_{|x| \geq 1/u} dF(x) \leq \frac{k}{u} \int_0^u [h(0) - \operatorname{Re} h(v)] dv.$$

PROOF.

$$\begin{aligned} \frac{1}{u} \int_0^u [h(0) - \operatorname{Re} h(v)] dv &= \frac{1}{u} \int_0^u \int_{-\infty}^{\infty} (1 - \cos vx) dF(x) dv \\ &= \int_{-\infty}^{\infty} \left[\frac{1}{u} \int_0^u (1 - \cos vx) dv \right] dF(x) \\ &\quad \text{by Fubini's theorem} \\ &= \int_{-\infty}^{\infty} \left(1 - \frac{\sin ux}{ux} \right) dF(x) \\ &\geq \inf_{|t| \geq 1} \left(1 - \frac{\sin t}{t} \right) \int_{|ux| \geq 1} dF(x) \\ &= \frac{1}{k} \int_{|x| \geq 1/u} dF(x). \quad \blacksquare \end{aligned}$$

In fact,

$$\inf_{|t| \geq 1} \left(1 - \frac{\sin t}{t} \right) = 1 - \sin 1 \geq \frac{1}{7},$$

so we may take $k = 7$.

8.2.8 Lévy's Theorem. Let $\{F_n\}$ be a bounded sequence of distribution functions on R , and let $\{h_n\}$ be the corresponding sequence of characteristic functions. If $F_n \xrightarrow{w} F$, where F is a distribution function with characteristic function h , then $h_n(u) \rightarrow h(u)$ for all u . Conversely, if h_n converges pointwise to a complex-valued function h , where h is continuous at $u = 0$, then h is the characteristic function of some bounded distribution function F , and $F_n \xrightarrow{w} F$.

PROOF. The first assertion follows from the definition of weak convergence, so assume $h_n(u) \rightarrow h(u)$ for all u , with h continuous at the origin. We claim that $\{F_n\}$ is tight. For by 8.2.7,

$$\begin{aligned} \int_{|x| \geq 1/u} dF_n(x) &\leq \frac{k}{u} \int_0^u [h_n(0) - \operatorname{Re} h_n(v)] dv, u > 0 \\ &\rightarrow \frac{k}{u} \int_0^u [h(0) - \operatorname{Re} h(v)] dv \quad \text{as } n \rightarrow \infty, \\ &\quad \text{by the dominated convergence theorem.} \end{aligned}$$

Since h is continuous at 0,

$$\frac{k}{u} \int_0^u [h(0) - \operatorname{Re} h(v)] dv \rightarrow 0 \quad \text{as } u \rightarrow 0;$$

hence, given $\varepsilon > 0$, we may choose u so small that

$$\int_{|x| \geq 1/u} dF_n(x) < \varepsilon \quad \text{for all } n,$$

proving tightness. By 8.2.6, F_n converges weakly to a distribution function F ; hence h_n converges pointwise to the characteristic function of F . But we know that $h_n \rightarrow h$, so that h is the characteristic function of F . ■

The following variation of Lévy's theorem is often useful.

8.2.9 Theorem. Let $\{F_n\}$ be a bounded sequence of distribution functions, and $\{h_n\}$ the corresponding sequence of characteristic functions. If F is a bounded distribution function with characteristic function h , then $F_n \xrightarrow{w} F$ iff $h_n(u) \rightarrow h(u)$ for all u , and in this case, h_n converges to h uniformly on bounded intervals.

PROOF. If $F_n \xrightarrow{w} F$, then $h_n(u) \rightarrow h(u)$ for all u by definition of weak convergence. If $h_n(u) \rightarrow h(u)$ for all u , then by 8.2.8, F_n converges weakly to the distribution function whose characteristic function is h , namely, F .

Now let I be a bounded interval of R . Then

$$h_n(u + \delta) - h_n(u) = \int_R (e^{i(u+\delta)x} - e^{iux}) dF_n(x);$$

hence

$$|h_n(u + \delta) - h_n(u)| \leq \int_I |e^{i\delta x} - 1| dF_n(x) + 2F_n(R - I).$$

Since $F_n \xrightarrow{w} F$, $\{F_n\}$ is relatively compact, and therefore tight by 8.2.4. Thus if $\varepsilon > 0$ is given, we may choose I so that $2F_n(R - I) < \varepsilon/2$ for all n . If $u \in R$, and M is a bound on $\{F_n(R), n = 1, 2, \dots\}$, then

$$\begin{aligned}|h_n(u + \delta) - h_n(u)| &\leq M \sup_{x \in I} |e^{i\delta x} - 1| + \frac{\varepsilon}{2} \\ &< \varepsilon \quad \text{for small enough } \delta = \delta(\varepsilon).\end{aligned}$$

Thus the h_n are equicontinuous on R . But equicontinuity and pointwise convergence imply uniform convergence on compact sets (see the appendix on general topology, proof of Theorem A8.5); hence $h_n \rightarrow h$ uniformly on I . ■

Problems

1. If X has density $(1/\pi)(1 - \cos x)/x^2$, the characteristic function of X is $h(u) = 1 - |u|$, $|u| \leq 1$; $h(u) = 0$, $|u| > 1$. [See Ash (1970, p. 166, Problem 8).] It is possible to construct a different characteristic function f that agrees with h on $[-1, 1]$, as follows.

Show that the Fourier series expansion on the interval $[-1, 1]$ of $f(u) = 1 - |u|$ (extended periodically) is

$$f(u) = \frac{1}{2} + \sum_{n=-\infty}^{\infty} \frac{2}{\pi^2(2n+1)^2} \exp(i(2n+1)\pi u).$$

[Since f is continuous and of bounded variation, and $f(-1) = f(1)$, the Fourier series converges uniformly to f ; see Titchmarsh (1939, p. 410).]

Thus if X is a discrete random variable with $P\{X = 0\} = \frac{1}{2}$, $P\{X = (2n+1)\pi\} = 2\pi^{-2}(2n+1)^{-2}$, n an integer, then X has characteristic function f . Since f is periodic, we have $f \neq h$, as desired.

2. Using Problem 1, give examples to show that the following results are possible (the h 's are characteristic functions of random variables).
- $h_1 h_2 = h_1 h_3$ does not imply $h_2 = h_3$.
 - $h_n \rightarrow h$ on $[-1, 1]$ does not imply $h_n \rightarrow h$ everywhere.
3. Give an example of a sequence of characteristic functions converging pointwise to a function that is not a characteristic function.
4. (a) If $\{a_n\}$ is a sequence of complex numbers and $\exp(iua_n)$ converges to a (finite) limit $g(u)$ for almost all u in the open interval $I \subset R$, show that $\{a_n\}$ converges.
- (b) If $X_n \equiv n$, so that $h_n(u) = e^{i\mu u}$, $n = 1, 2, \dots$, show that the sequence of characteristic functions h_n has no pointwise convergent subsequence. Thus the corresponding sequence of distribution functions has no weakly convergent subsequence, as was verified directly in 8.2.2.

5. Let F_0, F_1, F_2, \dots be distribution functions on R , and assume $F_n(-\infty) = 0, n = 0, 1, \dots, F_n(\infty) = 1, n = 1, 2, \dots$. Give examples to show that the following situations are possible.
- $F_n(x) \rightarrow F_0(x)$ for all $x \in R$ at which F is continuous, but F_n does not converge weakly to F_0 .
 - $F_n(b) - F_n(a) \rightarrow F_0(b) - F_0(a)$ for all $a, b \in R$ at which F_0 is continuous, but $\lim_{n \rightarrow \infty} F_n(x)$ does not exist for any $x \in R$. In particular, F_n does not converge weakly to F_0 .
6. Let F_0, F_1, F_2, \dots be distribution functions on R , and assume that $F_n(\infty) - F_n(-\infty) \leq M < \infty$ for all n , and $F_n(a, b] \rightarrow F_0(a, b]$ for all $a, b \in R$ at which F_0 is continuous. Show that F_n converges weakly to F_0 iff $\{F_n\}$ is tight.
7. Let F_0, F_1, F_2, \dots be distribution functions on R , and assume that $F_n(\infty) - F_n(-\infty) \leq M < \infty$ for all n . Show that F_n converges weakly to F_0 iff $F_n(a, b] \rightarrow F_0(a, b]$ for all $a, b \in R$ at which F_0 is continuous, and $F_n(\infty) - F_n(-\infty) \rightarrow F_0(\infty) - F_0(-\infty)$.
8. Let X, X_1, X_2, \dots be random variables, and assume that X_n converges to X in probability, that is, for every $\varepsilon > 0$, $P\{|X_n - X| \geq \varepsilon\} \rightarrow 0$ as $n \rightarrow \infty$. Show that $X_n \xrightarrow{d} X$. Give an example to show that the converse is not true in general; however, if X_n converges in distribution to a constant c , then $X_n \xrightarrow{P} c$.
9. Let Y_1, Y_2, \dots be independent random variables, and let \mathcal{F}_n be the σ -field $\mathcal{F}(Y_1, \dots, Y_n)$, $n = 1, 2, \dots$. If h_k is the characteristic function of Y_k and u is a fixed real number such that $h_k(u) \neq 0$, $k = 1, 2, \dots$, define

$$X_n = \prod_{k=1}^n \frac{1}{h_k(u)} \exp(iu Y_k).$$

- Show that $E(X_{n+1} | \mathcal{F}_n) = X_n$ a.e., and hence $\{X_n, \mathcal{F}_n\}$ is a complex-valued martingale. (In other words, $\{\operatorname{Re} X_n, \mathcal{F}_n\}$ and $\{\operatorname{Im} X_n, \mathcal{F}_n\}$ are martingales.)
- Assume that $\sum_{k=1}^n Y_k \xrightarrow{d} X$. Show that there is an open interval $I \subset R$, with $0 \in I$, such that for each $u \in I$, $\exp[iu \sum_{k=1}^n Y_k(\omega)]$ converges for almost every ω .
- Under the hypothesis of part (b), show that for almost every ω , $\exp[iu \sum_{k=1}^n Y_k(\omega)]$ converges for almost every $u \in I$ (Lebesgue measure). Thus by Problem 4, $\sum_{k=1}^\infty Y_k(\omega)$ converges a.e. to a finite limit.
- Conclude that for a series of independent random variables, convergence in distribution, convergence in probability, and convergence almost everywhere are equivalent.

10. If F_1, F_2, \dots are bounded distribution functions on R^k , with corresponding Lebesgue–Stieltjes measures μ_1, μ_2, \dots , we say that the sequence $\{F_n\}$ converges weakly iff $\{\mu_n\}$ converges weakly. If the F_n are the distribution functions of k -dimensional random vectors X_n , we also say that the sequence $\{X_n\}$ converges in distribution.

Show that the Lévy theorems 8.2.8 and 8.2.9 hold for distribution functions on R^k .

8.3 Convergence to a Normal Distribution

Let X_1, X_2, \dots be independent random variables, with each X_k having finite mean m_k and finite variance σ_k^2 . Let $S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \dots$; then $E(S_n) = \sum_{k=1}^n m_k$, $\text{Var } S_n = c_n^{-2} = \sum_{k=1}^n \sigma_k^2$. We consider the normalized sum $T_n = c_n^{-1}(S_n - E(S_n))$, which has mean 0 and variance 1. (To avoid degeneracy, we assume that $c_n > 0$ for sufficiently large n .) If X^* is a random variable having the normal distribution with mean 0 and variance 1, so that the distribution function of X^* is

$$F^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp[-\frac{1}{2}t^2] dt,$$

we ask for conditions under which T_n converges in distribution to X^* .

A long series of preliminary results were derived before a satisfactory solution was obtained, giving conditions that are sufficient and “almost” necessary for convergence to X^* . We consider sufficiency first.

8.3.1 Lindeberg's Theorem. Let $S_n = X_1 + \dots + X_n$, $n = 1, 2, \dots$, where the X_k are independent random variables with finite mean m_k and finite variance σ_k^2 . Let $T_n = c_n^{-1}(S_n - E(S_n))$, where $c_n^{-2} = \text{Var } S_n = \sum_{k=1}^n \sigma_k^2$, and let F_k be the distribution function of X_k . If for every $\varepsilon > 0$,

$$\frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x-m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

then T_n converges in distribution to a random variable X^* that is normal with mean 0 and variance 1.

Before proving the theorem, we examine some of its implications. Lindeberg's theorem implies that $T_n \xrightarrow{d} X^*$ under any one of the following conditions.

1. *The uniformly bounded case.* Assume $|X_k| \leq M$ for all k , and $c_n \rightarrow \infty$. Then

$$\begin{aligned} \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) &= E[(X_k - m_k)^2 I_{\{|X_k - m_k| \geq \varepsilon c_n\}}] \\ &\leq (2M)^2 P\{|X_k - m_k| \geq \varepsilon c_n\} \\ &\leq \frac{(2M)^2 \sigma_k^2}{\varepsilon^2 c_n^2} \end{aligned}$$

by Chebyshev's inequality. Thus

$$\frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \leq \frac{(2M)^2}{\varepsilon^2 c_n^2} \rightarrow 0.$$

2. *The identically distributed case.* Assume that the X_k are iid, with finite mean m and finite variance $\sigma^2 > 0$. If F is the distribution function of the X_k , then

$$\begin{aligned} \frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \\ &= \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{\{x: |x - m| \geq \varepsilon\sigma\sqrt{n}\}} (x - m)^2 dF(x) \\ &= \frac{1}{\sigma^2} \int_{\{x: |x - m| \geq \varepsilon\sigma\sqrt{n}\}} (x - m)^2 dF(x) \\ &\rightarrow 0 \quad \text{since } \sigma^2 \text{ is finite and } \{x: |x - m| \geq \varepsilon\sigma\sqrt{n}\} \downarrow \emptyset \quad \text{as } n \rightarrow \infty. \end{aligned}$$

3. *The Bernoulli case.* Let S_n be the number of successes in n Bernoulli trials, with probability of success p on a given trial. We may write

$$S_n = X_1 + \cdots + X_n,$$

where the X_k are independent and $P\{X_k = 1\} = p$, $P\{X_k = 0\} = q = 1 - p$. (We may take X_k as the indicator of a success on trial k .) Thus case 2 applies, with $m = E(X_k) = p$, $\sigma^2 = E(X_k^2) - [E(X_k)]^2 = p(1 - p)$, $E(S_n) = nm = np$, $c_n^2 = n\sigma^2 = np(1 - p)$. Thus

$$T_n = \frac{S_n - np}{(npq)^{1/2}}$$

and $T_n \xrightarrow{d} X^*$, that is, $P\{T_n \leq x\} \rightarrow F^*(x)$ for all x .

4. *Liapounov's condition.* Assume that

$$\frac{1}{c_n^{2+\delta}} \sum_{k=1}^n E[|X_k - m_k|^{2+\delta}] \rightarrow 0 \quad \text{for some } \delta > 0.$$

Then

$$\begin{aligned} E[|X_k - m_k|^{2+\delta}] &= \int_{-\infty}^{\infty} |x - m_k|^{2+\delta} dF_k(x) \\ &\geq \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} |x - m_k|^{\delta} |x - m_k|^2 dF_k(x) \\ &\geq \varepsilon^{\delta} c_n^{\delta} \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x). \end{aligned}$$

Thus

$$\begin{aligned} \frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) &\leq \frac{1}{c_n^2} \sum_{k=1}^n \frac{E[|X_k - m_k|^{2+\delta}]}{\varepsilon^{\delta} c_n^{\delta}} \\ &= \frac{\sum_{k=1}^n E[|X_k - m_k|^{2+\delta}]}{\varepsilon^{\delta} c_n^{2+\delta}} \\ &\rightarrow 0. \end{aligned}$$

PROOF OF THEOREM 8.3.1. We may assume without loss of generality that all $m_k = 0$. For if we have proved the theorem under this restriction, let

$$\begin{aligned} X'_k &= X_k - m_k \quad (\text{so that } EX'_k \equiv 0) \\ S'_n &= \sum_{k=1}^n X'_k \\ T'_n &= \frac{S'_n - ES'_n}{c'_n} = \frac{S_n - ES_n}{c_n} = T_n. \end{aligned}$$

Since

$$\begin{aligned} \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) &= E[(X_k - m_k)^2 I_{\{|X_k - m_k| \geq \varepsilon c_n\}}] \\ &= E[(X'_k)^2 I_{\{|X'_k| \geq \varepsilon c_n'\}}] \\ &= \int_{\{x: |x| \geq \varepsilon c_n'\}} x^2 dF'_k(x), \end{aligned}$$

the Lindeberg hypothesis applies to the random variables X'_k ; hence $T'_n \xrightarrow{d} X^*$. But then $T'_n \xrightarrow{d} X^*$, as desired.

The following estimates will be needed. If y is any real number and z a complex number with $|z| \leq \frac{1}{2}$,

$$e^{iy} = 1 + iy + \frac{\theta y^2}{2}, \tag{1}$$

$$e^{iy} = 1 + iy - \frac{y^2}{2} + \frac{\theta_1 |y|^3}{6}, \tag{2}$$

where θ and θ_1 depend on y , and $|\theta| \leq 1$, $|\theta_1| \leq 1$;

$$\text{Log}(1+z) = z + \theta' |z|^2, \quad (3)$$

where "Log" denotes the principal branch of the logarithm and $|\theta'| \leq 1$, θ' depending on z . [These formulas are exercises in calculus; for full details see Ash (1970, p. 173).]

Throughout the proof, h_k will denote the characteristic function of X_k , and u will be a fixed real number. The characteristic function of $T_n = S_n/c_n$ is

$$h_{T_n}(u) = E(e^{iuT_n}) = E(e^{iuS_n/c_n}) = h_{S_n}\left(\frac{u}{c_n}\right) = \prod_{k=1}^n h_k\left(\frac{u}{c_n}\right). \quad (4)$$

By (1) and (2),

$$\begin{aligned} h_k(u) &= \int_{-\infty}^{\infty} e^{iux} dF_k(x) \\ &= \int_{|x| \geq \varepsilon c_n} \left(1 + iux + \frac{\theta u^2 x^2}{2}\right) dF_k(x) \\ &\quad + \int_{|x| < \varepsilon c_n} \left(1 + iux - \frac{u^2 x^2}{2} + \frac{\theta_1 |u|^3 |x|^3}{6}\right) dF_k(x), \end{aligned}$$

$|\theta|, |\theta_1| \leq 1$, θ, θ_1 depending on ux . Since

$$\int_{-\infty}^{\infty} iux dF_k(x) = iuEX_k = 0,$$

we may drop the terms involving iux . Thus

$$\begin{aligned} h_k\left(\frac{u}{c_n}\right) &= 1 + \frac{u^2}{2c_n^2} \int_{|x| \geq \varepsilon c_n} \theta x^2 dF_k(x) \\ &\quad - \frac{u^2}{2c_n^2} \int_{|x| < \varepsilon c_n} x^2 dF_k(x) + \frac{|u|^3}{6c_n^3} \int_{|x| < \varepsilon c_n} \theta_1 |x|^3 dF_k(x). \end{aligned} \quad (5)$$

Now

$$\left| \frac{1}{2} \int_{|x| \geq \varepsilon c_n} \theta x^2 dF_k(x) \right| \leq \frac{1}{2} \int_{|x| \geq \varepsilon c_n} x^2 dF_k(x);$$

hence

$$\frac{1}{2} \int_{|x| \geq \varepsilon c_n} \theta x^2 dF_k(x) = \theta_2 \int_{|x| \geq \varepsilon c_n} x^2 dF_k(x), \quad \text{where } |\theta_2| \leq \frac{1}{2}.$$

Similarly,

$$\begin{aligned} \left| \frac{1}{6} \int_{|x| < \varepsilon c_n} \theta_1 |x|^3 dF_k(x) \right| &\leq \frac{1}{6} \int_{|x| < \varepsilon c_n} |x|^3 dF_k(x) \\ &\leq \frac{1}{6} \int_{|x| < \varepsilon c_n} \frac{\varepsilon c_n}{|x|} |x|^3 dF_k(x) \end{aligned}$$

(note that $|x| < \varepsilon c_n$ implies $\varepsilon c_n/|x| \geq 1$). Hence

$$\frac{1}{6} \int_{|x| < \varepsilon c_n} \theta_1 |x|^3 dF_k(x) = \theta_3 \int_{|x| < \varepsilon c_n} \varepsilon c_n x^2 dF_k(x), \quad |\theta_3| \leq \frac{1}{6}.$$

Thus if we set

$$\begin{aligned} \alpha_{nk} &= \frac{1}{c_n^2} \int_{|x| \geq \varepsilon c_n} x^2 dF_k(x), \\ \beta_{nk} &= \frac{1}{c_n^2} \int_{|x| < \varepsilon c_n} x^2 dF_k(x) \quad (\leq \varepsilon^2), \end{aligned}$$

Eq. (5) becomes

$$h_k\left(\frac{u}{c_n}\right) = 1 + \gamma_{nk}, \quad (6)$$

where

$$\gamma_{nk} = \theta_2 u^2 \alpha_{nk} - \frac{u^2}{2} \beta_{nk} + |u|^3 \varepsilon \theta_3 \beta_{nk}. \quad (7)$$

Now $\sum_{k=1}^n \alpha_{nk} \rightarrow 0$ by hypothesis, and

$$\sum_{k=1}^n (\alpha_{nk} + \beta_{nk}) = \frac{1}{c_n^2} \sum_{k=1}^n \sigma_k^2 = 1.$$

Thus by (7) and the fact that $\beta_{nk} \leq \varepsilon^2$, we have

$$\max_k |\gamma_{nk}| \leq \frac{u^2 \varepsilon^2}{2} + |u|^3 \varepsilon \quad \text{and} \quad \sum_k |\gamma_{nk}| \leq \frac{u^2}{2} + |u|^3 \varepsilon$$

for sufficiently large n . By (3),

$$\frac{u^2}{2} + \sum_{k=1}^n \log(1 + \gamma_{nk}) = \frac{u^2}{2} + \sum_{k=1}^n (\gamma_{nk} + \theta' |\gamma_{nk}|^2), \quad |\theta'| \leq 1.$$

Since

$$\sum_{k=1}^n |\gamma_{nk}|^2 \leq \left(\max_k |\gamma_{nk}| \right) \sum_{k=1}^n |\gamma_{nk}|,$$

it follows from (7) that if $\delta > 0$ is given, and ε is chosen sufficiently small (depending on δ and u), then

$$\left| \frac{u^2}{2} + \sum_{k=1}^n \text{Log}(1 + \gamma_{nk}) \right|$$

will be less than δ for sufficiently large n . Since the exponential function is continuous, we obtain from (6) that

$$\exp\left(\frac{u^2}{2}\right) \prod_{k=1}^n h_k\left(\frac{u}{c_n}\right) \rightarrow 1,$$

in other words [see Eq. (4)], $h_{T_n}(u) \rightarrow \exp(-u^2/2)$, the characteristic function of X^* . By 8.2.9, $T_n \xrightarrow{d} X^*$. ■

There is another proof of the Lindeberg theorem that uses weak convergence directly, rather than via Lévy's theorem; see Billingsley (1968, p. 42).

We now show that the Lindeberg hypothesis is not necessary for convergence to X^* . If the Lindeberg condition holds, we claim that for all $\varepsilon > 0$.

$$P\left\{\frac{|X_k - m_k|}{c_n} \geq \varepsilon\right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

uniformly in k. This is referred to as *uniform asymptotic negligibility* (uan) of the random variables $(X_k - m_k)/c_n$. For

$$\begin{aligned} \frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x-m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \\ \geq \varepsilon^2 \sum_{k=1}^n P\{|X_k - m_k| \geq \varepsilon c_n\} \\ \geq \varepsilon^2 \max_{1 \leq k \leq n} P\{|X_k - m_k| \geq \varepsilon c_n\}. \end{aligned}$$

Thus, intuitively, the contribution of each $(X_k - m_k)/c_n$ is small relative to the sum $(S_n - ES_n)/c_n$.

If we can construct an example of a sequence which is not uan but for which $T_n \xrightarrow{d} X^*$, we then have convergence to X^* without the Lindeberg condition; here is one possibility. Let X_1, X_2, \dots be independent, normally distributed random variables with zero mean. Let $\sigma_k^2 = 2^{k-2}$, $k \geq 2$; $\sigma_1^2 = 1$. Then $c_n^2 = \sum_{k=1}^n \sigma_k^2 = 2^{n-1}$; hence X_n/c_n is normal with mean 0 and variance $\frac{1}{2}$. Therefore

$$\max_{1 \leq k \leq n} P\left\{\left|\frac{X_k}{c_n}\right| \geq \varepsilon\right\} \geq P\left\{\left|\frac{X_n}{c_n}\right| \geq \varepsilon\right\},$$

which is a positive constant not depending on n . Thus the X_k/c_n are not uan, although $T_n \xrightarrow{d} X^*$. [In fact T_n is normal $(0, 1)$ for all n .]

If we impose the uan requirement, the Lindeberg condition becomes necessary and sufficient for convergence to X^* .

8.3.2 Theorem. Let X_1, X_2, \dots be independent random variables, with each X_k having finite mean m_k and finite variance σ_k^2 . Then the Lindeberg condition

$$\frac{1}{c_n^2} \sum_{k=1}^n \int_{\{x: |x - m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \rightarrow 0 \quad \text{for all } \varepsilon > 0$$

holds if and only if $T_n \xrightarrow{d} X^*$ and the $(X_k - m_k)/c_n$ are uan.

PROOF. The “only if” part follows from 8.3.1 and the above remarks. The “if” part, due to Feller, is rather lengthy, and is proved in Section 8.8. ■

In the Lindeberg theorem, we have normalized the sum S_n in a special way, that is, we have considered a sequence of random variables $a_n^{-1}(S_n - b_n)$, where b_n is the mean and a_n the standard deviation of S_n . We might ask whether a different choice of constants a_n and b_n would produce different results, for example, convergence to a nonnormal random variable. Questions of this nature may be handled by the “theorem on convergence of types,” which we now develop.

8.3.3 Definition. Two random variables X and Y (or their distribution functions G and F) are said to be of the *same positive type* iff X and $a^{-1}(Y - b)$ have the same distribution for some real a and b , $a > 0$, that is, $G(x) = F(ax + b)$ for all x ; X and Y are of the *same type* iff X and $a^{-1}(Y - b)$ have the same distribution for some real a, b with $a \neq 0$, not necessarily positive. (The notation $X \stackrel{d}{=} Y$ will indicate that X and Y have the same distribution.)

The notion of type is preserved under convergence in distribution, as the following basic theorem shows.

8.3.4 Convergence of Types Theorem. (a) Let $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, where for each n , X_n and Y_n are of the same positive type, with $X_n \stackrel{d}{=} a_n^{-1}(Y_n - b_n)$, $a_n > 0$. Assume X and Y are *nondegenerate*, that is, not a.e. constant. Then there are real numbers a and b , with $a > 0$, such that $a_n \rightarrow a$, $b_n \rightarrow b$, and $X \stackrel{d}{=} a^{-1}(Y - b)$; thus X and Y are of the same positive type.

(b) If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, where for each n , X_n and Y_n are of the same type, with $X_n \stackrel{d}{=} a_n^{-1}(Y_n - b_n)$, $a_n \neq 0$, and X and Y are nondegenerate, then there

are real numbers a and b , with $a \neq 0$, such that $|a_n| \rightarrow |a|$ and $X \stackrel{d}{=} a^{-1}(Y - b)$; thus X and Y are of the same type.

The proof is an intricate exercise in real analysis, and is done in Section 8.9.

We can now consider the question raised earlier about the normalizing constants. If $(a_n')^{-1}(S_n - b_n') \xrightarrow{d} Y$, and $a_n^{-1}(S_n - b_n) \xrightarrow{d} X$, where Y is normal and X is nondegenerate, then X must be normal. For

$$a_n^{-1}(S_n - b_n) = \frac{a_n'}{a_n} [(a_n')^{-1}(S_n - b_n') - (a_n')^{-1}(b_n - b_n')];$$

hence by the convergence of types theorem, X and $a^{-1}(Y - b)$ have the same distribution for some a, b , with $a \neq 0$. But a brief computation then shows that X is normal. (To do this, look at characteristic functions, or use the technique of 5.9.4.) Thus if one set of constants produces a normal limit, all nondegenerate limits are normal.

The restriction that X be nondegenerate is necessary. For given any real number c , it is always possible to choose the constants a_n and b_n so that $a_n^{-1}(S_n - b_n) \xrightarrow{d} c$ (see Problem 1).

There is a tricky aspect of Theorem 8.3.2 that is worth mentioning. Under the uan hypothesis, the Lindeberg condition is necessary and sufficient for $T_n \xrightarrow{d} X^*$; thus if the $c_n^{-1}(X_k - m_k)$ are uan and the Lindeberg condition fails, $c_n^{-1}(S_n - E(S_n))$ cannot converge in distribution to a normal $(0, 1)$ random variable. However, it is possible for $c_n^{-1}(S_n - E(S_n))$ to converge in distribution to a random variable X that is not degenerate and not normal $(0, 1)$. In Problem 4, an example is given in which X is normal with a variance unequal to 1.

In fact there are examples of independent (but not identically distributed) random variables X_1, X_2, \dots , each with finite mean and variance, such that for some constants a_n, b_n we have $a_n^{-1}(S_n - b_n) \xrightarrow{d} X$, where X is not degenerate and not normal. The construction of such examples is quite elaborate, and we give the reference only: Gnedenko and Kolmogorov (1954, p. 152).

Problems

- Let X_1, X_2, \dots be an arbitrary sequence of random variables. Given any real number c , show that constants a_n and b_n ($a_n > 0$) can always be chosen so that $a_n^{-1}(X_1 + \dots + X_n - b_n)$ converges in distribution to a random variable $X \equiv c$.
- Give an example of sequences of random variables $\{X_n\}$ and $\{Y_n\}$ such that $X_n \stackrel{d}{=} a_n^{-1}(Y_n - b_n)$ for real numbers a_n and b_n ($a_n \neq 0$), $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$ (so that X and Y are of the same type), but $|b_n|$ has no limit.

3. Let $\{X_{nk}, n = 1, 2, \dots, k = 1, \dots, n\}$ be a double sequence of random variables, and let h_{nk} be the characteristic function of X_{nk} . Show that the X_{nk} are uan (in other words,

$$\max_{1 \leq k \leq n} P\{|X_{nk}| \geq \varepsilon\} \rightarrow 0$$

as $n \rightarrow \infty$ for every $\varepsilon > 0$ iff

$$\max_{1 \leq k \leq n} |h_{nk}(u) - 1| \rightarrow 0$$

as $n \rightarrow \infty$, and in this case, the convergence is uniform on any bounded interval. (Use 8.2.7 in the "if" part.)

4. Let X_1, X_2, \dots be independent random variables, defined as follows:
 $X_1 = \pm 1$ with equal probability.
If $k > 1$, and c is a fixed real number greater than 1,

$$P\{X_k = 1\} = P\{X_k = -1\} = \frac{1}{2c},$$

$$P\{X_k = k\} = P\{X_k = -k\} = \frac{1}{2k^2} \left(1 - \frac{1}{c}\right),$$

$$P\{X_k = 0\} = 1 - \frac{1}{c} - \frac{1}{k^2} \left(1 - \frac{1}{c}\right).$$

Define

$$X'_{nk} = \begin{cases} X_k & \text{if } |X_k| \leq \sqrt{n}, \\ 0 & \text{if } |X_k| > \sqrt{n}. \end{cases}$$

Establish the following:

- (a) The X_k/c_n satisfy the uan condition.
- (b) The Lindeberg condition fails for the X_k , but holds for the X'_{nk} . Furthermore, if $S'_n = \sum_{k=1}^n X'_{nk}$, then $S'_n/c_n \xrightarrow{d}$ normal $(0, 1)$, where $(c_n')^2 = \text{Var } S'_n \sim n/c$.
- (c) If $S_n = \sum_{k=1}^n X_k$, then $P\{S_n \neq S'_n\} \rightarrow 0$ as $n \rightarrow \infty$.
- (d) $\sqrt{c} S_n/\sqrt{n} \xrightarrow{d}$ normal $(0, 1)$, but $S_n/\sqrt{n} \not\xrightarrow{d}$ normal $(0, 1)$.

8.4 Stable Distributions

If X_1, X_2, \dots are independent, identically distributed random variables, with finite mean m and finite variance σ^2 , we know from the previous section that $(S_n - nm)/\sigma\sqrt{n} \xrightarrow{d} X^*$ normal $(0, 1)$; hence any limiting distribution of a

sequence $a_n^{-1}(S_n - b_n)$ must be normal. If we drop the finite variance requirement, it is possible to obtain a nonnormal limit. For example, let the X_i have the Cauchy density $f(x) = \theta/\pi(x^2 + \theta^2)$, $x \in R$, θ a fixed positive constant. The corresponding characteristic function is $h(u) = e^{-\theta|u|}$ [see Ash (1970, p. 161), for the computation]. Therefore S_n has characteristic function $[h(u)]^n = e^{-n\theta|u|}$; hence $n^{-1}S_n$ has characteristic function $[h(u/n)]^n = e^{-\theta|u|}$. Thus $n^{-1}S_n \xrightarrow{d} X$, where X has the Cauchy density with parameter θ . Since $E(|X|) = \infty$, this does not contradict the previous results.

The following investigation is suggested. Let X_1, X_2, \dots be iid random variables. If $a_n^{-1}(S_n - b_n) \xrightarrow{d} X$, what are the possible distributions of X ? (We may assume that $a_n > 0$; for if negative a_n are allowed, we consider the two subsequences corresponding to $a_n > 0$ and $a_n < 0$.) In fact the possible limiting distributions may be completely characterized, as follows:

8.4.1 Definition. A random variable X (or its distribution function F , or its characteristic function h) is said to be *stable* iff, whenever X_1, \dots, X_n are iid random variables with distribution function F , then $S_n = X_1 + \dots + X_n$ is of the same positive type as X ; in other words, $X \stackrel{d}{=} a_n^{-1}(S_n - b_n)$, or equivalently $[h(u)]^n = \exp[ib_n u]h(a_n u)$, $u \in R$, for appropriate $a_n > 0$ and b_n .

A sequence $\{a_n^{-1}(S_n - b_n)\}$, where $a_n > 0$ and $S_n = \sum_{k=1}^n X_k$, the X_k iid, is called a sequence of *normed sums*.

8.4.2 Theorem. The random variable X is stable iff there is a sequence of normed sums converging in distribution to X .

PROOF. If X is stable, let X_1, \dots, X_n be iid with $X_j \stackrel{d}{=} X$. Then $a_n^{-1}(S_n - b_n) \xrightarrow{d} X$ for appropriate $a_n > 0$ and b_n ; in particular, $a_n^{-1}(S_n - b_n) \xrightarrow{d} X$.

Assume X_1, X_2, \dots iid, with $V_n = a_n^{-1}(S_n - b_n) \xrightarrow{d} X$. If X is degenerate, it is stable, so assume X nondegenerate. Fix the positive integer r , and define

$$\begin{aligned} S_n^{(1)} &= X_1 + \dots + X_n, \\ S_n^{(2)} &= X_{n+1} + \dots + X_{2n}, \\ &\vdots \\ S_n^{(r)} &= X_{(r-1)n+1} + \dots + X_{rn}. \end{aligned}$$

Then let

$$W_n^{(r)} = \frac{S_n^{(1)} - b_n}{a_n} + \frac{S_n^{(2)} - b_n}{a_n} + \dots + \frac{S_n^{(r)} - b_n}{a_n} = Z_n^{(1)} + \dots + Z_n^{(r)},$$

where $Z_n^{(1)}, \dots, Z_n^{(r)}$ are independent. Now $Z_n^{(i)} \stackrel{d}{=} Z_n^{(1)}$ for all i ; hence $Z_n^{(i)} \xrightarrow{d} X$

for each i . It follows from 8.2.9 that $W_n^{(r)} \xrightarrow{d} Z_1 + \cdots + Z_r$, where Z_1, \dots, Z_r are iid with $Z_i \stackrel{d}{=} X$. But we may also write

$$\begin{aligned} W_n^{(r)} &= \frac{X_1 + \cdots + X_{rn} - rb_n}{a_n} = \frac{a_{rn}}{a_n} \left(\frac{X_1 + \cdots + X_{rn} - b_{rn}}{a_{rn}} \right) + \frac{b_{rn} - rb_n}{a_n} \\ &= \alpha_n^{(r)} V_{rn} + \beta_n^{(r)}, \end{aligned}$$

where $\alpha_n^{(r)} = a_{rn}/a_n > 0$. To summarize:

$$V_{rn} = \frac{W_n^{(r)} - \beta_n^{(r)}}{\alpha_n^{(r)}}, \quad V_{rn} \xrightarrow{d} X, \quad W_n^{(r)} \xrightarrow{d} Z_1 + \cdots + Z_r.$$

By 8.3.4(a), $\alpha_n^{(r)}$ approaches a limit $\alpha_r > 0$, $\beta_n^{(r)}$ approaches a limit β_r , and $X \stackrel{d}{=} (\alpha_r)^{-1}(Z_1 + \cdots + Z_r - \beta_r)$. Thus X is stable. ■

In the above proof, we must have $\alpha_{rs} = \alpha_r \alpha_s$ for all positive integers r and s . For

$$\alpha_n^{(rs)} = \frac{a_{rsn}}{a_n} = \frac{a_{rn}}{a_n} \frac{a_{rsn}}{a_{rn}} = \alpha_n^{(r)} \alpha_n^{(s)},$$

and we may let $n \rightarrow \infty$ to obtain $\alpha_{rs} = \alpha_r \alpha_s$.

8.4.3 Examples. It can be shown [see, for example, Breiman (1968, p. 204)] that X is stable iff its characteristic function h can be expressed as $h = e^g$, where g has one of the following two forms:

$$g(u) = iu\beta - d|u|^\alpha \left(1 + i\theta \frac{u}{|u|} \tan \frac{\pi}{2} \alpha \right), \quad (1)$$

$$0 < \alpha < 1 \quad \text{or} \quad 1 < \alpha \leq 2, \quad \beta \in R, \quad d \geq 0, \quad |\theta| \leq 1, \quad (1)$$

$$g(u) = iu\beta - d|u| \left(1 + i\theta \frac{u}{|u|} \frac{2}{\pi} \ln |u| \right), \quad (2)$$

with β, d, θ as in (1). (Take $u/|u| = 0$ when $u = 0$.)

We shall prove only that a random variable with such a characteristic function must be stable. If X_1, \dots, X_n are iid, with each X_i having characteristic function $h = e^g$, with g of the form (1) or (2), let $\lambda = 1/\alpha$ in case (1), $\lambda = 1$ in case (2). Then in case (1),

$$g(n^\lambda u) = iun^\lambda \beta - dn|u|^\alpha \left[1 + i\theta \frac{u}{|u|} \tan \frac{\pi}{2} \alpha \right] = ng(u) - iu\beta(n - n^\lambda),$$

and in case (2),

$$\begin{aligned} g(n^\lambda u) &= g(nu) = iun\beta - dn|u| \left[1 + i\theta \frac{u}{|u|} \frac{2}{\pi} (\ln n + \ln |u|) \right] \\ &= ng(u) - iud\theta \frac{2}{\pi} n \ln n. \end{aligned}$$

Thus in either case,

$$[h(u)]^n = \exp[ng(u)] = \exp[g(n^\lambda u)] \exp[i b_n u] = h(n^\lambda u) \exp[i b_n u],$$

where $b_n = \beta(n - n^\lambda)$ in case (1), and $b_n = d\theta(2/\pi)n \ln n$ in case (2). Therefore $S_n \xrightarrow{d} a_n X + b_n$, with $a_n = n^\lambda = n^{1/\alpha}$; in particular, X is stable.

Now X has a symmetric distribution ($P\{X \in B\} = P\{-X \in B\}$ for all $B \in \mathcal{B}(R)$) iff h is real-valued [see 8.1.5(d)], so the general form of the symmetric stable characteristic function is

$$h(u) = \exp[-d|u|^\alpha], \quad d \geq 0, \quad 0 < \alpha \leq 2. \quad (3)$$

When $d = 0$, we have $X \equiv 0$. [Similarly, if $d = 0$ in (1) or (2), then $X \equiv \beta$.] Thus assume $d > 0$. When $\alpha = 2$, X is normal $(0, 2d)$; when $\alpha = 1$, X has the Cauchy density with parameter d .

If X is stable (not necessarily symmetric) and $0 < \alpha \leq 1$, then h is not differentiable at $u = 0$, so by 8.1.5(e), $E(|X|) = \infty$. In the symmetric case, $E(X)$ does not exist. For $X \xrightarrow{d} -X$; hence $X^+ \xrightarrow{d} (-X)^+ = X^-$, and therefore $E(X^+) = E(X^-)$, necessarily infinite. If $1 < \alpha < 2$, h can be differentiated once but not twice at $u = 0$, so that $E(X^2) = \infty$. This is to be expected, for if X has finite mean and variance, the fact that X can be obtained as a limit of a sequence of normed sums implies that X must be normal (see the opening paragraph of this section).

It can be shown [see Feller (1966, p. 215)] that if X is stable, X has a finite r th moment for all $r \in (0, \alpha)$.

Problem

- The following problem shows that the functions $h(u) = \exp[-d|u|^\alpha]$, $d \geq 0$, $0 < \alpha \leq 2$, are characteristic functions (see 8.4.3). Let X_1, \dots, X_n be independent random variables, each uniformly distributed between $-n$ and $+n$. Define

$$Y_n = k \sum_{i=1}^n \frac{\operatorname{sgn} X_i}{|X_i|^r}, \quad k > 0, \quad r > \frac{1}{2}.$$

If the X_i are the positions of masses distributed at random on $[-n, n]$, then Y_n is the gravitational force at the origin, assuming an inverse r th power law.

- (a) Show that the characteristic function of Y_n is

$$h_n(u) = \left(1 - \frac{1}{n} \left[\int_0^\infty [1 - \cos(kux^{-r})] dx - g(n) \right] \right)^n,$$

where $g(n) \rightarrow 0$ as $n \rightarrow \infty$.

- (b) Show that $h_n(u) \rightarrow h(u) = \exp(-\int_0^\infty [1 - \cos(kux^{-r})] dx)$ as $n \rightarrow \infty$.
(c) Make the change of variable $y = |u|^{1/r} k^{1/r} x^{-1}$ to show that $h(u)$ is of the form $\exp[-d|u|^\alpha]$, $d > 0$, $0 < \alpha < 2$. (The case $\alpha = 2$ corresponds to a normal distribution, and $d = 0$ to a degenerate distribution, so these characteristic functions are automatically realizable.)

8.5 Infinitely Divisible Distributions

There are limit laws that do not fit into any of the categories we have considered so far. For example, let T_n be the number of successes in n Bernoulli trials, with probability p_n of success on a given trial. Then T_n has the binomial distribution:

$$P\{T_n = k\} = \binom{n}{k} p_n^k (1 - p_n)^{n-k}, \quad k = 0, 1, \dots, n.$$

If we let $n \rightarrow \infty$, $p_n \rightarrow 0$, with $np_n \rightarrow \lambda$, then

$$P\{T_n = k\} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots$$

[see Ash (1970, p. 95), for details]. A discrete random variable X with $P\{X = k\} = e^{-\lambda} \lambda^k / k!$, $k = 0, 1, \dots$, is said to have the *Poisson* distribution. In this case,

$$F_{T_n}(k) = P\{T_n \leq k\} = \sum_{j=0}^k P\{T_n = j\} \rightarrow \sum_{j=0}^k P\{X = j\} = P\{X \leq k\} = F_X(k);$$

hence $T_n \xrightarrow{d} X$.

Now this can be regarded as a limit law for sums of independent random variables. We may represent T_n as $X_{n1} + X_{n2} + \dots + X_{nn}$, where X_{ni} , the number of successes on trial i , or equivalently, the indicator of the event {successes on trial i }, is 1 with probability p_n and 0 with probability $1 - p_n$, and the X_{ni} are independent. The difference between this case and the previous ones is that we are no longer dealing with a *single sequence* of random variables; T_n is not simply $X_1 + \dots + X_n$, where X_1, X_2, \dots are independent. Instead, for each n we have a different sequence X_{n1}, \dots, X_{nn} .

We may construct a model that includes this case as well as all previous results, as follows. Consider a *triangular array*:

$$\begin{bmatrix} X_{11} & & & & \\ X_{21} & X_{22} & & & \\ X_{31} & X_{32} & X_{33} & & \\ \vdots & & & & \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nn} \\ \vdots & & & & \end{bmatrix}.$$

We assume that for each n , X_{n1}, \dots, X_{nn} are independent. (We say nothing as yet about any relation between rows.) We set $T_n = X_{n1} + \dots + X_{nn}$; we want to investigate convergence in distribution of the sequence $\{T_n\}$.

Notice that if we are interested in sequences $\{a_n^{-1}(S_n - b_n)\}$, where S_n is the sum of independent random variables X_1, \dots, X_n , we may construct an appropriate triangular array; take

$$X_{ni} = \frac{X_i}{a_n} - \frac{b_n}{na_n};$$

then

$$T_n = \sum_{i=1}^n X_{ni} = \frac{S_n - b_n}{a_n}.$$

Thus the triangular array scheme includes the previous models we have considered.

Note also that the Lindeberg theorem holds for triangular arrays. If the X_{nk} have finite mean m_{nk} and finite variance σ_{nk}^2 ,

$$c_n^{-2} = \text{Var}\left(\sum_{k=1}^n X_{nk}\right) = \sum_{k=1}^n \sigma_{nk}^2,$$

and for every $\varepsilon > 0$,

$$\frac{1}{c_n^{-2}} \sum_{k=1}^n \int_{\{x: |x - m_{nk}| \geq \varepsilon c_n\}} (x - m_{nk})^2 dF_{nk}(x) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

then

$$c_n^{-1} \sum_{k=1}^n (X_{nk} - m_{nk}) \xrightarrow{d} X^* \quad \text{normal } (0, 1).$$

The proof is the same as in 8.3.1, with the distribution function F_k replaced by F_{nk} .

A natural question is the characterization of the possible limiting distributions of a triangular array; this problem was solved for normed sums in Section 8.4. However, as it stands, the question is not sensible, even if we

require that the triangular array come from a single sequence of random variables. For let X be an arbitrary random variable, and take $X_1 = X$, $X_n = 0$ for $n \geq 2$, $b_n \equiv 0$, $a_n \equiv 1$. Then $a_n^{-1}(S_n - b_n) \equiv X$, so any limit distribution is possible. Thus some restriction must be imposed.

One way to take care of this difficulty is to assume the hypothesis of uniform asymptotic negligibility, as we did in considering the converse of the Lindeberg theorem:

$$\max_{1 \leq i \leq n} P\{|X_{ni}| \geq \varepsilon\} \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{for every } \varepsilon > 0.$$

However, we are going to sacrifice generality for simplicity, and assume that for each n , $X_{n1}, X_{n2}, \dots, X_{nn}$ are identically distributed. We may then characterize the possible limiting distributions.

8.5.1 Definition. A random variable X (or its distribution function F , or its characteristic function h) is said to be *infinitely divisible* iff for each n , X has the same distribution as the sum of n independent, identically distributed random variables. In other words, for each n , we may write $h = (h_n)^n$, where h_n is the characteristic function of a random variable.

8.5.2 Theorem. The random variable X is infinitely divisible iff there is a triangular array, with X_{n1}, \dots, X_{nn} iid for each n , such that $T_n = \sum_{k=1}^n X_{nk} \xrightarrow{d} X$.

PROOF. Let X be infinitely divisible. For each n , we may write $X = X_{n1} + \dots + X_{nn}$, where the X_{ni} are iid. Then $T_n = \sum_{i=1}^n X_{ni} \equiv X$; hence $T_n \xrightarrow{d} X$.

The converse is another application of Prokhorov's weak compactness theorem. Assume we have a triangular array with the X_{nk} iid for each n and $T_n \xrightarrow{d} X$. Fix the positive integer r ; then

$$T_{rn} = Z_n^{(1)} + \dots + Z_n^{(r)},$$

where

$$\begin{aligned} Z_n^{(1)} &= X_{rn,1} + \dots + X_{rn,n}, \\ Z_n^{(2)} &= X_{rn,n+1} + \dots + X_{rn,2n}, \\ &\vdots \\ Z_n^{(r)} &= X_{rn,(r-1)n+1} + \dots + X_{rn,rn}. \end{aligned}$$

Since $T_{rn} \xrightarrow{d} X$ as $n \rightarrow \infty$, it follows that $\{T_{rn}, n = 1, 2, \dots\}$ is relatively compact. (This means that the associated sequence of distribution functions is relatively compact.) By 8.2.4, $\{T_{rn}, n = 1, 2, \dots\}$ is tight. But

$$\begin{aligned} (P\{Z_n^{(1)} > z\})^r &= P\{Z_n^{(1)} > z, \dots, Z_n^{(r)} > z\} \\ &\quad \text{by independence of the } Z_n^{(i)} \\ &\leq P\{T_{rn} > rz\} \end{aligned}$$

and similarly,

$$(P\{Z_n^{(1)} < -z\})^r \leq P\{T_{rn} < -rz\}.$$

It follows that $\{Z_n^{(1)}, n = 1, 2, \dots\}$ is tight, and hence relatively compact by 8.2.4. Thus we have a subsequence $\{Z_n^{(1)}, n = n_1, n_2, \dots\}$ converging in distribution to a random variable Y . But the $Z_n^{(i)}$, $i = 1, \dots, r$, are iid; hence $\{Z_n^{(i)}, n = n_1, n_2, \dots\} \xrightarrow{d} Y$. By 8.2.9, $T_{rn} \xrightarrow{d} Y_1 + \dots + Y_r$, where Y_1, \dots, Y_r are iid with $Y_i \xrightarrow{d} Y$. But $T_{rn} \xrightarrow{d} X$; hence $X \xrightarrow{d} Y_1 + \dots + Y_r$. ■

It can be shown (Gnedenko and Kolmogorov, 1954) that Theorem 8.5.1 still holds if the condition that for each n , the X_{ni} have the same distribution, is replaced by the uan condition.

8.5.3 Examples of Infinitely Divisible Random Variables. (a) Every *stable* random variable is infinitely divisible. This may be seen from the fact that every stable X is a limit in distribution of a sequence of normed sums, hence a limit of row sums of a triangular array in which the X_{ni} , $i = 1, 2, \dots, n$, have the same distribution. Alternatively, if $X_1 + \dots + X_n \xrightarrow{d} a_n X + b_n$, then

$$X \xrightarrow{d} \sum_{i=1}^n \left(\frac{X_i}{a_n} - \frac{b_n}{na_n} \right).$$

(b) A random variable of the *Poisson type* is infinitely divisible. Let Y have the Poisson distribution: $P\{Y = k\} = e^{-\lambda} \lambda^k / k!$, $k = 0, 1, \dots$. The characteristic function of Y is

$$h(u) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{iu})^k}{k!} = \exp[\lambda(e^{iu} - 1)]$$

and it follows that if Y_1, \dots, Y_n are independent, with Y_i Poisson with parameter λ_i , $i = 1, \dots, n$, then $Y_1 + \dots + Y_n$ is Poisson with parameter $\lambda_1 + \dots + \lambda_n$. In particular, if Y is Poisson with parameter λ , then $Y \xrightarrow{d} Y_1 + \dots + Y_n$, where the Y_i are iid, each Poisson with parameter λ/n . Thus Y is infinitely divisible. Now if Y is infinitely divisible, so is $aY + b$ (a similar statement holds for stable random variables); hence a random variable of the Poisson type ($aY + b$, Y Poisson, $a \neq 0$) is infinitely divisible. The characteristic function of $aY + b$ is $\exp[ibu + \lambda(e^{iau} - 1)]$.

(c) A random variable with the *gamma distribution* is infinitely divisible. Let X have density

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\alpha, \beta > 0$. The characteristic function of X is

$$h(u) = (1 - i\beta u)^{-\alpha} = [(1 - i\beta u)^{-\alpha/n}]^n;$$

hence X is the sum of n independent gamma-distributed random variables with parameters α/n and β .

We now develop some general properties of infinitely divisible distributions.

8.5.4 Theorem. If h_1 and h_2 are infinitely divisible characteristic functions, so is $h_1 h_2$. If h is infinitely divisible, then \bar{h} , the complex conjugate of h , and $|h|^2$ are infinitely divisible.

PROOF. If $h_i = (h_{in})^n$, $i = 1, 2$, then $h_1 h_2 = (h_{1n} h_{2n})^n$; since $h_{1n} h_{2n}$ is the characteristic function of the sum of two independent random variables with characteristic functions h_{1n} and h_{2n} , the first assertion is proved. If X has characteristic function h , then $-X$ has characteristic function \bar{h} [see 8.1.5(c)]; thus if $h = (h_n)^n$, then $\bar{h} = (\bar{h}_n)^n$; hence \bar{h} is infinitely divisible. Since $|h|^2 = h\bar{h}$, $|h|^2$ is also infinitely divisible. ■

If \mathcal{C} is the entire class of characteristic functions of random variables, the proof of 8.5.4 shows that if $h_1, h_2 \in \mathcal{C}$; then $h_1 h_2 \in \mathcal{C}$.

Also, if $h \in \mathcal{C}$, then $\bar{h} \in \mathcal{C}$ and $|h|^2 \in \mathcal{C}$. Furthermore, 8.2.8 implies that if $h_n \in \mathcal{C}$, $n = 1, 2, \dots$, and $h_n(u) \rightarrow h(u)$ for all u , where h is continuous at the origin, then $h \in \mathcal{C}$. A similar result holds for infinitely divisible characteristic functions.

8.5.5 Theorem. If h_n is an infinitely divisible characteristic function for each $n = 1, 2, \dots$, and $h_n(u) \rightarrow h(u)$ for all u , where h is a characteristic function, then h is infinitely divisible.

PROOF. Let Z_n be a random variable with characteristic function h_n , $n = 1, 2, \dots$. If r is a fixed positive integer, then $Z_n \stackrel{d}{=} Z_n^{(1)} + \dots + Z_n^{(r)}$, where the $Z_n^{(i)}$, $i = 1, \dots, r$, are iid. If Z is a random variable with characteristic function h , then $Z_n \stackrel{d}{\rightarrow} Z$ by 8.2.9, so that $\{Z_n\}$ is relatively compact, and hence tight by 8.2.4. Just as in the proof of 8.5.2., it follows that $\{Z_n^{(1)}\}$ is tight. By 8.2.4, we have a subsequence $\{Z_n^{(1)}, n = n_1, n_2, \dots\}$ converging in distribution to a random variable Y ; hence (again as in 8.5.2) $Z_n \stackrel{d}{\rightarrow} Y_1 + \dots + Y_r$, where

Y_1, \dots, Y_r are iid with $Y_i \stackrel{d}{=} Y$. But $Z_n \xrightarrow{d} Z$; hence $Z \stackrel{d}{=} Y_1 + \dots + Y_r$. In other words, h is infinitely divisible. ■

Now if h is infinitely divisible, a uniqueness question arises; namely, can h be represented in two different ways as the n th power of a characteristic function? This is actually an exercise in complex variables, as follows. Let f and g be continuous complex-valued functions on the connected set S , with $f^n = g^n$; assume that $f(u) = g(u)$ for at least one $u \in S$. [In our case $S = \mathbb{R}$ and $f^n = g^n = h$; since f and g are characteristic functions of random variables, $f(0) = g(0) = 1$.] If f and g are never 0 on S , then $f \equiv g$. For since $(f/g)^n = 1$, f/g is a continuous map of S into $\{\exp(i2\pi k/n), k = 0, 1, \dots, n-1\}$. Since the image of S under f/g is connected, it must consist of a single point; thus f/g is a constant, necessarily 1 because f and g agree at one point. Thus the representation of h as the n th power of a characteristic function is unique, provided we can establish that an infinitely divisible characteristic function never vanishes.

8.5.6 Theorem. If h is an infinitely divisible characteristic function, then h is never 0.

PROOF. If $h = (h_n)^n$, then $|h|^2 = |h_n|^{2n}$. Since $|h|^2$ is infinitely divisible by 8.5.4, we may as well assume that h and the h_n are real and nonnegative. Thus $h_n = h^{1/n} (= \exp[(1/n)\ln h])$, so if $h(u) > 0$, then $h_n(u) \rightarrow 1$, and if $h(u) = 0$, then $h_n(u) = 0$ for all n . But $h(0) = 1$; hence $h(u) > 0$ in some neighborhood of the origin. Thus h_n converges to a function g that is 1 in a neighborhood of the origin. By 8.2.8, g is a characteristic function, and hence continuous everywhere. But g takes on only the values 0 and 1, and hence $g \equiv 1$. Thus for any u , $h_n(u) \rightarrow 1$, so that $h_n(u) \neq 0$ for sufficiently large n . Therefore $h(u) = [h_n(u)]^n \neq 0$. ■

Example 8.5.3(b) is basic in the sense that random variables of the Poisson type can be used as building blocks for arbitrary infinitely divisible random variables.

8.5.7 Theorem. The random variable X is infinitely divisible iff there is a sequence of sums $\sum_{k=1}^{r(n)} X_{nk} \xrightarrow{d} X$, where for each n , the X_{nk} are independent (not necessarily identically distributed) and each X_{nk} is of the Poisson type.

PROOF. The “if” part follows from 8.5.3(b), the first assertion of 8.5.4, and 8.5.5, so assume X infinitely divisible. Since h , the characteristic function of X ,

is continuous and never 0 (by 8.5.6), h has a continuous logarithm, to be denoted by $\log h$. If we specify that $\log h(0) = \log 1 = 0$, the logarithm is determined uniquely. [See Ash (1971, p. 49ff.) for a discussion of continuous logarithms.] If $h = (h_n)^n$, where h_n is a characteristic function, then

$$(h_n)^n = \left[\exp\left(\frac{1}{n} \log h\right) \right]^n;$$

hence as in the discussion before 8.5.6,

$$h_n = \exp\left(\frac{1}{n} \log h\right),$$

so for any fixed $u \in R$,

$$\begin{aligned} n(h_n(u) - 1) &= n \left(\exp\left[\frac{1}{n} \log h(u)\right] - 1 \right) \\ &= n\left(\frac{1}{n} \log h(u) + o\left(\frac{1}{n}\right)\right) \quad \text{since } e^z = 1 + z + o(z) \\ &\rightarrow \log h(u) \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus

$$\log h(u) = \lim_{n \rightarrow \infty} n(h_n(u) - 1) = \lim_{n \rightarrow \infty} n \int_{-\infty}^{\infty} (e^{iux} - 1) dF_n(x),$$

where F_n is the distribution function corresponding to h_n . It follows from the dominated convergence theorem that for each n we may select a positive number $m = m(n)$ such that $m \rightarrow \infty$ as $n \rightarrow \infty$ and

$$\left| \int_{-\infty}^{\infty} (e^{iux} - 1) dF_n(x) - \int_{-m}^m (e^{iux} - 1) dF_n(x) \right| \leq \frac{1}{n^2} \quad \text{for all } u,$$

and we may then choose a positive integer $r = r(n)$ such that

$$\left| \int_{-m}^m (e^{iux} - 1) dF_n(x) - \sum_{k=1}^r (e^{iux_k} - 1)[F_n(x_k) - F_n(x_{k-1})] \right| \leq \frac{1}{n^2}$$

for all $u \in (-m, m)$, where $x_k = -m + 2mk/r$, $k = 0, 1, \dots, r$.

It follows that we may obtain $h(u)$ as a pointwise limit of terms of the form

$$\prod_{k=1}^{r(n)} \exp[\lambda_{nk}(\exp(i\alpha_{nk} u) - 1)],$$

where $\lambda_{nk} = n[F_n(x_k) - F_n(x_{k-1})]$ and $\alpha_{nk} = x_k$. ■

We conclude this section by mentioning the *Lévy–Khintchine representation*: The characteristic function h is infinitely divisible iff

$$\log h(u) = iu\beta - \frac{u^2\sigma^2}{2} + \int_R \left(e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{1+x^2}{x^2} d\lambda(x),$$

where $\beta \in R$, $\sigma^2 \geq 0$, and λ is a finite measure on $\mathcal{B}(R)$ such that $\lambda\{0\} = 0$.

The result is basic for a deeper study of the central limit theorem, in particular for deriving conditions for convergence to a particular infinitely divisible distribution, analogous to the results on normal convergence in Section 8.3. Full details are given by Gnedenko and Kolmogorov (1954). Proofs of the Lévy–Khintchine representation are also given by Chung (1968) and Tucker (1967).

Problems

1. The random variable X is said to have the *geometric distribution* iff $P\{X = k\} = q^{k-1}p$, $k = 1, 2, \dots$, where $0 < p < 1$, $q = 1 - p$. Show that the associated characteristic function, given by

$$h(u) = pe^{iu} \sum_{k=1}^{\infty} (qe^{iu})^{k-1} = \frac{pe^{iu}}{1 - qe^{iu}},$$

is infinitely divisible (use 8.5.7).

2. Let $g(s) = \sum_{n=1}^{\infty} n^{-s}$, $\text{Re } s > 1$, be the Riemann zeta function. The series converges uniformly for $\text{Re } s \geq 1 + \varepsilon$, any $\varepsilon > 0$; also,

$$g(s) = \prod_{k=1}^{\infty} \frac{1}{1 - p_k^{-s}},$$

where p_n is the n th prime. [See Ash (1971, Chapter 6) for details.] If c is a fixed real number greater than 1, show that $h(u) = g(c + iu)/g(c)$ is an infinitely divisible characteristic function (use 8.5.7).

3. Give an example of an infinitely divisible characteristic function that is not stable.
4. When characteristic functions are not easy to compute, the following technique is sometimes useful for actually finding the distribution of a sum of independent random variables.

Let X and Y be independent random variables, and let $Z = X + Y$. If X , Y , and Z have distribution functions F_1 , F_2 , and F_3 , show that F_3 is the *convolution* of F_1 and F_2 (notation: $F_3 = F_1 * F_2$), that is,

$$F_3(z) = \int_{-\infty}^{\infty} F_1(z - y) dF_2(y).$$

F_3 is also the convolution of F_2 and F_1 , that is,

$$F_3(z) = \int_{-\infty}^{\infty} F_2(z - x) dF_1(x).$$

If X [respectively Y] has density f_1 [respectively f_2], show that Z has density f_3 , where

$$f_3(z) = \int_{-\infty}^{\infty} f_1(z - y) dF_2(y) = \int_{-\infty}^{\infty} f_2(z - x) dF_1(x).$$

[If both X and Y have densities, replace $dF_1(x)$ by $f_1(x) dx$ and $dF_2(y)$ by $f_2(y) dy$.]

Intuitively, the probability that X falls in $(x, x + dx)$ is $dF_1(x)$; given that $X = x$, we have $Z \leq z$ iff $Y \leq z - x$, and this happens with probability $F_2(z - x)$. Integrate over x to obtain the total probability that $Z \leq z$, namely, $F_3 = F_2 * F_1$. The other formulas have a similar interpretation.

Note also that convolution is associative, that is, $F_1 * (F_2 * F_3) = (F_1 * F_2) * F_3$. This is somewhat messy to prove directly, but a probabilistic interpretation makes it transparent. For if X_1 , X_2 , and X_3 are independent random variables with distribution functions F_1 , F_2 , and F_3 , respectively, then $F_1 * (F_2 * F_3)$ is the distribution function of $X_1 + (X_2 + X_3)$, and $(F_1 * F_2) * F_3$ is the distribution function of $(X_1 + X_2) + X_3$.

Finally, we note that if F is the distribution function of a random variable, then F is infinitely divisible iff for each n there is a distribution function F_n (of a random variable) such that $F = F_n * F_n * \cdots * F_n$ (n times).

5. If $\lambda \geq 0$ and f is the characteristic function of a random variable, show that $\exp[\lambda(f - 1)]$ is an infinitely divisible characteristic function.

8.6 Uniform Convergence in the Central Limit Theorem

Let X_1, X_2, \dots be independent random variables with finite mean and variance, and suppose that $T_n = c_n^{-1}(S_n - E(S_n)) \xrightarrow{d} X^*$ normal $(0, 1)$. Very often the statement is made that “for large n , S_n is approximately normal with mean $a_n = E(S_n)$ and variance c_n^2 ,” that is,

$$F_{S_n}(x) - \int_{-\infty}^x \frac{1}{\sqrt{2\pi c_n}} \exp\left[-\frac{(t - a_n)^2}{2c_n^2}\right] dt \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Let us try to prove this. If X is normal (a_n, c_n^2) and X^* is normal $(0, 1)$, then

$$\begin{aligned} |P\{S_n \leq x\} - P\{X \leq x\}| &= \left| P\left\{T_n \leq \frac{x - a_n}{c_n}\right\} - P\left\{X^* \leq \frac{x - a_n}{c_n}\right\} \right| \\ &= \left| F_{T_n}\left(\frac{x - a_n}{c_n}\right) - F^*\left(\frac{x - a_n}{c_n}\right) \right| \end{aligned}$$

and this will approach 0 as $n \rightarrow \infty$ if $F_{T_n} \rightarrow F^*$ uniformly on R . In fact this does happen; the proof rests on the following two results.

8.6.1 Theorem. Let F, F_1, F_2, \dots be bounded distribution functions on R , S a dense subset of R containing all the discontinuity points of F .

If

$$\begin{aligned} F_n(\infty) &\rightarrow F(\infty), \\ F_n(-\infty) &\rightarrow F(-\infty), \\ F_n(x) &\rightarrow F(x) \quad \text{for all } x \in S, \\ F_n(x^-) &\rightarrow F(x^-) \quad \text{for all } x \in S, \end{aligned}$$

then $F_n \rightarrow F$ uniformly on R .

PROOF. Let $\varepsilon > 0$ be given. We wish to obtain a partition $-\infty = y_0 < y_1 < \dots < y_m = \infty$ with

$$y_j \in S, \quad 1 \leq j \leq m-1,$$

$$F(y_{j+1}^-) < F(y_j) + \frac{\varepsilon}{2}, \quad 0 \leq j \leq m-1,$$

$$F(y_{j+1}) \geq F(y_j) + \frac{\varepsilon}{3}, \quad 0 \leq j \leq m-2$$

(take $y_m^- = \infty$).

Set $y_0 = -\infty$ and define $z_1 = \sup\{x > y_0 : F(x) - F(y_0) \leq \varepsilon/3\}$. If $z_1 < \infty$, then $F(x) - F(y_0) \leq \varepsilon/3$ for $y_0 < x < z_1$; hence $F(z_1^-) \leq F(y_0) + (\varepsilon/3) < F(y_0) + (\varepsilon/2)$. Also, $F(z_1) \geq F(y_0) + (\varepsilon/3)$; for if not, $F(z_1') < F(y_0) + (\varepsilon/3)$ for some $z_1' > z_1$ (by right-continuity), contradicting the definition of z_1 .

Now if $F(z_1^-) < F(z_1)$, then $z_1 \in S$ by hypothesis, and we set $y_1 = z_1$. If $F(z_1^-) = F(z_1)$, then since S is dense and F is right-continuous, we can find $y_1 \in S$ such that $y_1 > z_1$ and $F(y_1) < F(y_0) + (\varepsilon/2)$. Thus in either case we obtain $y_1 \in S$ such that $F(y_1^-) < F(y_0) + (\varepsilon/2)$ and $F(y_1) \geq F(y_0) + (\varepsilon/3)$. Continue by defining $z_2 = \sup\{x > y_1 : F(x) - F(y_1) \leq \varepsilon/3\}$ and proceed as above. Since $F(y_{j+1}) \geq F(y_j) + (\varepsilon/3)$ and F is bounded, the process will terminate in a finite number of steps and produce the desired partition.

Let $x \in R$; say $y_j \leq x < y_{j+1}$. Then

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(y_{j+1}^-) - F(y_j) \\ &< F(y_{j+1}^-) + \frac{\varepsilon}{2} - F(y_j) \\ &\quad \text{for large } n, \text{ since } y_{j+1} \in S \quad \text{or} \quad y_{j+1} = \infty \\ &< \varepsilon. \end{aligned}$$

Also,

$$\begin{aligned} F(x) - F_n(x) &\leq F(y_{j+1}^-) - F_n(y_j) \\ &\leq F(y_{j+1}^-) - F(y_j) + \frac{\varepsilon}{2} \\ &\quad \text{for large } n \text{ since } y_j \in S \quad \text{or} \quad y_j = -\infty \\ &< \varepsilon. \end{aligned}$$

The “large n ” depends only on j and not on x , and it follows that $F_n \rightarrow F$ uniformly on R . ■

8.6.2 Theorem. Let F, F_1, F_2, \dots be bounded distribution functions on R . Assume that $F_n(-\infty) = 0$ for all n , $F(-\infty) = 0$, F is continuous everywhere, and F_n converges weakly to F . Then F_n converges to F uniformly on R .

PROOF. In 8.6.1, take $S = R$. Since F_n converges weakly to F and F is continuous everywhere, $F_n(\infty) \rightarrow F(\infty)$ and $F_n(x) \rightarrow F(x)$ for all $x \in R$. Since $F_n(-\infty) \rightarrow F(-\infty)$ by assumption, it remains to show that $F_n(x^-) \rightarrow F(x^-)$ for all x . If $y < x$,

$$F(y) = \lim_n F_n(y) \leq \liminf_n F_n(x^-) \leq \limsup_n F_n(x^-) \leq \lim_n F_n(x) = F(x).$$

If $y \rightarrow x$, then $F(y) \rightarrow F(x)$ by continuity of F ; hence $F_n(x^-) \rightarrow F(x) = F(x^-)$. ■

Problem

1. (*Glivenko–Cantelli Theorem*) Let X_1, X_2, \dots be independent, identically distributed random variables with common distribution function F . Let $F_n(x, \omega)$, $x \in R$, $\omega \in \Omega$, be the *empirical distribution function* of the X_i , based on n trials, that is,

$$\begin{aligned} F_n(x, \omega) &= \frac{1}{n} \left[\begin{array}{l} \text{the number of terms among } X_1(\omega), \dots, X_n(\omega) \\ \text{that are less than or equal to } x \end{array} \right] \\ &= \frac{1}{n} \sum_{k=1}^n I_{\{X_k \leq x\}}(\omega). \end{aligned}$$

For example, if $n = 3$, $X_1(\omega) = 2$, $X_2(\omega) = \frac{1}{2}$, $X_3(\omega) = 7$, then

$$F_3(x, \omega) = \begin{cases} 0, & x < \frac{1}{2} \\ \frac{1}{3}, & \frac{1}{2} \leq x < 2 \\ \frac{2}{3}, & 2 < x \leq 7 \\ 1, & x \geq 7. \end{cases}$$

Intuitively, for large n , $F_n(x, \omega)$ should approximate $F(x)$. Show that there is a set A of probability 0 such that for $\omega \notin A$, $F_n(x, \omega) \rightarrow F(x)$ uniformly for $x \in R$.

8.7 Proof of the Inversion Formula (8.1.4)

Intuitively, if $h(u) = \int_{-\infty}^{\infty} e^{iux} dF(x)$, then $h(u) = \int_{-\infty}^{\infty} e^{iux} F'(x) dx$, so

$$F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(u) e^{-iux} du.$$

Thus

$$F(a, b] = \int_a^b F'(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(u) \left[\int_a^b e^{-iux} dx \right] du$$

and since

$$\int_a^b e^{-iux} dx = \frac{e^{-iua} - e^{-iub}}{iu},$$

this leads to the inversion formula. For a formal proof, let

$$I_c = \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iua} - e^{-iub}}{iu} h(u) du, \quad a < b.$$

Then

$$I_c = \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iua} - e^{-iub}}{iu} \left[\int_{-\infty}^{\infty} e^{iux} dF(x) \right] du.$$

Now

$$\left| \frac{(e^{-iua} - e^{-iub})}{iu} e^{iux} \right| = \left| \frac{e^{-iua} - e^{-iub}}{iu} \right| = \left| \int_a^b e^{-iux} dx \right| \leq b - a$$

and

$$\int_{-c}^c \left[\int_{-\infty}^{\infty} (b - a) dF(x) \right] du = 2c(b - a)[F(\infty) - F(-\infty)] < \infty.$$

By Fubini's theorem, the order of integration may be interchanged to obtain

$$I_c = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-c}^c \frac{(e^{-iu a} - e^{-iu b})}{iu} e^{iux} du \right] dF(x) = \int_{-\infty}^{\infty} J_c(x) dF(x),$$

where

$$J_c(x) = \frac{1}{2\pi} \int_{-c}^c \frac{\sin u(x-a) - \sin u(x-b)}{u} du.$$

[Note that

$$\int_{-c}^c \frac{\cos u(x-a) - \cos u(x-b)}{iu} du = 0$$

since the integrand is an odd function.] Let $v = u(x-a)$ and $w = u(x-b)$ to obtain

$$J_c(x) = \frac{1}{2\pi} \int_{-c(x-a)}^{c(x-a)} \frac{\sin v}{v} dv - \frac{1}{2\pi} \int_{-c(x-b)}^{c(x-b)} \frac{\sin w}{w} dw.$$

Now

$$\int_r^s \frac{\sin v}{v} dv \rightarrow \pi \quad \text{as } s \rightarrow \infty \quad \text{and } r \rightarrow -\infty,$$

and since the integral is continuous in r and s , it is bounded uniformly in r and s . Thus for some $M < \infty$, $|J_c(x)| \leq M$ for all c and x ; furthermore, as $c \rightarrow \infty$, $J_c(x) \rightarrow J(x)$, where

$$J(x) = \begin{cases} 0 & \text{if } x < a \text{ or } x > b, \\ 1 & \text{if } a < x < b, \\ \frac{1}{2} & \text{if } x = a \text{ or } x = b. \end{cases}$$

By the dominated convergence theorem,

$$\begin{aligned} \lim_{c \rightarrow \infty} I_c &= \int_{-\infty}^{\infty} J(x) dF(x) \\ &= F(b^-) - F(a) + \frac{1}{2} [F(a) - F(a^-) + F(b) - F(b^-)] \\ &= \frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2} = F(b) - F(a) \end{aligned}$$

if F is continuous at a and b . This proves the formula.

Now assume h integrable on $(-\infty, \infty)$. Let

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} h(u) du, \quad -\infty < x < \infty.$$

Since h is integrable, f is well-defined; furthermore, f is continuous by the dominated convergence theorem. Now by Fubini's theorem,

$$\begin{aligned} \int_a^b f(x) dx &= \frac{1}{2\pi} \int_{-\infty}^{\infty} h(u) \left[\int_a^b e^{-iux} dx \right] du \\ &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c h(u) \left[\int_a^b e^{-iux} dx \right] du \\ &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu a} - e^{-iu b}}{iu} h(u) du \\ &= F(b) - F(a) \end{aligned}$$

by the inversion formula if a and b are continuity points of F . Thus $F(b) - F(a) = \int_a^b f(x) dx$ at continuity points of F , hence for all a and b since the integral is a continuous function of its limits.

Since f is continuous, it follows that F is differentiable everywhere and its derivative is f . Since F is increasing, f is everywhere nonnegative. Thus f is a density for F . ■

8.8 Completion of the Proof of Theorem 8.3.2

The “if” part of the theorem remains to be proved. We may assume without loss of generality that $EX_k \equiv 0$. The uan condition is equivalent to the statement that $h_k(u/c_n) \rightarrow 1$ as $n \rightarrow \infty$ uniformly in $k = 1, \dots, n$, that is,

$$\max_{1 \leq k \leq n} \left| 1 - h_k \left(\frac{u}{c_n} \right) \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Furthermore, in this case

$$\max_{1 \leq k \leq n} \left| 1 - h_k \left(\frac{u}{c_n} \right) \right| \rightarrow 0$$

uniformly for u in any bounded interval (see Problem 3, Section 8.3).

(a) We show first that

$$\frac{u^2}{2} + \sum_{k=1}^n \left[h_k \left(\frac{u}{c_n} \right) - 1 \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Let h_Y denote the characteristic function of the random variable Y . By the normal convergence,

$$h_{T_n}(u) = \prod_{k=1}^n h_k \left(\frac{u}{c_n} \right) = \prod_{k=1}^n \left[1 + \left(h_k \left(\frac{u}{c_n} \right) - 1 \right) \right] \rightarrow \exp \left[\frac{-u^2}{2} \right]$$

and the convergence is uniform for u in any bounded interval (see 8.2.9). If “Log” denotes the principal branch of the logarithm, then

$$\sum_{k=1}^n \text{Log} \left[1 + \left(h_k \left(\frac{u}{c_n} \right) - 1 \right) \right]$$

is the (necessarily unique) continuous logarithm of $h_{T_n}(u)$ having the value 0 at $u = 0$. Thus

$$\sum_{k=1}^n \text{Log} \left[1 + \left(h_k \left(\frac{u}{c_n} \right) - 1 \right) \right] \rightarrow \frac{-u^2}{2}.$$

Therefore,

$$\frac{u^2}{2} + \sum_{k=1}^n \left[h_k \left(\frac{u}{c_n} \right) - 1 \right] + \sum_{k=1}^n \theta_k \left| h_k \left(\frac{u}{c_n} \right) - 1 \right|^2 \rightarrow 0,$$

where each $|\theta_k|$ is at most 1. But if $a_n + b_n \rightarrow 0$, then $\limsup_{n \rightarrow \infty} |a_n| = \limsup_{n \rightarrow \infty} |b_n|$ (eventually $|a_n| < |b_n| + \varepsilon$, and thus $\limsup_{n \rightarrow \infty} |a_n| \leq \limsup_{n \rightarrow \infty} |b_n|$); hence

$$\begin{aligned} & \limsup_n \left| \frac{u^2}{2} + \sum_{k=1}^n \left[h_k \left(\frac{u}{c_n} \right) - 1 \right] \right| \\ &= \limsup_n \left| \sum_{k=1}^n \theta_k \left| h_k \left(\frac{u}{c_n} \right) - 1 \right|^2 \right| \\ &\leq \limsup_n \max_{1 \leq k \leq n} \left| h_k \left(\frac{u}{c_n} \right) - 1 \right| \sum_{k=1}^n \left| h_k \left(\frac{u}{c_n} \right) - 1 \right|. \end{aligned}$$

Now

$$\max_{1 \leq k \leq n} \left| h_k \left(\frac{u}{c_n} \right) - 1 \right| \rightarrow 0$$

and

$$\begin{aligned} \sum_{k=1}^n \left| h_k \left(\frac{u}{c_n} \right) - 1 \right| &= \sum_{k=1}^n \left| \int_{-\infty}^{\infty} \left[\exp \left(\frac{iux}{c_n} \right) - 1 \right] dF_k(x) \right| \\ &= \sum_{k=1}^n \left| \int_{-\infty}^{\infty} \theta \left(\frac{u^2 x^2}{2c_n^2} \right) dF_k(x) \right| \\ &\leq \frac{u^2}{2c_n^2} \sum_{k=1}^n \sigma_k^2 = \frac{u^2}{2} < \infty. \end{aligned}$$

This proves (a).

(b) For any $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \left[1 - \frac{1}{c_n^2} \sum_{k=1}^n \int_{|x| < \varepsilon c_n} x^2 dF_k(x) \right] \leq \frac{4}{\varepsilon^2 u^2}.$$

For by (a),

$$\frac{u^2}{2} - \sum_{k=1}^n \operatorname{Re} \left[1 - h_k \left(\frac{u}{c_n} \right) \right] \rightarrow 0,$$

that is,

$$\frac{u^2}{2} - \sum_{k=1}^n \int_{-\infty}^{\infty} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \rightarrow 0,$$

or equivalently,

$$\frac{u^2}{2} - \sum_{k=1}^n \int_{|x| < \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) - \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \rightarrow 0.$$

Again, noting that $a_n + b_n \rightarrow 0$ implies $\limsup_{n \rightarrow \infty} |a_n| = \limsup_{n \rightarrow \infty} |b_n|$, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} & \left| \frac{u^2}{2} - \sum_{k=1}^n \int_{|x| < \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \right| \\ &= \limsup_{n \rightarrow \infty} \left| \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \right|. \end{aligned} \quad (1)$$

But

$$\begin{aligned} \sum_{k=1}^n \int_{|x| < \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) &\leq \sum_{k=1}^n \int_{|x| < \varepsilon c_n} \frac{u^2 x^2}{2 c_n^2} dF_k(x) \\ &\leq \frac{u^2}{2 c_n^2} \sum_{k=1}^n \sigma_k^2 = \frac{u^2}{2} \end{aligned}$$

and thus the absolute values around the “ \limsup ” terms in (1) may be dropped. [This also shows that the quantity inside the brackets in (b) is nonnegative.] Consequently,

$$\begin{aligned} \limsup_{n \rightarrow \infty} & \left(\frac{u^2}{2} \left[1 - \frac{1}{c_n^2} \sum_{k=1}^n \int_{|x| < \varepsilon c_n} x^2 dF_k(x) \right] \right) \\ &\leq \limsup_{n \rightarrow \infty} \left[\frac{u^2}{2} - \sum_{k=1}^n \int_{|x| < \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \right] \\ &= \limsup_{n \rightarrow \infty} \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n} \right) dF_k(x) \quad \text{by (1).} \end{aligned}$$

But

$$\begin{aligned}
 \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} \left(1 - \cos \frac{ux}{c_n}\right) dF_k(x) &\leq 2 \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} dF_k(x) \\
 &= 2 \sum_{k=1}^n P\{|X_k| \geq \varepsilon c_n\} \\
 &\leq 2 \sum_{k=1}^n \frac{\sigma_k^2}{\varepsilon^2 c_n^2} \quad \text{by Chebyshev's inequality} \\
 &= \frac{2}{\varepsilon^2}, \quad \text{proving (b).}
 \end{aligned}$$

The proof of the theorem is completed by letting $u \rightarrow \infty$ in (b). ■

8.9 Proof of the Convergence of Types Theorem (8.3.4)

(a) Let G_n, G, F_n, F be the distribution functions of X_n, X, Y_n, Y , respectively.

We may select convergent subsequences $\{a_{n_k}\}, \{b_{n_k}\}$ such that

$$a_{n_k} \rightarrow a, \quad b_{n_k} \rightarrow b, \quad \text{where } 0 \leq a \leq \infty, \quad -\infty \leq b \leq \infty.$$

We first show that $a < \infty$.

Suppose that $a = \infty$. Let $E = \{x \in R : \limsup_{k \rightarrow \infty} (a_{n_k} x + b_{n_k}) < \infty\}$ and let $c = \sup E$. (Take $c = -\infty$ if $E = \emptyset$.)

(i) If $x \in R, x < c$, then $a_{n_k} x + b_{n_k} \rightarrow -\infty$.

PROOF. Let $x < u < c$, with $u \in E$ (u exists since $c = \sup E$). Then $u \in E, x < u$ implies $x \in E$ by definition of E .

Now

$$a_{n_k} x + b_{n_k} = a_{n_k}(x - u) + a_{n_k} u + b_{n_k};$$

$$\limsup_{k \rightarrow \infty} (a_{n_k} u + b_{n_k}) < \infty$$

since $u \in E$, and $a_{n_k}(x - u) \rightarrow -\infty$ since $a_{n_k} \rightarrow a = \infty$ and $x - u < 0$. This proves (i).

(ii) If $x \in R, x < c$, then $G(x) = 0$.

PROOF. Let $\varepsilon > 0$ be arbitrary. Choose $z \in R$ such that $F(z) < \varepsilon$ and z is a continuity point of F . Then $F_{n_k}(z) \rightarrow F(z)$, so eventually $F_{n_k}(z) < \varepsilon$. By (i), eventually $a_{n_k} x + b_{n_k} < z$, so

$$G_{n_k}(x) = F_{n_k}(a_{n_k} x + b_{n_k}) \leq F_{n_k}(z) < \varepsilon$$

for large enough k . Thus $G_{n_k}(x) \rightarrow 0$, $x < c$. Let $x < x' < c$, x' a continuity point of G . Then $G(x') = \lim_{k \rightarrow \infty} G_{n_k}(x') = 0$, and since $G(x) \leq G(x')$, we have $G(x) = 0$, proving (ii).

(iii) If $x \in R$, $x > c$, then $G(x) = 1$.

PROOF. By definition of E , we find a subsequence $\{r_j\}$ such that $a_{r_j}x + b_{r_j} \rightarrow \infty$. Choose a continuity point w of F such that $F(w) > 1 - \varepsilon$. Eventually, $a_{r_j}x + b_{r_j} > w$, so $G_{r_j}(x) = F_{r_j}(a_{r_j}x + b_{r_j}) \geq F_{r_j}(w) > 1 - \varepsilon$. Thus $G_{r_j}(x) \rightarrow 1$ for $x > c$. If $c < y < x$, y a continuity point of G , we have $G(x) \geq G(y) = \lim_{j \rightarrow \infty} G_{r_j}(y) = 1$, proving (iii).

It follows from (ii) and (iii) that G is degenerate, a contradiction.

Next we show that b is finite.

If $b_{n_k} \rightarrow \infty$, then $a_{n_k}x + b_{n_k} \rightarrow \infty$ for every $x \in R$, so the argument of (iii) may be repeated to show $G(x) \equiv 1$, a contradiction. If $b_{n_k} \rightarrow -\infty$, then $a_{n_k}x + b_{n_k} \rightarrow -\infty$ for each $x \in R$, so the argument of (ii) shows $G(x) \equiv 0$, a contradiction.

Now we show that $a > 0$.

Let x be a continuity point of G , and let $\varepsilon_1, \varepsilon_2 > 0$ such that $ax + b + \varepsilon_1$ and $ax + b - \varepsilon_2$ are continuity points of F . Then $a_{n_k}x + b_{n_k} \rightarrow ax + b$, so eventually $ax + b - \varepsilon_2 \leq a_{n_k}x + b_{n_k} \leq ax + b + \varepsilon_1$; hence

$$F_{n_k}(ax + b - \varepsilon_2) \leq F_{n_k}(a_{n_k}x + b_{n_k}) = G_{n_k}(x) \leq F_{n_k}(ax + b + \varepsilon_1).$$

Let $k \rightarrow \infty$ to obtain $F(ax + b - \varepsilon_2) \leq G(x) \leq F(ax + b + \varepsilon_1)$. Since ε_1 and ε_2 may be chosen arbitrarily small, $F(ax + b)^- \leq G(x) \leq F(ax + b)$ for all continuity points x of G .

If $a = 0$, then $F(b^-) \leq G(x) \leq F(b)$ for all continuity points, and hence for all $x \in R$. But then $F(b) = 1$, $F(b^-) = 0$ since $G(\infty) = 1$, $G(-\infty) = 0$. Thus F is degenerate at b , a contradiction.

Finally, if G is continuous at x and F is continuous at $ax + b$, we have just seen that $F(ax + b)^- \leq G(x) \leq F(ax + b)$, so $G(x) = F(ax + b)$. Since there are only countably many real numbers y such that G is discontinuous at y or F is discontinuous at $ay + b$, it follows that $G(x) = F(ax + b)$ for all $x \in R$.

Now if we have other convergent subsequences $a_{m_i} \rightarrow a'$, $b_{m_i} \rightarrow b'$, the above argument shows that $0 < a' < \infty$, $-\infty < b' < \infty$, and $G(x) = F(ax + b) = F(a'x + b')$ for all $x \in R$, so that $a^{-1}(Y - b) \stackrel{d}{=} (a')^{-1}(Y - b')$.

Since random variables with the same distribution have the same characteristic function,

$$\exp\left(\frac{-iub}{a}\right)h_Y\left(\frac{u}{a}\right) = \exp\left(\frac{-iub'}{a'}\right)h_Y\left(\frac{u}{a'}\right) \quad \text{for all } u.$$

Say $a < a'$, and set $k = a/a'$. Let $v = u/a$ to obtain

$$|h_Y(v)| = \left|h_Y\left(\frac{av}{a'}\right)\right| = |h_Y(kv)|$$

for all v . Thus

$$|h_Y(v)| = |h_Y(kv)| = |h_Y(k^2v)| = \dots = |h_Y(k^n v)| \rightarrow |h_Y(0)| = 1.$$

It follows that Y is degenerate, a contradiction (see Problem 4, Section 8.1). Thus $a = a'$, so $e^{-iub} = e^{-iub'}$ for all sufficiently small u ; hence $b = b'$. Therefore, $a_n \rightarrow a$, $b_n \rightarrow b$, proving 8.3.4(a).

Comment. If $a^{-1}(Y - b) \xrightarrow{d} (a')^{-1}(Y - b')$, where a and a' are nonzero but not necessarily positive, then if $|a| < |a'|$ and we set $k = a/a'$, we obtain a contradiction as above. We can conclude that $|a| = |a'|$ but not that $a = a'$.

(b) If $a_n > 0$, X_n and Y_n are of the same positive type, and if $a_n < 0$, $X_n \xrightarrow[d]{n \in S_1} -a_n^{-1}(-Y_n + b_n)$; hence X_n and $-Y_n$ are of the same positive type.

Let $S_1 = \{n: a_n > 0\}$, $S_2 = \{n: a_n < 0\}$. If S_1 is infinite, part (a) shows that X and $a^{-1}(Y - b)$ have the same distribution for some real a, b , $a > 0$, and

$$\lim_{\substack{n \rightarrow \infty \\ n \in S_1}} a_n = a, \quad \lim_{\substack{n \rightarrow \infty \\ n \in S_1}} b_n = b.$$

Now suppose that S_2 is infinite. Then $Y_n \xrightarrow{d} Y$ implies $-Y_n \xrightarrow{d} -Y$ (use 8.2.9), and it follows from part (a) that for some real a', b' , with $a' < 0$, we have $X \xrightarrow{d} -(a')^{-1}(-Y + b')$, and

$$\lim_{\substack{n \rightarrow \infty \\ n \in S_2}} a_n = a', \quad \lim_{\substack{n \rightarrow \infty \\ n \in S_2}} b_n = b'.$$

Now there are three possibilities:

Case 1. S_1 and S_2 are both infinite. Then since $a^{-1}(Y - b) \xrightarrow{d} (a')^{-1}(Y - b') \xrightarrow{d} X$, we have $|a| = |a'|$ [see the comment after the proof of (a)]. Thus $|a_n| \rightarrow |a|$ and the result follows.

Case 2. S_1 is infinite, S_2 finite. Then $a_n \rightarrow a$, $b_n \rightarrow b$, and $X \xrightarrow{d} a^{-1}(Y - b)$, proving the result.

Case 3. S_1 is finite, S_2 infinite. Then $a_n \rightarrow a'$, $b_n \rightarrow b'$, and $X \xrightarrow{d} (a')^{-1}(Y - b')$, and the result follows. ■

8.10 References

An exhaustive treatment of the one-dimensional central limit theorem is given by Gnedenko and Kolmogorov (1954). Conditions are derived for convergence in distribution of the row sums of a triangular array to a given infinitely divisible distribution. If X is a given stable random variable with distribution function G and X_1, X_2, \dots are iid random variables with distribution function F , conditions on F are given that are necessary and sufficient for there to exist a sequence of normed sums (formed from the X_i) converging in distribution to X . (In this case, F is said to belong to the *domain of attraction* of G .) Also, results on rates of convergence are given; in other words, if $Y_n \xrightarrow{d} Y$, there are estimates as to how fast the distribution function of Y_n approaches that of Y .

Prokhorov's weak compactness theorem actually holds in a complete, separable metric space; see Billingsley (1968) for a proof. An n -dimensional extension of Lindeberg's theorem is given by Gikhman and Skorokhod (1969). Aspects of the central limit theorem for random variables with values in abstract spaces, for example, Hilbert spaces or locally compact groups, are discussed by Parthasarathy (1967).

Appendix on General Topology

A1 Introduction

The reader is assumed to be familiar with elementary set theory, including basic properties of ordinal and cardinal numbers [see, for example, Halmos (1960)]. Also, an undergraduate course in point-set topology is assumed; Simmons (1963) is a suitable text for such a course. In this appendix, we shall concentrate on aspects of general topology that are useful in functional analysis and probability. A good reference for collateral reading is Dugundji (1966).

Before proceeding, we mention one result, which, although usually covered in a first course in topology, deserves to be stated explicitly because of its fundamental role in the construction of topological vector spaces (see 3.5.1). Throughout the appendix, a *neighborhood* of a point x is an open set containing x , an *overneighborhood* of x is an overset of a neighborhood of x .

A1.1 Theorem. Let Ω be a set, and suppose that for each $x \in \Omega$, we are given a nonempty collection $\mathcal{V}(x)$ of subsets of Ω satisfying the following:

- (a) $x \in$ each $U \in \mathcal{V}(x)$.
- (b) If $U_1, U_2 \in \mathcal{V}(x)$, then $U_1 \cap U_2 \in \mathcal{V}(x)$.
- (c) If $U \in \mathcal{V}(x)$ and $U \subset V$, then $V \in \mathcal{V}(x)$.

(d) If $U \in \mathcal{V}(x)$, there is a set $V \in \mathcal{V}(x)$ such that $V \subset U$ and $U \in \mathcal{V}(y)$ for each $y \in V$.

Then there is a unique topology on Ω such that for each x , $\mathcal{V}(x)$ is the system of overneighborhoods of x .

PROOF. If such a topology exists, a set U will be open iff U is an overneighborhood of each of its points, that is, $U \in \mathcal{V}(x)$ for each $x \in U$. Thus it suffices to show that $\mathcal{T} = \{U \subset \Omega : U \in \mathcal{V}(x) \text{ for each } x \in U\}$ is a topology, and for each x , the overneighborhood system $\mathcal{N}(x)$ coincides with $\mathcal{V}(x)$. If $U, V \in \mathcal{T}$, then by (b), $U \cap V \in \mathcal{V}(x)$ for each $x \in U \cap V$, so $U \cap V \in \mathcal{T}$. If $U_i \in \mathcal{T}$ for each $i \in I$, then (c) implies that $\bigcup\{U_i : i \in I\} \in \mathcal{T}$; (c) yields $\Omega \in \mathcal{T}$ as well. Since $\emptyset \in \mathcal{T}$ trivially, \mathcal{T} is a topology.

If $U \in \mathcal{N}(x)$, there is a set $V \in \mathcal{T}$ with $x \in V \subset U$. But then $U \in \mathcal{V}(x)$ by (c). Conversely, if $U \in \mathcal{V}(x)$, let $W = \{x' \in U : U \in \mathcal{V}(x')\}$. If $x' \in W$, then by (d), there is a set $V \in \mathcal{V}(x')$ with $V \subset U$ and $U \in \mathcal{V}(y)$ for each $y \in V$. But then $V \subset W$, so by (c), $W \in \mathcal{V}(x')$; consequently, $W \in \mathcal{T}$ by definition of \mathcal{T} , and furthermore $x \in W$ by (a). Thus if $U \in \mathcal{V}(x)$, there is a set $W \in \mathcal{T}$ with $x \in W \subset U$; hence $U \in \mathcal{N}(x)$. ■

For the remainder of the appendix, $\mathcal{U}(x)$ will always stand for the collection of neighborhoods of x .

A2 Convergence

If Ω is a metric space, the topology of Ω can be described entirely in terms of convergence of sequences. For example, a subset A of Ω is closed iff whenever $\{x_n, n = 1, 2, \dots\}$ is a sequence of points in A and $x_n \rightarrow x$, we have $x \in A$. Also, $x \in \bar{A}$, the closure of A , iff there is a sequence of points in A converging to x . This result does not generalize to arbitrary topological spaces.

A2.1 Example. Let α be the first uncountable ordinal, and let Ω be the set of all ordinals less than or equal to α (recall that for ordinals, $a < b$ means $a \in b$). Put the order topology on Ω ; this topology has as a base the sets $\Omega \cap (a, b) = \{x \in \Omega : a < x < b\}$ where a and b are arbitrary ordinals.

We show that α belongs to the closure of $\Omega - \{\alpha\}$, but no sequence in $\Omega - \{\alpha\}$ converges to α . If U is a neighborhood of α , then for some a, b , we have $\alpha \in \Omega \cap (a, b) \subset U$. Now $a < \alpha < b$; hence a is countable, and therefore so is $a + 1$. Thus $a + 1 \in U \cap (\Omega - \{\alpha\})$, proving that $\alpha \in \overline{\Omega - \{\alpha\}}$. But if $x_n \in \Omega - \{\alpha\}$, $n = 1, 2, \dots$, then each x_n is countable; hence so is $c = \sup x_n$.

Since $c < \alpha$, $V = (c, \alpha]$ is a neighborhood of α and x_n is never in V , so that the sequence cannot possibly converge to α .

In order to describe an arbitrary topology in terms of convergence, we must consider objects more general than sequences. A sequence is a function on the positive integers; the desired generalization, called a "net" or "generalized sequence," is a function on a directed set.

A2.2 Definitions. A *directed set* is a set D on which there is defined a preordering (a reflexive and transitive relation), denoted by \leq , with the property that whenever $a, b \in D$, there is a $c \in D$ with $a \leq c$ and $b \leq c$. A *net* in a topological space Ω is a function from a directed set D into Ω . A net will be denoted by $\{x_n, n \in D\}$ or simply by $\{x_n\}$. The net $\{x_n\}$ is said to *converge* to the point x iff for every neighborhood U of x , the net is *eventually* in U , that is, there is an $n_0 \in D$ such that $x_n \in U$ for all $n \in D$ such that $n \geq n_0$.

The basic relations between convergence and topology in metric spaces can now be generalized to arbitrary topological spaces.

A2.3 Theorem. Let A be a subset of the topological space Ω .

- (a) A point $x \in \Omega$ belongs to \bar{A} iff there is a net $\{x_n\}$ in A such that $x_n \rightarrow x$.
- (b) A is closed iff for every net $\{x_n\}$ in A such that $x_n \rightarrow x$, we have $x \in A$.
- (c) A point $x \in \Omega$ is a cluster point of A (that is, every neighborhood of x contains a point of A other than x) iff there is a net in $A - \{x\}$ converging to x .

PROOF. (a) If $x_n \in A$ and $x_n \rightarrow x$, let $U \in \mathcal{U}(x)$. Then x_n is eventually in U , in particular $U \cap A \neq \emptyset$, and thus $x \in \bar{A}$. Conversely, if $x \in \bar{A}$, then for each $U \in \mathcal{U}(x)$, choose $x_U \in U \cap A$. Then $\{x_U\}$ becomes a net in A (with $U \leq V$ iff $U \supset V$) and $x_U \rightarrow x$.

(b) Suppose that A is closed. If $x_n \in A$, $x_n \rightarrow x$, then by (a), $x \in \bar{A}$; thus by hypothesis, $x \in A$. Conversely, if A is not closed, let $x \in \bar{A} - A$. By (a) we can find $x_n \in A$, $x_n \rightarrow x$. Since $x \notin A$, the result follows.

(c) If $x_n \in A - \{x\}$, $x_n \rightarrow x$, then if $U \in \mathcal{U}(x)$, x_n is eventually in U , hence $U \cap (A - \{x\}) \neq \emptyset$, and thus x is a cluster point of A . Conversely, if x is a cluster point of A , then for each $U \in \mathcal{U}(x)$, choose $x_U \in U \cap (A - \{x\})$; the x_U form a net in $A - \{x\}$ converging to x . ■

A2.4 Comments. The reason that sequences are not adequate in A2.3 is that in choosing a point x_U in each neighborhood U of x , we are in general forced to make uncountably many choices. This would not be necessary if Ω were

first countable. (A first countable space is one with a countable base of neighborhoods at each point; that is, if $x \in \Omega$, there are countably many neighborhoods V_1, V_2, \dots of x such that for each neighborhood U of x , we have $V_n \subset U$ for some n ; by replacing V_n by $\bigcap_{i=1}^n V_i$, we may assume without loss of generality that $V_{n+1} \subset V_n$ for all n .) In A2.3, the proof will go through if we simply choose $x_n \in V_n$ for $n = 1, 2, \dots$. Thus *in a first countable space, sequences are adequate to describe the topology*; in other words, A2.3 holds with “net” replaced by “sequence.”

A criterion for openness in terms of nets may also be given: The set V is open iff for every net $\{x_n\}$ in Ω such that $x_n \rightarrow x \in V$, we have $x_n \in V$ eventually. [The “only if” part follows from the definition of convergence; for the “if” part, apply A2.3(b) to V^c .]

A2.5 Definitions. Let $\{x_n, n \in D\}$ be a net, and suppose that we are given a directed set E and a map $k \rightarrow n_k$ of E into D . Then $\{x_{n_k}, k \in E\}$ is called a *subnet* of $\{x_n, n \in D\}$, provided that “as k becomes large, so does n_k ”; that is, given $n_0 \in D$, there is a $k_0 \in E$ such that $k \geq k_0$ implies $n_k \geq n_0$. If $E = D$ = the positive integers, we obtain the usual notion of a *subsequence*.

If $\{x_n, n \in D\}$ is a net in the topological space Ω , the point $x \in \Omega$ is called an *accumulation point* of the net iff for each neighborhood U of x , x_n is frequently in U ; in other words, given $n \in D$, there is an $m \in D$ with $m \geq n$ and $x_m \in U$.

A2.6 Theorem. Let $\{x_n, n \in D\}$ be a net in the topological space Ω . If $x \in \Omega$, x is an accumulation point of $\{x_n\}$ iff there is a subnet $\{x_{n_k}, k \in E\}$ converging to x .

PROOF. If $x_{n_k} \rightarrow x$, $U \in \mathcal{U}(x)$ and $n \in D$, then for some $k_0 \in E$, we have $x_{n_k} \in U$ for $k \geq k_0$. But by definition of subnet, there is a $k_1 \in E$ such that $k \geq k_1$ implies $n_k \geq n$. Thus if $k \geq k_0$ and $k \geq k_1$, we have $n_k \geq n$ and $x_{n_k} \in U$, proving that x is an accumulation point.

Conversely, let x be an accumulation point of $\{x_n, n \in D\}$. Let E be the collection of pairs (n, U) , where $n \in D$, U is a neighborhood of x , and $x_n \in U$. Direct E by setting $(n, U) \leq (m, V)$ iff $n \leq m$ and $U \supset V$. If $k = (m, V) \in E$ let $n_k = m$. Given n and U , then if $k = (m, V) \geq (n, U)$, we have $n_k = m \geq n$, so that $\{x_{n_k}, k \in E\}$ is a subnet of $\{x_n, n \in D\}$. Now if U is a neighborhood of x , then $x_n \in U$ for some $n \in D$. If $k = (m, V) \geq (n, U)$, then $x_{n_k} = x_m \in V \subset U$; therefore $x_{n_k} \rightarrow x$. ■

For some purposes, it is more convenient to specify convergence in a topological space by means of filters rather than nets. If $\{x_n, n \in D\}$ is a net

in Ω , and $a \in D$, let $T_a = \{n \in D : n \geq a\}$, and let $x(T_a)$ be the set of all x_n , $n \geq a$. The $x(T_a)$, $a \in D$, are called the *tails* of the net. The collection \mathcal{A} of tails is an example of a filterbase, which we now define.

A2.7 Definitions and Comments. Let \mathcal{A} be a nonempty family of subsets of a set Ω . Then \mathcal{A} is called a *filterbase* in Ω iff

- (a) each $U \in \mathcal{A}$ is nonempty;
- (b) if $U, V \in \mathcal{A}$, there is a $W \in \mathcal{A}$ with $W \subset U \cap V$.

If \mathcal{F} is a nonempty family of subsets of Ω such that

- (c) each $U \in \mathcal{F}$ is nonempty,
- (d) if $U, V \in \mathcal{F}$, then $U \cap V \in \mathcal{F}$, and
- (e) if $U \in \mathcal{F}$ and $U \subset V$, then $V \in \mathcal{F}$,

then \mathcal{F} is called a *filter* in Ω . If \mathcal{A} is a filterbase, then $\mathcal{F} = \{U \subset \Omega : U \supset V \text{ for some } V \in \mathcal{A}\}$ is a filter, called the *filter generated by \mathcal{A}* . If \mathcal{A} is the collection of neighborhoods of a given point x in a topological space, \mathcal{A} is a filterbase, and the filter generated by \mathcal{A} is the system of overneighborhoods of x .

A filterbase \mathcal{A} in a topological space Ω is said to *converge* to the point x (notation $\mathcal{A} \rightarrow x$) iff for each $U \in \mathcal{U}(x)$ there is a set $A \in \mathcal{A}$ such that $A \subset U$. A filter \mathcal{F} in Ω is said to converge to x iff each $U \in \mathcal{U}(x)$ belongs to \mathcal{F} . Thus a filterbase \mathcal{A} converges to x iff the filter generated by \mathcal{A} converges to x .

If $\{x_n, n \in D\}$ is a net, then $x_n \rightarrow x$ iff for each $U \in \mathcal{U}(x)$ we have $x(T_a) \subset U$ for some $a \in D$, that is, $x_n \rightarrow x$ iff the associated filterbase converges to x .

Convergence in a topological space may be described using filterbases instead of nets. The analog of Theorem A2.3 is the following:

A2.8 Theorem. Let B be a subset of the topological space Ω .

- (a) A point $x \in \Omega$ belongs to \bar{B} iff there is a filterbase \mathcal{A} in B such that $\mathcal{A} \rightarrow x$.
- (b) B is closed iff for every filterbase \mathcal{A} in B such that $\mathcal{A} \rightarrow x$, we have $x \in B$.
- (c) A point $x \in \Omega$ is a cluster point of B iff there is a filterbase in $B - \{x\}$ converging to x .

PROOF. (a) If $\mathcal{A} \rightarrow x$ and $U \in \mathcal{U}(x)$, then $A \subset U$ for some $A \in \mathcal{A}$, in particular, $U \cap B \neq \emptyset$; thus $x \in \bar{B}$. Conversely, if $x \in \bar{B}$, then $U \cap B \neq \emptyset$ for each $U \in \mathcal{U}(x)$. Let \mathcal{A} be the collection of sets $U \cap B$, $U \in \mathcal{U}(x)$. Then \mathcal{A} is a filterbase in B and $\mathcal{A} \rightarrow x$.

(b) If B is closed, \mathcal{A} is a filterbase in B , and $\mathcal{A} \rightarrow x$, then $x \in \bar{B}$ by (a), hence $x \in B$ by hypothesis. Conversely, if B is not closed and $x \in \bar{B} - B$, by (a) there is a filterbase \mathcal{A} in B with $\mathcal{A} \rightarrow x$. Since $x \notin B$, the result follows.

(c) If there is such a filterbase \mathcal{A} and $U \in \mathcal{U}(x)$, then $U \supset A$ for some $A \in \mathcal{A}$; in particular, $U \cap (B - \{x\}) \neq \emptyset$, so x is a cluster point of B . Conversely, if x is a cluster point of B , let \mathcal{A} consist of all sets $U \cap (B - \{x\})$, $U \in \mathcal{U}(x)$. Then \mathcal{A} is a filterbase in $B - \{x\}$ and $\mathcal{A} \rightarrow x$. ■

If Ω is first countable, the filterbases in A2.8 may be formed using a countable system of neighborhoods of x , so that in a first countable space, the topology may be described by filterbases containing countably many sets.

A2.9 Definitions. The filterbase \mathcal{B} is said to be *subordinate* to the filterbase \mathcal{A} iff for each $A \in \mathcal{A}$ there is a $B \in \mathcal{B}$ with $B \subset A$; this means that the filter generated by \mathcal{A} is included in the filter generated by \mathcal{B} .

If $\{x_{n_k}, k \in E\}$ is a subnet of $\{x_n, n \in D\}$, the filterbase determined by the subnet is subordinate to the filterbase determined by the original net. For if $n_0 \in D$, there is a $k_0 \in E$ such that $k \geq k_0$ implies $n_k \geq n_0$. Therefore

$$\{x_{n_k}: k \in E, k \geq k_0\} \subset \{x_n: n \in D, n \geq n_0\}.$$

If \mathcal{A} is a filterbase in the topological space Ω , the point $x \in \Omega$ is called an *accumulation point* of \mathcal{A} iff $U \cap A \neq \emptyset$ for all $U \in \mathcal{U}(x)$ and all $A \in \mathcal{A}$, in other words, $x \in \bar{A}$ for all $A \in \mathcal{A}$. We may now prove the analog of Theorem A2.6.

A2.10 Theorem. Let \mathcal{A} be a filterbase in the topological space Ω . If $x \in \Omega$, x is an accumulation point of \mathcal{A} iff there is a filterbase \mathcal{B} subordinate to \mathcal{A} with $\mathcal{B} \rightarrow x$; in other words, some overfilter of \mathcal{A} converges to x .

PROOF. If \mathcal{B} is subordinate to \mathcal{A} and $\mathcal{B} \rightarrow x$, let $U \in \mathcal{U}(x)$, $A \in \mathcal{A}$. Then $U \supset B$ and $A \supset B_1$ for some $B, B_1 \in \mathcal{B}$; hence $U \cap A \supset B \cap B_1$, which is nonempty since \mathcal{B} is a filterbase. Therefore $x \in \bar{A}$. Conversely, if x is an accumulation point of \mathcal{A} , let \mathcal{B} consist of all sets $U \cap A$, $U \in \mathcal{U}(x)$, $A \in \mathcal{A}$. Then $\mathcal{A} \subset \mathcal{B}$ (take $U = \Omega$), hence \mathcal{B} is subordinate to \mathcal{A} ; since $\mathcal{B} \rightarrow x$, the result follows. ■

A2.11 Definition. An *ultrafilter* is a maximal filter, that is, a filter included in no properly larger filter. (By Zorn's lemma, every filter is included in an ultrafilter.)

A2.12 Theorem. Let \mathcal{F} be a filter in the set Ω .

- (a) \mathcal{F} is an ultrafilter iff for each $A \subset \Omega$ we have $A \in \mathcal{F}$ or $A^c \in \mathcal{F}$.
- (b) If \mathcal{F} is an ultrafilter and $p: \Omega \rightarrow \Omega'$, the filter \mathcal{G} generated by the filterbase $p(\mathcal{F}) = \{p(F): F \in \mathcal{F}\}$ is an ultrafilter in Ω' .
- (c) If Ω is a topological space and \mathcal{F} is an ultrafilter in Ω , \mathcal{F} converges to each of its accumulation points.

PROOF. (a) If \mathcal{F} is an ultrafilter and $A \notin \mathcal{F}$, necessarily $A \cap B = \emptyset$ for some $B \in \mathcal{F}$. For if not, let \mathcal{A} consist of all sets $A \cap B$, $B \in \mathcal{F}$; then \mathcal{A} is a filterbase generating a filter larger than \mathcal{F} . But $A \cap B = \emptyset$ implies $B \subset A^c$; hence $A^c \in \mathcal{F}$. Conversely, if the condition is satisfied, let \mathcal{F} be included in the filter \mathcal{G} . If $A \in \mathcal{G}$ and $A \notin \mathcal{F}$, then $A^c \in \mathcal{F} \subset \mathcal{G}$, a contradiction since $A \cap A^c = \emptyset$.

(b) Let $A \subset \Omega'$; by (a), either $p^{-1}(A) \in \mathcal{F}$ or $p^{-1}(A^c) \in \mathcal{F}$. If $p^{-1}(A) \in \mathcal{F}$, then $A \supset pp^{-1}(A) \in p(\mathcal{F})$; hence $A \in \mathcal{G}$. Similarly, if $p^{-1}(A^c) \in \mathcal{F}$, then $A^c \in \mathcal{G}$. By (a), \mathcal{G} is an ultrafilter.

(c) Let x be an accumulation point of \mathcal{F} . If $U \in \mathcal{U}(x)$ and $U \notin \mathcal{F}$, then $U^c \in \mathcal{F}$ by (a). But $U \cap U^c = \emptyset$, contradicting the fact that x is an accumulation point of \mathcal{F} . ■

We have associated with each net $\{x_n, n \in D\}$ the filterbase $\{x(T_a): a \in D\}$ of tails of the net, and have seen that convergence of the net is equivalent to convergence of the filterbase. We now prove a converse result.

A2.13 Theorem. If \mathcal{A} is a filterbase in the set Ω , there is a net in Ω such that the collection of tails of the net coincides with \mathcal{A} .

PROOF. Let D be all ordered pairs (a, A) where $a \in A$ and $A \in \mathcal{A}$; define $(a, A) \leq (b, B)$ iff $B \subset A$. If (a, A) and (b, B) belong to D , choose $C \in \mathcal{A}$ with $C \subset A \cap B$; for any $c \in C$ we have $(c, C) \geq (a, A)$ and $(c, C) \geq (b, B)$, hence D is directed. If we set $x_{(a, A)} = a$ we obtain a net in Ω with $x(T_{(a, A)}) = A$. ■

We conclude this section with a characterization of continuity.

A2.14 Theorem. Let $f: \Omega \rightarrow \Omega'$, where Ω and Ω' are topological spaces. The following are equivalent:

- (a) The function f is continuous on Ω ; that is, $f^{-1}(V)$ is open in Ω whenever V is open in Ω' .

(b) For every net $\{x_n\}$ in Ω converging to the point $x \in \Omega$, the net $\{f(x_n)\}$ converges to $f(x)$.

(c) For every filterbase \mathcal{A} in Ω converging to the point $x \in \Omega$, the filterbase $f(\mathcal{A})$ converges to $f(x)$.

PROOF. Let $\{x_n\}$ be a net and \mathcal{A} a filterbase such that the tails of the net coincide with the elements of the filterbase. If, say, $x(T_a) = A \in \mathcal{A}$, then $f(A) = \{f(x_n): n \in D, n \geq a\}$. Thus the tails of the net $\{f(x_n)\}$ coincide with the elements of $f(\mathcal{A})$. It follows that (b) and (c) are equivalent.

If f is continuous and $x_n \rightarrow x$, let V be a neighborhood of $f(x)$. Then $f^{-1}(V)$ is a neighborhood of x ; hence x_n is eventually in $f^{-1}(V)$, so that $f(x_n)$ is eventually in V . Thus (a) implies (b). Conversely, if (b) holds and C is closed in Ω' , let $\{x_n\}$ be a net in $f^{-1}(C)$ converging to x . Then $f(x_n) \rightarrow f(x)$ by (b), and since C is closed we have $f(x) \in C$ by A2.3(b). Thus $x \in f^{-1}(C)$, hence $f^{-1}(C)$ is closed, proving continuity of f . ■

A3 Product and Quotient Topologies

In the Euclidean plane R^2 , a base for the topology may be formed from sets $U \times V$, where U and V are open subsets of R ; in fact U and V can be taken to be open intervals, so that $U \times V$ is an open rectangle. If

$$\{(x_n, y_n), \quad n = 1, 2, \dots\}$$

is a sequence in R^2 , then $(x_n, y_n) \rightarrow (x, y)$ iff $x_n \rightarrow x$ and $y_n \rightarrow y$, that is, convergence in R^2 is “pointwise” or “coordinatewise” convergence.

In general, given an arbitrary collection of topological spaces Ω_i , $i \in I$, let Ω be the cartesian product $\prod_{i \in I} \Omega_i$, which is the collection of all families $(x_i, i \in I)$; that is, all functions on I such that $x_i \in \Omega_i$ for each i . We shall place a topology on Ω such that convergence in the topology coincides with pointwise convergence.

A3.1 Definition. The *product topology* (also called the *topology of pointwise convergence*) on $\Omega = \prod_{i \in I} \Omega_i$ has as a base all sets of the form

$$\{x \in \Omega: x_{i_k} \in U_{i_k}, \quad k = 1, \dots, n\}$$

where the U_{i_k} are open in Ω_{i_k} and n is an arbitrary positive integer. (Since the intersection of two sets of this type is a set of this type, the sets do in fact form a base.)

If p_i is the projection of Ω onto Ω_i , the product topology is the weakest topology making each p_i continuous; in other words, the product topology is included in any topology that makes each p_i continuous.

The product topology has the following properties:

A3.2 Theorem. Let $\Omega = \prod_{i \in I} \Omega_i$, with the product topology.

- (a) If $\{x^{(n)}, n \in D\}$ is a net in Ω and $x \in \Omega$, then $x^{(n)} \rightarrow x$ iff $x_i^{(n)} \rightarrow x_i$ for each i .
- (b) A map f from a topological space Ω_0 into Ω is continuous iff $p_i \circ f$ is continuous for each i .
- (c) If $f_i: \Omega_0 \rightarrow \Omega_i$, $i \in I$, and we define $f: \Omega_0 \rightarrow \Omega$ by $f(x) = (f_i(x), i \in I)$, then f is continuous iff each f_i is continuous.
- (d) The projections p_i are open maps of Ω onto Ω_i .

PROOF. (a) If $x^{(n)} \rightarrow x$, then $x_i^{(n)} = p_i(x^{(n)}) \rightarrow p_i(x) = x_i$ by continuity of the p_i . Conversely, assume $x_i^{(n)} \rightarrow x_i$ for each i . Let

$$V = \{y \in \Omega: y_{i_k} \in U_{i_k}, \quad k = 1, \dots, r\},$$

be a basic neighborhood of x . Since $x_{i_k} \in U_{i_k}$, there is an $n_k \in D$ with $x_{i_k}^{(n)} \in U_{i_k}$ for $n \geq n_k$. Therefore, if $n \in D$ and $n \geq n_k$ for all $k = 1, \dots, r$, we have $x \in V$, so that $x^{(n)} \rightarrow x$.

(b) The “only if” part follows by continuity of the p_i . Conversely, assume each $p_i \circ f$ continuous. If $x^{(n)} \rightarrow x$, then $p_i(f(x^{(n)})) \rightarrow p_i(f(x))$ by hypothesis; hence $f(x^{(n)}) \rightarrow f(x)$ by (a).

(c) We have $f_i = p_i \circ f$, so (b) applies.

(d) The result follows from the observation that

$$p_i\{x \in \Omega: x_{i_k} \in U_{i_k}, \quad k = 1, \dots, n\} = \begin{cases} \Omega_i & \text{if } i \neq \text{any } i_k, \\ U_{i_k} & \text{if } i = i_k \text{ for some } k. \end{cases} \blacksquare$$

Note that if Ω is the collection of all functions from a topological space S to a topological space T , then $\Omega = \prod_{i \in I} \Omega_i$, where $I = S$ and $\Omega_i = T$ for all i . If $\{f_n\}$ is a net of functions from S to T , and $f: S \rightarrow T$, then $f_n \rightarrow f$ in the product topology iff $f_n(s) \rightarrow f(s)$ for all $s \in S$.

We now consider quotient spaces.

A3.3 Definition. Let Ω_0 be a topological space, and p a map of Ω_0 onto a set Ω . The *identification topology* on Ω is the strongest topology making p continuous, that is, the open subsets of Ω are the sets U such that $p^{-1}(U)$ is

open in Ω_0 . When Ω has the identification topology, it is called an *identification space*, and p is called an *identification map*.

A quotient topology may be regarded as a particular identification topology. Let R be an equivalence relation on the topological space Ω_0 , Ω_0/R the set of equivalence classes, and $p: \Omega_0 \rightarrow \Omega_0/R$ the canonical projection: $p(x) = [x]$, the equivalence class containing x . The *quotient space* of Ω_0 by R is Ω_0/R with the identification topology determined by p .

In fact, any identification space may be regarded as a quotient space. To see this, we need two preliminary results.

A3.4 Lemma. If Ω has the identification topology determined by $p: \Omega_0 \rightarrow \Omega$, then $g: \Omega \rightarrow \Omega_1$ is continuous iff $g \circ p: \Omega_0 \rightarrow \Omega_1$ is continuous.

PROOF. The “only if” part follows from the continuity of p . If $g \circ p$ is continuous and V is open in Ω_1 , then $(g \circ p)^{-1}(V) = p^{-1}(g^{-1}(V))$ is open in Ω_0 . By definition of the identification topology, $g^{-1}(V)$ is open in Ω . ■

A3.5 Lemma. Let $p: \Omega_0 \rightarrow \Omega$ be an identification, and let $h: \Omega_0 \rightarrow \Omega_1$ be continuous. Assume that $h \circ p^{-1}$ is single-valued, in other words, $p(x) = p(y)$ implies $h(x) = h(y)$. Then $h \circ p^{-1}$ is continuous.

PROOF. Since $h = (h \circ p^{-1}) \circ p$, the result follows from A3.4. (Note that p^{-1} is defined on all of Ω since p is onto.) ■

A3.6 Theorem. Let $f: \Omega_0 \rightarrow \Omega$ be an identification. Define an equivalence relation R on Ω_0 by calling x and y equivalent iff $f(x) = f(y)$. Let p be the canonical projection of Ω_0 onto Ω_0/R . Then Ω_0/R , with the quotient topology, is homeomorphic to Ω .

PROOF. We have $f(x) = f(y)$ iff $p(x) = p(y)$, so by A3.5, $f \circ p^{-1}$ and $p \circ f^{-1}$ are both continuous. Since these functions are inverses of each other, they define a homeomorphism of Ω and Ω_0/R . ■

The following result gives conditions under which a given topology arises from an identification.

A3.7 Theorem. Let p be a map of Ω_0 onto Ω . If p is continuous and either open or closed, it is an identification, that is, the identification topology on Ω determined by p coincides with the original topology on Ω .

PROOF. Since p is continuous and the identification topology is the largest one making p continuous, the original topology is included in the identification topology. If p is an open map and U is an open subset of Ω in the identification topology, $p^{-1}(U)$ is open in Ω_0 , hence $p(p^{-1}(U)) = U$ is open in the original topology. If p is a closed map, the same argument applies, with “open” replaced by “closed.” ■

A4 Separation Properties and Other Ways of Classifying Topological Spaces

Topological spaces may be classified as to how well disjoint sets may be separated, as follows. (The results of this section are generally discussed in a first course in topology, and will not be proved.)

A4.1 Definitions. A topological space Ω is said to be a T_0 space iff given any two distinct points x and y , *at least one* point has a neighborhood not containing the other; Ω is a T_1 space iff *each* point has a neighborhood not containing the other; Ω is a T_2 (*or Hausdorff*) space iff x and y have disjoint neighborhoods. This is equivalent to uniqueness of limits of nets (or filter-bases). Also, Ω is said to be a T_3 (*or regular*) space iff Ω is T_2 and for each closed set C and point $x \notin C$, there are disjoint open sets U and V with $x \in U$ and $C \subset V$; Ω is said to be a T_4 (*or normal*) space iff Ω is T_2 and for each pair of disjoint closed sets A and B there are disjoint open sets U and V with $A \subset U$ and $B \subset V$.

It follows from the definitions that T_i implies T_{i-1} , $i = 4, 3, 2, 1$. Also, the T_1 property is equivalent to the statement that $\{x\}$ is closed for each x . The space Ω is regular iff it is Hausdorff and for each open set U and point $x \in U$, there is a $V \in \mathcal{U}(x)$ with $\bar{V} \subset U$. The space Ω is normal iff it is Hausdorff and for each closed set A and open set $U \supset A$, there is an open set V with $A \subset V \subset \bar{V} \subset U$.

A metric space is T_4 , for if A and B are disjoint closed sets, we may take $U = \{x: d(x, A) - d(x, B) < 0\}$, $V = \{x: d(x, A) - d(x, B) > 0\}$, where $d(x, A) = \inf\{d(x, y): y \in A\}$.

A4.2 Urysohn's Lemma. Let Ω be a Hausdorff space. Then Ω is normal iff for each pair of disjoint closed sets A and B , there is a continuous function $f: \Omega \rightarrow [0, 1]$ with $f = 0$ on A and $f = 1$ on B .

A4.3 Tietze Extension Theorem. Let Ω be a Hausdorff space. Then Ω is normal iff for every closed set $A \subset \Omega$ and every continuous real-valued

function f defined on A , f has an extension to a continuous real-valued function F on Ω . Furthermore, if $|f| < c$ (respectively, $|f| \leq c$) on A , then $|F|$ can be taken less than c (respectively, less than or equal to c) on Ω .

A4.4 Theorem. Let A be a closed subset of the normal space Ω . There is a continuous $f: \Omega \rightarrow [0, 1]$ such that $A = f^{-1}\{0\}$ iff A is a G_δ , that is, a countable intersection of open sets.

A4.5 Definitions and Comments. A topological space Ω is *second countable* iff there is a countable base for the topology, *first countable* iff there is a countable base at each point (see A2.4). Second countability implies first countability but not conversely. Any metric space is first countable.

If Ω is second countable, it is necessarily *separable*, that is, there is a countable dense subset of Ω . Furthermore, if Ω is second countable, it is *Lindelöf*, that is, for every family of open sets V_i , $i \in I$, such that $\bigcup_i V_i = \Omega$, there is a countable subfamily whose union is Ω (for short, every open covering of Ω has a countable subcovering).

In a metric space, the separable, second countable, and Lindelöf properties are equivalent as follows. Second countability always implies separability and Lindelöf. If Ω is separable with a countable dense set $\{x_1, x_2, \dots\}$, then the balls $B(x_i, r) = \{y \in \Omega: d(y, x_i) < r\}$, $i = 1, 2, \dots$, r rational, form a countable base. If Ω is Lindelöf, the cover by balls $B(x, 1/n)$, $x \in \Omega$, has a countable subcover $\{B(x_{ni}, 1/n)\}$, $i = 1, 2, \dots$, and the sets $B(x_{ni}, 1/n)$, $i, n = 1, 2, \dots$, form a countable base.

This result implies that any space that is separable but not second countable (or not Lindelöf), or Lindelöf but not second countable (or not separable) cannot be metrizable, that is, there is no metric whose topology coincides with the original one.

A4.6 Definitions and Comments. A topological space Ω is said to be *completely regular* iff Ω is Hausdorff and for each $x \in \Omega$ and closed set $C \subset \Omega$ with $x \notin C$, there is a continuous $f: \Omega \rightarrow [0, 1]$ such that $f(x) = 1$ and $f = 0$ on C .

If A is a subset of the normal space Ω , then A , with the relative topology, is completely regular (this follows quickly from Urysohn's lemma). Also, complete regularity implies regularity. Thus complete regularity is in between regularity and normality; for this reason, completely regular spaces are sometimes called $T_{3\frac{1}{2}}$ spaces.

Now if Ω_0 is a Hausdorff space and \mathcal{F} is the family of continuous maps $f: \Omega_0 \rightarrow [0, 1]$, let $\Omega = \prod \{I_f: f \in \mathcal{F}\}$ where each $I_f = [0, 1]$. Let $e: \Omega_0 \rightarrow \Omega$ be the evaluation map, that is, $e(x) = (f(x), f \in \mathcal{F})$.

With the product topology on Ω , e is continuous, and e is one-to-one iff \mathcal{F} distinguishes points, in other words, given $x, y \in \Omega_0$, $x \neq y$, there is an $f \in \mathcal{F}$ such that $f(x) \neq f(y)$. Finally, if \mathcal{F} distinguishes points from closed sets, that is, if Ω_0 is completely regular, then e is an open map of Ω_0 onto $e(\Omega_0) \subset \mathcal{F}$. [If $\{x_n\}$ is a net and $e(x_n) \rightarrow e(x)$ but $x_n \not\rightarrow x$, there is a neighborhood U of x such that x_n is not eventually in U ; that is, given m , there is an $n \geq m$ with $x_n \notin U$. Choose $f \in \mathcal{F}$ with $f(x) = 1$ and $f = 0$ on $\Omega_0 - U$. Then for each m , we have $f(x_n) = 0$ for some $n \geq m$, so that $f(x_n) \not\rightarrow f(x)$. But then $e(x_n) \not\rightarrow e(x)$, a contradiction.]

It follows that if Ω_0 is completely regular, it is homeomorphic to a subset of a normal space. [Since Ω is a product of Hausdorff spaces, it is Hausdorff; the Tychonoff theorem, to be proved later, shows that Ω is compact, and hence normal (see A5.3(d) and A5.4).]

Since $e(\Omega_0)$ is determined completely by \mathcal{F} , we may say that the continuous functions are adequate to describe the topology of Ω_0 .

A5 Compactness

The notion of compactness appears in virtually all areas of mathematics. The original compactness result was the Heine–Borel theorem: If $[0, 1] \subset \bigcup_i V_i$, where the V_i are open subsets of \mathbb{R} , then in fact $[0, 1]$ is covered by finitely many V_i , that is, $[0, 1] \subset \bigcup_{k=1}^n V_{i_k}$ for some V_{i_1}, \dots, V_{i_n} . In general, we have the following definition:

A5.1 Definition. The topological space Ω is *compact* iff every open covering of Ω has a finite subcovering.

There are several ways of expressing this idea.

A5.2 Theorem. If Ω is a topological space, the following are equivalent:

- (a) Ω is compact.
- (b) Each family of closed sets $C_i \subset \Omega$ with the finite intersection property (all finite intersections of the C_i are nonempty) has nonempty intersection. Equivalently, for every family of closed subsets of Ω with empty intersection, there is a finite subfamily with empty intersection.
- (c) Every net in Ω has an accumulation point in Ω ; in other words (by A2.6), every net in Ω has a subnet converging to a point of Ω .
- (d) Every filterbase in Ω has an accumulation point in Ω , that is (by A2.10), every filterbase in Ω has a convergent filterbase subordinate to it, or equally well, every filter in Ω has an overfilter converging to a point of Ω .
- (e) Every ultrafilter in Ω converges to a point of Ω .

PROOF. Parts (a) and (b) are equivalent by the duality between open and closed sets. If $\{x_n\}$ is a net and \mathcal{A} is a filterbase whose elements coincide with the tails of the net, then the accumulation points of the net and of the filterbase coincide, so that (c) and (d) are equivalent. Now (d) implies (e) by A2.12(c), and (e) implies (d) since every filter is included in an ultrafilter. To prove that (b) implies (d), observe that if \mathcal{A} is a filterbase, the sets $A \in \mathcal{A}$, hence the sets \bar{A} , $A \in \mathcal{A}$, have the finite intersection property, so by (b), there is a point $x \in \bigcap \{\bar{A} : A \in \mathcal{A}\}$. Finally, we prove that (d) implies (b). If the closed sets C_i have the finite intersection property, the finite intersections of the C_i form a filterbase, which by (d) has an accumulation point x . But then x belongs to all C_i . ■

It is important to note that if $A \subset \Omega$, the statement that every covering of A by sets open in Ω has a finite subcovering is equivalent to the statement that every covering of A by sets open in A (in the relative topology inherited from Ω) has a finite subcovering. Thus when we talk about a compact subset of a topological space, there is no ambiguity.

The following results follow quickly from the definition of compactness.

A5.3 Theorem. (a) If Ω is compact and f is continuous on Ω , then $f(\Omega)$ is compact.

- (b) A closed subset C of a compact space Ω is compact.
- (c) If A and B are disjoint compact subsets of the Hausdorff space Ω , there are disjoint open sets U and V such that $A \subset U$ and $B \subset V$. In particular (take $A = \{x\}$), a compact subset of a Hausdorff space is closed.
- (d) A compact Hausdorff space is normal.
- (e) If A is a compact subset of the regular space Ω , and A is a subset of the open set U , there is an open set V with $A \subset V \subset \overline{V} \subset U$.

PROOF. (a) This is immediate from the definition of compactness.

(b) If C is covered by sets U open in Ω , the sets U together with $\Omega - C$ cover Ω . By compactness there is a finite subcover.

(c) If $x \notin B$ and $y \in B$, there are disjoint neighborhoods $U_y(x)$ and $V(y)$ of x and y . The $V(y)$ cover B ; hence there is a finite subcover $V(y_i)$, $i = 1, \dots, n$. Then $U' = \bigcap_{i=1}^n U_{y_i}(x)$ and $V' = \bigcup_{i=1}^n V(y_i)$ are disjoint open sets with $x \in U'$ and $B \subset V'$. If we repeat the process for each $x \in A$, we obtain disjoint open sets $U(x)$ and $V(x)$ as above. The $U(x)$ cover A ; hence there is a finite subcover $U(x_i)$, $i = 1, \dots, m$. Take $U = \bigcup_{i=1}^m U(x_i)$, $V = \bigcap_{i=1}^m V(x_i)$.

(d) If A and B are disjoint closed sets, they are compact by (b); the result then follows from (c).

(e) If $x \in A$, regularity yields an open set $V(x)$ with $x \in V(x)$ and $\overline{V(x)} \subset U$. The $V(x)$ cover A ; so for some x_1, \dots, x_n , we have

$$A \subset \bigcup_{i=1}^n V(x_i) \subset \bigcup_{i=1}^n \overline{V(x_i)} = \overline{\bigcup_{i=1}^n V(x_i)} \subset U. \blacksquare$$

The following is possibly the most important compactness result.

A5.4 Tychonoff Theorem. If Ω_i is compact for each $i \in I$, then $\Omega = \prod_{i \in I} \Omega_i$ is compact in the product topology.

PROOF. Let \mathcal{F} be an ultrafilter in Ω . If p_i is the projection of Ω onto Ω_i , then by A2.12(b), $p_i(\mathcal{F})$ is a filterbase that generates an ultrafilter in Ω_i . By hypothesis, $p_i(\mathcal{F})$ converges to some $x_i \in \Omega_i$, and it follows that $\mathcal{F} \rightarrow x = (x_i, i \in I)$. [To see this, observe that if \mathcal{A} is a filterbase in Ω and $\{x^n\}$ is a net whose tails are the elements of \mathcal{A} , then the tails of $\{p_i(x^n)\}$ are the elements of $p_i(\mathcal{A})$. Since $x^n \rightarrow x$ iff $p_i(x^n) \rightarrow p_i(x)$ for all i , by A3.2(a), it follows that $\mathcal{A} \rightarrow x$ iff $p_i(\mathcal{A}) \rightarrow p_i(x)$ for all i .] The Tychonoff theorem now follows from A5.2(e). \blacksquare

The following result is often used to infer that the inverse of a particular one-to-one continuous map is continuous.

A5.5 Theorem. Let $f: \Omega \rightarrow \Omega_1$ where Ω is compact, Ω_1 is Hausdorff, and f is continuous. Then f is a closed map; consequently if f is one-to-one onto, it is a homeomorphism.

PROOF. By A5.3(a), (b) and (c). \blacksquare

A5.6 Corollary. Let $p: \Omega_0 \rightarrow \Omega$ be an identification, and let $h: \Omega_0 \rightarrow \Omega_1$ be continuous; assume $h \circ p^{-1}$ is single valued, and hence continuous by A3.5. Assume also that Ω_0 is compact (hence so is Ω because p is onto) and Ω_1 is Hausdorff. If $h \circ p^{-1}$ is one-to-one onto, it is a homeomorphism.

PROOF. Apply A5.5 with $f = h \circ p^{-1}$. \blacksquare

Corollary A5.6 is frequently applied in constructing quotient spaces. For example, if one pair of opposite edges of a rectangle are identified, we obtain a cylinder. Formally, let $I^2 = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Define an

equivalence relation R on I^2 by specifying that $(0, y)$ be equivalent to $(1, y)$, $0 \leq y \leq 1$, with the other equivalence classes consisting of single points.

Let $h(x, y) = (e^{i2\pi x}, y)$, $(x, y) \in I^2$; h maps I^2 onto a cylinder C . If p is the canonical projection of I^2 onto I^2/R , then A5.6 implies that $h \circ p^{-1}$ is a homeomorphism of I^2/R and C .

In some situations, for example in metric spaces, there are alternative ways of expressing the idea of compactness.

A5.7 Definition. A topological space Ω is said to be *countably compact* iff every countable open covering of Ω has a finite subcover.

A5.8 Theorem. For any topological space Ω , the following properties (a)-(d) are equivalent, and each implies (e). If Ω is T_1 , then all five properties are equivalent.

- (a) Ω is countably compact.
- (b) Each sequence of closed subsets of Ω with the finite intersection property has nonempty intersection.
- (c) Every sequence in Ω has an accumulation point.
- (d) Every countable filterbase in Ω has an accumulation point.
- (e) Every infinite subset of Ω has a cluster point.

PROOF. The equivalence of (a), (b), and (d) is proved exactly as in A5.2, and (d) implies (c) because the tails of a sequence form a countable filterbase. To prove that (c) implies (d), let $\mathcal{A} = \{A_1, A_2, \dots\}$ be a countable filterbase, and choose $x_n \in \bigcap_{i=1}^n A_i$, $n = 1, 2, \dots$. If x is an accumulation point of $\{x_n\}$ and $U \in \mathcal{U}(x)$, then for each n there is an $m \geq n$ such that $x_m \in U$, hence $U \cap \bigcap_{i=1}^m A_i \neq \emptyset$. It follows that $U \cap A_n \neq \emptyset$ for all n , and consequently x is an accumulation point of \mathcal{A} . To prove that (c) implies (e), pick a sequence $\{x_n\}$ of distinct points from the infinite set A and observe that if x is an accumulation point of the sequence, then x is a cluster point of A . Finally, we show that (e) implies (a) if Ω is T_1 . Say U_1, U_2, \dots form a countable open covering of Ω with no finite subcover. Choose $x_1 \notin U_1$; having chosen distinct x_1, \dots, x_k with $x_j \notin \bigcup_{i=1}^j U_i$, $j = 1, \dots, k$, then x_1, \dots, x_k all belong to some finite union $\bigcup_{i=1}^n U_i$, $n \geq k + 1$; choose $x_{k+1} \notin \bigcup_{i=1}^n U_i$ (hence $x_{k+1} \notin \bigcup_{i=1}^{k+1} U_i$ and x_1, \dots, x_{k+1} are distinct). In this way we form an infinite set $A = \{x_1, x_2, \dots\}$ with no cluster point. For if x is such a point, x belongs to U_n for some n . Since Ω is T_1 , there is a set $U \in \mathcal{U}(x)$ such that $U \subset U_n$ and $x_i \notin U$, $i = 1, 2, \dots, n - 1$ (unless $x_i = x$). Since $x_k \notin U_n$ for $k \geq n$, U contains no point of A distinct from x . ■

A5.9 Definitions and Comments. The topological space Ω is said to be *sequentially compact* iff every sequence in Ω has a convergent subsequence. By A5.8(c), sequential compactness implies countable compactness. In a first countable space, countable and sequential compactness are equivalent. For if x is an accumulation point of the sequence $\{x_n\}$ and V_1, V_2, \dots (with $V_{n+1} \subset V_n$ for all n) form a countable base at x , then for each k we may find $n_k \geq k$ such that $x_{n_k} \in V_k$. Thus we have a subsequence converging to x .

A5.10 Theorem. In a second countable space or a metric space, compactness, countable compactness, and sequential compactness are equivalent.

PROOF. A second countable space is Lindelöf (see A4.5), so compactness and countable compactness are equivalent. It is first countable, so countable compactness and sequential compactness are equivalent; this result holds in a metric space also, because a metric space is first countable.

Now a sequentially compact metric space Ω is *totally bounded*, that is, for each $\varepsilon > 0$, Ω can be covered by finitely many balls of radius ε . (If not, inductively pick x_1, x_2, \dots with $x_{n+1} \notin \bigcup_{i=1}^n B(x_i, \varepsilon)$; then $\{x_n\}$ can have no convergent subsequence.) Thus for each positive integer n , Ω can be covered by finitely many balls $B(x_{ni}, 1/n)$, $i = 1, 2, \dots, k_n$. If $\{U_j, j \in J\}$ is an arbitrary open covering of Ω , for each ball $B(x_{ni}, 1/n)$ we choose, if possible, a set U_{ni} of the covering such that $U_{ni} \supset B(x_{ni}, 1/n)$. If $x \in \Omega$, then x belongs to a ball $B(x, \varepsilon)$ included in some U_j ; hence $x \in B(x_{ni}, 1/n) \subset B(x, \varepsilon) \subset U_j$ for some n and i ; therefore $x \in U_{ni}$. Thus the U_{ni} form a countable subcover, and Ω , which is countably compact, must in fact be compact. ■

Note that a compact metric space is Lindelöf, hence (see A4.5) is second countable and separable.

A5.11 Definition. A Hausdorff space is said to be *locally compact* iff each $x \in \Omega$ has a *relatively compact* neighborhood, that is, a neighborhood whose closure is compact. (It follows that a compact Hausdorff space is locally compact.)

A5.12 Theorem. The following are equivalent, for a Hausdorff space Ω :

- (a) Ω is locally compact.
- (b) For each $x \in \Omega$ and $U \in \mathcal{U}(x)$, there is a relatively compact open set V with $x \in V \subset \overline{V} \subset U$. (Thus a locally compact space is regular; furthermore, the relatively compact open sets form a base for the topology.)

- (c) If K is compact, U is open, and $K \subset U$, there is a relatively compact open set V with $K \subset V \subset \overline{V} \subset U$.

PROOF. It is immediate that (c) implies (a), and (b) implies (c) is proved by applying (b) to each point of K and using compactness. To prove that (a) implies (b), let x belong to the open set U . By (a), there is a neighborhood V_1 of x such that $K = \overline{V}_1$ is compact. Now K is compact Hausdorff, and hence regular, and $x \in U \cap V_1$, which is open in Ω , and hence open in K . Thus (see A4.1) there is a set W open in Ω such that $x \in W \cap K$ and the closure of $W \cap K$ in K , namely, $\overline{W \cap K}$, is a subset of $U \cap V_1$.

Now $x \in W \cap V_1$ and $\overline{W \cap V_1} \subset \overline{W \cap K} \subset U$, so $V = W \cap V_1$ is the desired relatively compact neighborhood. ■

The following properties of locally compact spaces are often useful:

A5.13 Theorem. (a) Let Ω be a locally compact Hausdorff space. If $K \subset U \subset \Omega$, with K compact and U open, there is a continuous $f: \Omega \rightarrow [0, 1]$ such that $f = 0$ on K and $f = 1$ on $\Omega - U$. In particular, a locally compact Hausdorff space is completely regular.

(b) Let Ω be locally compact Hausdorff, or, more generally, completely regular. If A and B are disjoint subsets of Ω with A compact and B closed, there is a continuous $f: \Omega \rightarrow [0, 1]$ such that $f = 0$ on A and $f = 1$ on B .

(c) Let Ω be locally compact Hausdorff, and let $A \subset U \subset \Omega$, with A compact and U open. Then there are sets B and V with $A \subset V \subset B \subset U$, where V is open and σ -compact (a countable union of compact sets) and B is compact and is also a G_δ (a countable intersection of open sets). Consequently (take $A = \{x\}$) the σ -compact open sets form a base for the topology.

PROOF. (a) Let $K \subset V \subset \overline{V} \subset U$, with V open, \overline{V} compact [see A5.12(c)]. \overline{V} is normal, so there is a continuous $g: \overline{V} \rightarrow [0, 1]$ with $g = 0$ on K , $g = 1$ on $\overline{V} - V$. Define $f = g$ on \overline{V} , $f = 1$ on $\Omega - V$. (On $\overline{V} - V$, $g = 1$ so f is well-defined.) Now f is continuous on Ω (look at preimages of closed sets), so f is the desired function.

(b) By complete regularity, for each $x \in A$, there is a continuous $f_x: \Omega \rightarrow [0, 1]$ with $f_x(x) = 0$, $f_x = 1$ on B . By compactness,

$$A \subset \bigcup_{i=1}^n \{x: f_{x_i}(x) < \frac{1}{2}\}$$

for some x_1, \dots, x_n . Let $g = \prod_{i=1}^n f_{x_i}$; then $g = 1$ on B and $0 \leq g < \frac{1}{2}$ on A , so $f = \max(2g - 1, 0)$ is the desired function.

(c) By A.5.12(c), we may assume without loss of generality that U is included in a compact set K . By (b) there is a continuous $f: \Omega \rightarrow [0, 1]$ with $f = 0$ on A , $f = 1$ off U . Let $V = \{x: f(x) < \frac{1}{2}\}$, $B = \{x: f(x) \leq \frac{1}{2}\}$. Then V is open, B is closed, and $A \subset V \subset B \subset U \subset K$; hence B is compact. Now B is a G_δ because

$$B = \bigcap_{n=1}^{\infty} \left\{ x: f(x) < \frac{1}{2} + \frac{1}{n} \right\}, \quad \text{and} \quad V = \bigcup_{n=1}^{\infty} \left\{ x: f(x) \leq \frac{1}{2} - \frac{1}{n} \right\} \subset K,$$

so V is σ -compact. (In fact V is a countable union of compact G_δ 's.) ■

A5.14 Corollary. Let Ω be completely regular, A a compact G_δ subset of Ω . There is a continuous $f: \Omega \rightarrow [0, 1]$ such that $f^{-1}\{0\} = A$.

PROOF. If $A = \bigcap_{n=1}^{\infty} U_n$, the U_n open, by A5.13(b) there are continuous $f_n: \Omega \rightarrow [0, 1]$ with $f_n = 0$ on A and $f_n = 1$ off U_n . Let $f = \sum_{n=1}^{\infty} f_n 2^{-n}$. ■

A5.15 Theorem. If Ω is Hausdorff, the following are equivalent:

- (a) Ω is locally compact Lindelöf.
- (b) Ω can be expressed as $\bigcup_{n=1}^{\infty} U_n$ where \bar{U}_n is a compact subset of U_{n+1} for each n .
- (c) Ω is locally compact and σ -compact.

PROOF. (a) implies (b): The relatively compact open sets form a base by A5.12(b), in particular they cover Ω . Extract a countable subcover $\{V_1, V_2, \dots\}$ using Lindelöf, and let $U_1 = V_1$, and for $n > 1$, $U_n = V_n \cup W_n$ where W_n is a relatively compact open set and $W_n \supset \bigcup_{i=1}^{n-1} \bar{U}_i$.

(b) implies (c): We have $\Omega = \bigcup_{n=1}^{\infty} \bar{U}_n$, proving σ -compactness. If $x \in \Omega$, then $x \in U_n$ for some n , with \bar{U}_n compact, proving local compactness.

(c) implies (a): If $\Omega = \bigcup_{n=1}^{\infty} K_n$, K_n compact, and $\{U_i\}$ is an open covering of Ω , extract a finite subcovering of each K_n , and put the sets together to form a countable subcovering of Ω . ■

The *Urysohn metrization theorem*, which we shall not prove, states that for a second countable space Ω , metrizability and regularity are equivalent. This result yields the following corollary to A5.15.

A5.16 Theorem. If Ω is locally compact Hausdorff, then Ω is second countable iff Ω is metrizable and σ -compact.

Proof. If Ω is second countable, it is metrizable by A5.12(b) and the Urysohn metrization theorem. Also, Ω is Lindelöf (see A4.5), hence is σ -compact by A5.15. Conversely, if Ω is metrizable and σ -compact, then Ω is Lindelöf by A5.15, hence second countable (see A4.5). ■

Finally, we consider the important *one-point compactification*.

A5.17 Theorem. Let Ω be locally compact Hausdorff, and let $\Omega^* = \Omega \cup \{\infty\}$, where ∞ stands for any element not belonging to Ω . Put the following topology on Ω^* : U is open in Ω^* iff U is open in Ω or U is the complement (in Ω^*) of a compact subset of Ω . Then:

- (a) If $U \subset \Omega$, U is open in Ω iff U is open in Ω^* ; thus Ω^* induces the original topology on Ω .
- (b) Ω^* is compact Hausdorff.

PROOF. Part (a) follows from the definition of the topology. To prove (b), let $\{U_i\}$ be an open covering of Ω^* . Then ∞ belongs to some U_i , and the remaining U_j cover the compact set $\Omega^* - U_i$, so the cover reduces to a finite subcover. Since Ω is Hausdorff, distinct points of Ω have disjoint neighborhoods. If $x \in \Omega$, let V be an open subset of Ω with $x \in V$ and $\bar{V} \subset \Omega$, where \bar{V} , the closure of V in Ω , is compact [see A 5.12(b)]. Then V and $\Omega^* - \bar{V}$ are disjoint neighborhoods of x and ∞ . ■

A6 Semicontinuous Functions

If f_1, f_2, \dots are continuous maps from the topological space Ω to the extended reals \bar{R} , and $f_n(x)$ increases to a limit $f(x)$ for each x , f need not be continuous; however, f is lower semicontinuous. Functions of this type play an important role in many aspects of analysis and probability.

A6.1 Definition. Let Ω be a topological space. The function $f: \Omega \rightarrow \bar{R}$ is said to be *lower semicontinuous* (LSC) on Ω iff $\{x \in \Omega: f(x) > a\}$ is open in Ω for each $a \in \bar{R}$, *upper semicontinuous* (USC) on Ω iff $\{x \in \Omega: f(x) < a\}$ is open in Ω for each $a \in \bar{R}$. Thus f is LSC iff $-f$ is USC. Note that f is continuous iff it is both LSC and USC.

We have the following criterion for semicontinuity.

A6.2 Theorem. The function f is LSC on Ω iff, for each net $\{x_n\}$ converging to a point $x \in \Omega$, we have $\liminf_n f(x_n) \geq f(x)$, where $\liminf_n f(x_n)$ means

$\sup_n \inf_{k \geq n} f(x_k)$. Hence f is USC iff $\lim \sup_n f(x_n) \leq f(x)$ whenever $x_n \rightarrow x$. (In a first countable space, “net” may be replaced by “sequence.”)

PROOF. Let f be LSC. If $x_n \rightarrow x$ and $b < f(x)$, then $x \in f^{-1}(b, \infty]$, an open subset of Ω , hence eventually $x_n \in f^{-1}(b, \infty]$, that is, $f(x_n) > b$ eventually. Thus $\liminf_n f(x_n) \geq f(x)$. Conversely if $x_n \rightarrow x$ implies

$$\liminf_n f(x_n) \geq f(x),$$

we show that $V = \{x : f(x) > a\}$ is open. Let $x_n \rightarrow x$, where $f(x) > a$. Then $\liminf_n f(x_n) > a$, hence $f(x_n) > a$ eventually, that is, $x_n \in V$ eventually. Thus (see A2.4) V is open. ■

We now prove a few properties of semicontinuous functions.

A6.3 Theorem. Let f be LSC on the compact space Ω . Then f attains its infimum. (Hence if f is USC on the compact space Ω , f attains its supremum.)

PROOF. If $b = \inf f$, there is a sequence of points $x_n \in \Omega$ with $f(x_n) \rightarrow b$. By compactness, we have a subnet x_{n_k} converging to some $x \in \Omega$. Since f is LSC, $\liminf_k f(x_{n_k}) \geq f(x)$. But $f(x_{n_k}) \rightarrow b$, so that $f(x) \leq b$; consequently $f(x) = b$. ■

A6.4 Theorem. If f_i is LSC on Ω for each $i \in I$, then $\sup_i f_i$ is LSC; if I is finite, then $\min_i f_i$ is LSC. (Hence if f_i is USC for each i , then $\inf_i f_i$ is USC, and if I is finite, then $\max_i f_i$ is USC.)

PROOF. Let $f = \sup_i f_i$; then $\{x : f(x) > a\} = \bigcup_{i \in I} \{x : f_i(x) > a\}$; hence $\{x : f(x) > a\}$ is open. If $g = \min(f_1, f_2, \dots, f_n)$, then

$$\{x : g(x) > a\} = \bigcap_{i=1}^n \{x : f_i(x) > a\}$$

is open. ■

A6.5 Theorem. Let $f: \Omega \rightarrow \bar{\mathbb{R}}$, Ω any topological space, f arbitrary. Define

$$f(x) = \liminf_{y \rightarrow x} f(y), \quad x \in \Omega;$$

that is,

$$f(x) = \sup_V \inf_{y \in V} f(y),$$

where V ranges over all neighborhoods of x . [If Ω is a metric space, then $f(x) = \sup_{n=1, 2, \dots} \inf_{d(x, y) < 1/n} f(y)$.]

Then f is LSC on Ω and $f \leq f$; furthermore if g is LSC on Ω and $g \leq f$, then $g \leq f$. Thus f , called the *lower envelope* of f , is the sup of all LSC functions that are less than or equal to f (there is always at least one such function, namely the function constant at $-\infty$).

Similarly, if $\bar{f}(x) = \limsup_{y \rightarrow x} f(y) = \inf_V \sup_{y \in V} f(y)$, then \bar{f} , the *upper envelope* of f , is USC and $\bar{f} \geq f$; in fact \bar{f} is the inf of all USC functions that are greater than or equal to f .

PROOF. It suffices to consider f . Let $\{x_n\}$ be a net in Ω with $x_n \rightarrow x$ and $\liminf_n f(x_n) < b < f(x)$. If V is a neighborhood of x , we can choose n such that $x_n \in V$ and $f(x_n) < b$. Since V is also a neighborhood of x_n , we have

$$b > f(x_n) \geq \inf_{y \in V} f(y),$$

so

$$f(x) = \sup_V \inf_{y \in V} f(y) \leq b < f(x),$$

a contradiction. By A6.2, f is LSC, and $f \leq f$ by definition of f . Finally if g is LSC, $g \leq f$, then $f(x) = \liminf_{y \rightarrow x} f(y) \geq \liminf_{y \rightarrow x} g(y) \geq g(x)$ since g is LSC. [If $\sup_V \inf_{y \in V} g(y) < b < g(x)$, then for each V pick $x_V \in V$ with $g(x_V) < b$. If $V_1 \leq V_2$ means that $V_2 \subset V_1$, the x_V form a net converging to x , while $\liminf_V g(x_V) \leq b < g(x)$, contradicting A6.2.] ■

It can be shown that if Ω is completely regular, every LSC function on Ω is the sup of a family of continuous functions. If Ω is a metric space, the family can be assumed countable, as we now prove.

A6.6 Theorem. Let Ω be a metric space, f a LSC function on Ω . There is a sequence of continuous functions $f_n: \Omega \rightarrow \bar{\mathbb{R}}$ such that $f_n \uparrow f$. (Thus if f is USC, there is a sequence of continuous functions $f_n \downarrow f$.) If $|f| \leq M < \infty$, the f_n may be chosen so that $|f_n| \leq M$ for all n .

PROOF [following Hausdorff (1962)]. First assume $f \geq 0$ and finite-valued. If d is the metric on Ω , define $g(x) = \inf\{f(z) + td(x, z): z \in \Omega\}$, where $t > 0$ is fixed; then $0 \leq g \leq f$ since $g(x) \leq f(x) + td(x, x) = f(x)$.

If $x, y \in \Omega$, then $f(z) + td(x, z) \leq f(z) + td(y, z) + td(x, y)$. Take the inf over z to obtain $g(x) \leq g(y) + td(x, y)$. By symmetry,

$$|g(x) - g(y)| \leq td(x, y),$$

hence g is continuous on Ω .

Now set $t = n$; in other words let $f_n(x) = \inf\{f(z) + nd(x, z) : z \in \Omega\}$. Then $0 \leq f_n \uparrow h \leq f$. But given $\varepsilon > 0$, for each n we can choose $z_n \in \Omega$ such that

$$f_n(x) + \varepsilon > f(z_n) + nd(x, z_n) \geq nd(x, z_n).$$

But $f_n(x) + \varepsilon \leq f(x) + \varepsilon$, and it follows that $d(x, z_n) \rightarrow 0$. Since f is LSC, $\liminf_{n \rightarrow \infty} f(z_n) \geq f(x)$; thus $f(z_n) > f(x) - \varepsilon$ eventually. But now

$$\begin{aligned} f_n(x) &> f(z_n) - \varepsilon + nd(x, z_n) \geq f(z_n) - \varepsilon \\ &> f(x) - 2\varepsilon \quad \text{for large enough } n. \end{aligned}$$

It follows that $0 \leq f_n \uparrow f$. If $|f| \leq M < \infty$, then $f + M$ is LSC and nonnegative; if $0 \leq g_n \uparrow f + M$, then $f_n = g_n - M \uparrow f$ and $|f_n| \leq M$.

In general, let $h(x) = \frac{1}{2}\pi + \arctan x$, $x \in \bar{R}$; then h is an order-preserving homeomorphism of \bar{R} and $[0, \pi]$. If f is LSC, then $h \circ f$ is finite-valued, LSC, and nonnegative, so that we can find continuous functions g_n such that $g_n \uparrow h \circ f$. Let $f_n = h^{-1} \circ g_n$; then $f_n \uparrow f$. ■

A7 The Stone-Weierstrass Theorem

In this section, we consider the family $C(\Omega)$ of continuous *real-valued* functions on a compact Hausdorff space Ω . The set $C(\Omega)$ becomes a Banach space under the sup norm $\|f\| = \sup\{|f(x)| : x \in \Omega\}$. If $A \subset C(\Omega)$, A is called a *subalgebra* of $C(\Omega)$ iff for all $f_1, f_2 \in A$, $a, b \in R$, we have $af_1 + bf_2 \in A$ and $f_1 f_2 \in A$. Thus A is a vector space under addition and scalar multiplication, and a ring under ordinary multiplication. The main theorem of this section gives conditions under which A is dense in $C(\Omega)$. We assume throughout that Ω has at least two points; the Stone-Weierstrass theorem will be seen to be trivial if Ω has only a single point.

A7.1 Lemma. The function given by $f(x) = |x|$, $-1 \leq x \leq 1$, can be uniformly approximated by polynomials having no constant term.

PROOF. If $0 \leq y \leq a \leq 1$, then $0 \leq y \leq y + \frac{1}{2}(a^2 - y^2) \leq a$. (For $y^2 - 2y + 2a - a^2$ decreases when $0 \leq y \leq 1$, and its value when $y = a$ is 0.) If we define $y_0 = 0$, $y_{n+1} = y_n + \frac{1}{2}(a^2 - y_n^2)$, the above argument shows that $0 \leq y_n \leq y_{n+1} \leq a$ for all n . Let $n \rightarrow \infty$ to obtain $y_n \rightarrow y$, where $y = y + \frac{1}{2}(a^2 - y^2)$; hence $y = a$.

Thus, changing the notation, define

$$p_0(x) = 0, p_{n+1}(x) = p_n(x) + \frac{1}{2}(x^2 - p_n^2(x)), \quad -1 \leq x \leq 1.$$

By induction, p_n is a polynomial having no constant term, and $0 \leq p_n(x) \leq p_{n+1}(x) \rightarrow (x^2)^{1/2} = |x|$. Since the domain $[-1, 1]$ is compact and p_n converges monotonically to a continuous limit p , the convergence is uniform by Dini's theorem (see 4.3.9 for the details). ■

A7.2 Lemma. Let A be a subset of $C(\Omega)$ that is closed under the lattice operations; that is, if $f, g \in A$, then $\max(f, g) \in A$ and $\min(f, g) \in A$.

If $f \in C(\Omega)$ and f can be approximated at each pair of points by functions in A [if $x \neq y$, there is a sequence of functions $f_n \in A$ with $f_n(x) \rightarrow f(x)$ and $f_n(y) \rightarrow f(y)$], then $f \in \bar{A}$.

PROOF. Given $\delta > 0$, $x, y \in \Omega$, there is a function $f_{xy} \in A$ with $|f - f_{xy}| < \delta$ at the points x and y . Let $U(x, y) = \{z \in \Omega : f_{xy}(z) < f(z) + \delta\}$, $V(x, y) = \{z \in \Omega : f_{xy}(z) > f(z) - \delta\}$. Since $x \in U(x, y)$, the $U(x, y)$ cover Ω for each fixed y ; by compactness, $\Omega = \bigcup_{i=1}^n U(x_i, y)$ for some x_1, \dots, x_n .

Let $f_y = \min\{f_{x_i y} : 1 \leq i \leq n\} \in A$ by hypothesis, and let

$$V(y) = \bigcap_{i=1}^n V(x_i, y).$$

If $x \in \Omega$, then $x \in U(x_i, y)$ for some i , so $f_y(x) \leq f_{x_i y}(x) < f(x) + \delta$; if $x \in V(y)$, then $f_{x_i y}(x) > f(x) - \delta$ for all i , hence $f_y(x) > f(x) - \delta$.

But the $V(y)$ cover Ω [note $y \in V(y)$] so by compactness, $\Omega = \bigcup_{i=1}^m V(y_i)$ for some y_1, \dots, y_m . Define $f_\delta = \max\{f_{y_i} : 1 \leq i \leq m\} \in A$ by hypothesis.

If $x \in \Omega$, then $f_\delta(x) < f(x) + \delta$; since $x \in V(y_i)$ for some i , we have $f_\delta(x) > f(x) - \delta$. Thus $\|f - f_\delta\| \leq \delta$ and, since δ is arbitrary, it follows that $f \in \bar{A}$. ■

A7.3 Lemma. Let A be a subalgebra of $C(\Omega)$, and assume that A is a closed subset of $C(\Omega)$. Then $f \in A$ implies $|f| \in A$; also, $f, g \in A$ implies $\max(f, g) \in A$ and $\min(f, g) \in A$.

PROOF. If $g \in A$ and p is a polynomial with no constant term, then $p \circ g \in A$. If $f \in A$ and $\|f\| = M$, set $g = f/M$. Then $|g(x)| \leq 1$ for all x , so by A7.1, $\|p \circ g - |g|\|$ can be made arbitrarily small by appropriate choice of p ; consequently $|g|$, hence $|f|$, belongs to A .

Now $f + g = \max(f, g) + \min(f, g)$, $|f - g| = \max(f, g) - \min(f, g)$. Addition and then subtraction of these equations completes the proof. ■

A7.4 Lemma. Let A be a subalgebra of $C(\Omega)$ that separates points; that is, if $x \neq y$, there is an $f \in A$ with $f(x) \neq f(y)$. Assume that for each $x \in \Omega$, there is an $f \in A$ with $f(x) \neq 0$.

If $x, y \in \Omega$, $x \neq y$, and r and s are arbitrary real numbers, there is an $f \in A$ with $f(x) = r, f(y) = s$.

PROOF. If this is not the case, then for all $f_1, f_2 \in A$, the equations

$$af_1(x) + bf_2(x) = r \quad \text{and} \quad af_1(y) + bf_2(y) = s$$

have no solution for a and b . Thus the determinant $f_1(x)f_2(y) - f_1(y)f_2(x)$ is 0. In particular, if $f_1 = f, f_2 = f^2$, we obtain $f(x)f^2(y) = f(y)f^2(x)$, or $f(x)f(y)(f(y) - f(x)) = 0$ for all $f \in A$. Thus if $g \in A$ and $g(x) \neq g(y)$, we have $g(x)g(y) = 0$. If, say, $g(x) = 0$, then (taking $f_1 = g$ in the above determinant) $g(x)f_2(y) = g(y)f_2(x)$ for all $f_2 \in A$. Since $g(x) = 0$ we must have $g(y) \neq 0$; hence $f_2(x) = 0$ for all $f_2 \in A$, contradicting the hypothesis. The argument is symmetrical if $g(y) = 0$. ■

A7.5 Stone-Weierstrass Theorem. Let A be a subalgebra of $C(\Omega)$ that separates points.

- (a) If for each $x \in \Omega$, there is an $f \in A$ with $f(x) \neq 0$ (in particular, if the constant functions belong to A), then $\bar{A} = C(\Omega)$; thus A is (uniformly) dense in $C(\Omega)$.
- (b) If all $f \in A$ vanish at a particular $x \in \Omega$ (there can be at most one such point since A separates points), then $\bar{A} = \{g \in C(\Omega) : g(x) = 0\}$.

PROOF. (a) Since A is an algebra, so is \bar{A} ; by A7.3, \bar{A} is closed under the lattice operations. Let $g \in C(\Omega)$; if $x, y \in \Omega$, $x \neq y$, then by A7.4 there is an $f \in A \subset \bar{A}$ such that $f(x) = r = g(x)$ and $f(y) = s = g(y)$. Thus g can be approximated at each pair of points by functions in \bar{A} . By A7.2, $g \in \bar{A}$; in other words, $\bar{A} = C(\Omega)$.

(b) If $f(x) = 0$ for all $f \in C(\Omega)$, let B be the algebra generated by A and the constant functions; thus $B = \{f + c : f \in A, c \text{ a constant function}\}$. Then B satisfies the hypothesis of (a), so $\bar{B} = C(\Omega)$. Now if $\delta > 0$, $g \in C(\Omega)$, and $g(x) = 0$, then for some $f \in A$ and some c we have $\|g - (f + c)\| < \delta/2$. But $g(x) = f(x) = 0$; hence $|c| < \delta/2$, and therefore $\|g - f\| < \delta$. Thus $g \in \bar{A}$, so that $\{g \in C(\Omega) : g(x) = 0\} \subset \bar{A}$. But the reverse inclusion holds by hypothesis, completing the proof. ■

A7.6 Corollary. The Stone-Weierstrass theorem holds equally well with $C(\Omega)$ taken as the collection of all continuous *complex-valued* functions on Ω , if we add the hypothesis that whenever $f \in A$, the complex conjugate \bar{f} also belongs to A .

PROOF. (a) Let $B \subset A$ be the set of functions in A that assume only real values. If $x \neq y$, there is an $f = f_1 + if_2 \in A$ with $f(x) \neq f(y)$. Thus either $f_1(x) \neq f_1(y)$ or $f_2(x) \neq f_2(y)$. Now $f_1 = (f + \bar{f})/2$ and $f_2 = (f - \bar{f})/2i$ belong to A by hypothesis; hence $f_1, f_2 \in B$ since f_1 and f_2 are real-valued. Thus B separates points. Furthermore, if $x \in \Omega$, then $f(x) \neq 0$ for some $f \in A$; but $|f|^2 = f\bar{f} \in B$, so B satisfies the hypothesis of A7.5(a). If $f = f_1 + if_2 \in C(\Omega)$, then f_1 and f_2 can be uniformly approximated by functions in B ; hence f can be uniformly approximated by functions in A .

(b) This follows from (a) exactly as in A7.5. ■

If Ω is the unit disk in the complex plane and A is the collection of all analytic functions on Ω , then A is a closed subalgebra of $C(\Omega)$ that separates points and contains the constant functions; but $\bar{A} \neq C(\Omega)$. Thus the hypothesis $f \in A$ implies $\bar{f} \in A$ cannot be dropped in A7.6.

A8 Topologies on Function Spaces

In this section we examine spaces consisting of functions from an arbitrary set Ω to another set Ω_1 . Some structure will be imposed on Ω_1 ; it can be a metric space or, more generally, a gauge space, which we now define.

A8.1 Definitions and Comments. A *gauge space* is a space Ω_1 whose topology is determined by a family $\mathcal{D} = \mathcal{D}(\Omega_1)$ of pseudometrics; thus a subbase for the topology is formed by the sets $B_d(x, \delta) = \{y \in \Omega_1 : d(x, y) < \delta\}$, $x \in \Omega_1$, $\delta > 0$, $d \in \mathcal{D}$. [A pseudometric has all the properties of a metric except that $d(x, y)$ can be 0 for $x \neq y$.] Convergence of x_n to x in Ω_1 means $d(x_n, x) \rightarrow 0$ for all $d \in \mathcal{D}$.

Since $B_{d_1}(x, \delta) \cap B_{d_2}(x, \delta) = B_d(x, \delta)$, where d is the pseudometric $\max(d_1, d_2)$, it follows that the sets $B_{d^+}(x, \delta)$, $d^+ = \max(d_{i_1}, \dots, d_{i_n})$, $n = 1, 2, \dots$, the $d_{i_k} \in \mathcal{D}$, form a base for the topology. Denote by $\mathcal{D}^+ = \mathcal{D}^+(\Omega_1)$ the collection of all pseudometrics d^+ .

The gauge space Ω_1 is Hausdorff iff whenever $x \neq y$, there is a $d \in \mathcal{D}$ with $d(x, y) \neq 0$; in this case, \mathcal{D} is said to be *separating*.

The Hausdorff gauge spaces coincide with the completely regular spaces. To see this, let x belong to the open set U in the gauge space Ω_1 . If $d \in \mathcal{D}^+$ and $B_d(x, \delta) \subset U$, let $f(y) = \min(1, d(x, y)/\delta)$. Then f is continuous, $0 \leq f \leq 1$, $f(x) = 0$, and $f = 1$ off U . Conversely, if Ω_1 is completely regular, it can be embedded in a product of closed unit intervals (see A4.6). Now a product of gauge spaces is a gauge space [take $d(x, y) = d_i(x_i, y_i)$, where the d_i are

pseudometrics for the i th coordinate space], and a subspace of a gauge space is a gauge space, and the result follows.

If \mathcal{D} consists of countably many pseudometrics d_1, d_2, \dots , let

$$d = \sum_{n=1}^{\infty} \min(d_n, 2^{-n});$$

the single pseudometric d includes the topology of Ω_1 ; hence Ω_1 is pseudo-metrizable (metrizable if Ω_1 is Hausdorff).

We now discuss function space topologies.

A8.2 Definitions and Comments. Let $F(\Omega, \Omega_1)$ be the collection of all functions from the set Ω to the gauge space Ω_1 . If \mathcal{A} is a family of subsets of Ω , the topology $\mathcal{T}_{\mathcal{A}}$ of uniform convergence on members of \mathcal{A} has as subbase the sets

$$\{g \in F(\Omega, \Omega_1) : \sup_{x \in A} d(g(x), f(x)) < \delta\},$$

$$f \in F(\Omega, \Omega_1), \quad A \in \mathcal{A}, \quad \delta > 0, \quad d \in \mathcal{D}.$$

Convergence relative to $\mathcal{T}_{\mathcal{A}}$ means uniform convergence on each $A \in \mathcal{A}$.

For example, if $\mathcal{A} = \{\Omega\}$, we obtain the topology of uniform convergence on Ω , also called the *uniform topology*; if Ω is a topological space and \mathcal{A} consists of all compact subsets, we obtain the topology of uniform convergence on compact sets. If \mathcal{A} is the collection of all singletons $\{x\}$, $x \in \Omega$, we have the topology of pointwise convergence.

Adaptations of standard proofs in metric spaces show the following:

A8.3 Theorem. If $C(\Omega, \Omega_1)$ is the collection of all continuous maps from the topological space Ω to the gauge space Ω_1 , then $C(\Omega, \Omega_1)$ is closed in $F(\Omega, \Omega_1)$ relative to the topology of uniform convergence on Ω . In other words, a uniform limit of continuous functions is continuous.

PROOF. If $f_n \in C(\Omega, \Omega_1)$, $f_n \rightarrow f$ uniformly on Ω , $x_0 \in \Omega$, and $d \in \mathcal{D}$, then

$$d(f(x), f(x_0)) \leq d(f(x), f_n(x)) + d(f_n(x), f_n(x_0)) + d(f_n(x_0), f(x_0)).$$

The result follows just as in the metric space proof. ■

A8.4 Theorem. If f is a continuous mapping from the compact gauge space Ω to the gauge space Ω_1 , then f is *uniformly continuous*; that is, if $d \in \mathcal{D}(\Omega_1)$ and $\varepsilon > 0$, there is a $d^+ \in \mathcal{D}^+(\Omega)$ and a $\delta > 0$ such that if $x_1, x_2 \in \Omega$ and $d^+(x_1, x_2) < \delta$, then $d(f(x_1), f(x_2)) < \varepsilon$.

PROOF. If f is not uniformly continuous, there is an $\varepsilon > 0$ and $e \in \mathcal{D}(\Omega_1)$ such that for all $\delta > 0$, $d^+ \in \mathcal{D}^+(\Omega)$, there are points $x, y \in \Omega$ with $d^+(x, y) < \delta$ but $e(f(x), f(y)) \geq \varepsilon$. If we choose such (x, y) for each $\delta > 0$ and $d \in \mathcal{D}(\Omega)$, we obtain a net (x_n, y_n) with $d(x_n, y_n) \rightarrow 0$ but $e(f(x_n), f(y_n)) \geq \varepsilon$ for all n . (Take $(\delta_1, d) \geq (\delta_2, d')$ iff $\delta_1 \leq \delta_2$.) By compactness we find a subnet with $(x_{n_k}, y_{n_k}) \rightarrow (x_0, y_0)$ for some $x_0, y_0 \in \Omega$; thus $f(x_{n_k}) \rightarrow f(x_0), f(y_{n_k}) \rightarrow f(y_0)$ by continuity. But $d(x_{n_k}, y_{n_k}) \rightarrow 0$, so $d(x_0, y_0) = 0$ and consequently, $e(f(x_0), f(y_0)) = 0$ by continuity of f , a contradiction. ■

We now prove a basic compactness theorem in function spaces.

A8.5 Arzela–Ascoli Theorem. Let Ω be a compact topological space, Ω_1 a Hausdorff gauge space, and $G \subset C(\Omega, \Omega_1)$, with the uniform topology. Then G is compact iff the following three conditions are satisfied:

- (a) G is closed,
- (b) $\{g(x) : g \in G\}$ is a relatively compact subset of Ω_1 for each $x \in \Omega$, and
- (c) G is equicontinuous at each point of Ω ; that is, if $\varepsilon > 0$, $d \in \mathcal{D}(\Omega_1)$, $x_0 \in \Omega$, there is a neighborhood U of x_0 such that if $x \in U$, then

$$d(g(x), g(x_0)) < \varepsilon$$

for all $g \in G$.

PROOF. We first note two facts about equicontinuity.

(1) If $M \subset F(\Omega, \Omega_1)$, where Ω is a topological space and Ω_1 is a gauge space, and M is equicontinuous at x_0 , the closure of M in the topology of pointwise convergence is also equicontinuous at x_0 .

(2) If M is equicontinuous at all $x \in \Omega$, then on M , the topology of pointwise convergence coincides with the topology of uniform convergence on compact subsets.

To prove (1), let $f_n \in M, f_n \rightarrow f$ pointwise; if $d \in \mathcal{D}(\Omega_1)$, we have

$$d(f(x), f(x_0)) \leq d(f(x), f_n(x)) + d(f_n(x), f_n(x_0)) + d(f_n(x_0), f(x_0)).$$

If $\delta > 0$, the third term on the right will eventually be less than $\delta/3$ by the pointwise convergence, and the second term will be less than $\delta/3$ for x in some neighborhood U of x_0 , by equicontinuity. If $x \in U$, the first term is eventually less than $\delta/3$ by pointwise convergence, and the result follows.

To prove (2), let $f_n \in M, f_n \rightarrow f$ pointwise, and let K be a compact subset of Ω ; fix $\delta > 0$ and $d \in \mathcal{D}(\Omega_1)$. If $x \in K$, equicontinuity yields a neighborhood

$U(x)$ such that $y \in U(x)$ implies $d(f_n(y), f_n(x)) < \delta/3$ for all n . By compactness, $K \subset \bigcup_{i=1}^r U(x_i)$ for some x_1, \dots, x_r . Then

$$d(f(x), f_n(x)) \leq d(f(x), f(x_i)) + d(f(x_i), f_n(x_i)) + d(f_n(x_i), f_n(x)).$$

If $x \in K$, then $x \in U(x_i)$ for some i ; thus the third term on the right is less than $\delta/3$ for all n , so that the first term is less than or equal to $\delta/3$. The second term is eventually less than $\delta/3$ by pointwise convergence, and it follows that $f_n \rightarrow f$ uniformly on K .

Now assume (a)–(c) hold. Since $G \subset \overline{\prod_{x \in \Omega} \{g(x) : g \in G\}}$, which is pointwise compact by (b) and the Tychonoff theorem A5.4, the pointwise closure G_0 of G is pointwise compact. Thus if $\{g_n\}$ is a net in G , there is a subnet converging pointwise to some $g \in G_0$. By (c) and (1), G_0 is equicontinuous at each point of Ω ; hence by (2), the subnet converges uniformly to g . But g is continuous by A8.3; hence $g \in G$ by (a).

Conversely, assume G compact. Since Ω_1 is Hausdorff, so is $C(\Omega, \Omega_1)$ [as well as $F(\Omega, \Omega_1)$], hence G is closed, proving (a). The map $g \rightarrow g(x)$ of G into Ω_1 is continuous, and (b) follows from A5.3(a). Finally, if G is not equicontinuous at x , there is an $\varepsilon > 0$ and a $d \in \mathcal{D}(\Omega_1)$ such that for each neighborhood U of x there is an $x_U \in U$ and $g_U \in G$ with $d(g_U(x), g_U(x_U)) \geq \varepsilon$. If $U \geq V$ means $U \subset V$, the g_U form a net in G , so there is a subnet converging uniformly to a limit $g \in G$. But $x_U \rightarrow x$; hence $g_U(x_U) \rightarrow g(x)$, a contradiction.

[The last step follows from the fact that the map $(x, g) \rightarrow g(x)$ of $\Omega \times G$ into Ω_1 is continuous. To see this, let $x_n \rightarrow x$, and $g_n \rightarrow g$ uniformly on Ω ; if $d \in \mathcal{D}(\Omega_1)$, then

$$d(g_n(x_n), g(x)) \leq d(g_n(x_n), g(x_n)) + d(g(x_n), g(x)).$$

The first term approaches 0 by uniform convergence, and the second term approaches 0 by continuity of g .] ■

If G satisfies (b) and (c) of A8.5, but not necessarily (a), the closure of G in the uniform topology satisfies all three hypotheses, and hence is compact. In the special case when Ω_1 is the set of complex numbers, $C(\Omega, \Omega_1) = C(\Omega)$ is a Banach space, and in particular, compactness and sequential compactness are equivalent. We thus obtain the most familiar form of the Arzela–Ascoli theorem: If f_1, f_2, \dots is a sequence of continuous complex-valued functions on the compact space Ω , and if the f_n are pointwise bounded and equicontinuous, there is a uniformly convergent subsequence. In fact the f_n must be uniformly bounded by the continuity of the map $(x, g) \rightarrow g(x)$ referred to above.

A9 Complete Metric Spaces and Category Theorems

A9.1 Definitions and Comments. A metric space (or the associated metric d) is said to be *complete* iff each Cauchy sequence (that is, each sequence such that $d(x_n, x_m) \rightarrow 0$ as $n, m \rightarrow \infty$) converges to a point in the space.

Any compact metric space is complete since a Cauchy sequence with a convergent subsequence converges.

It is important to recognize that *completeness is not a topological property*. In other words, two metrics, only one of which is complete, may be equivalent, that is, induce the same topology. As an example, the chordal metric d' on the complex plane [$d'(z_1, z_2)$ is the Euclidean distance between the stereographic projections of z_1 and z_2 on the Riemann sphere] is not complete, but the equivalent Euclidean metric d is complete. Note also that the sequence $z_n = n$ is d' -Cauchy but not d -Cauchy, so the notion of a Cauchy sequence is not topological.

The subset A of the topological space Ω is said to be *nowhere dense* iff the interior of \bar{A} is empty, in other words, iff the complement of \bar{A} is dense. A set $B \subset \Omega$ is said to be of *category 1 in Ω* iff B can be expressed as a countable union of nowhere dense subsets of Ω ; otherwise B is of *category 2 in Ω* .

The following result is the best known category theorem.

A9.2 Baire Category Theorem. Let (Ω, d) be a complete metric space. If A_n is closed in Ω for each $n = 1, 2, \dots$, and $\bigcup_{n=1}^{\infty} A_n = \Omega$, then the interior A_n^0 is nonempty for some n . Therefore Ω is of category 2 in itself.

PROOF. Assume $A_n^0 = \emptyset$ for all n . Then $A_1 \neq \Omega$, and since $\Omega - A_1$ is open, there is a ball $B(x_1, \delta_1) \subset \Omega - A_1$ with $0 < \delta_1 < \frac{1}{2}$. Now $B(x_1, \frac{1}{2}\delta_1) \not\subset A_2$ since $A_2^0 = \emptyset$; therefore $B(x_1, \frac{1}{2}\delta_1) - A_2$ is a nonempty open set, hence includes a ball $B(x_2, \delta_2)$ with $0 < \delta_2 < \frac{1}{4}$. Inductively, we find $B(x_n, \delta_n)$, $0 < \delta_n < 2^{-n}$, such that $B(x_n, \delta_n) \subset B(x_{n-1}, \frac{1}{2}\delta_{n-1}) - A_n$ [in particular, $B(x_n, \delta_n) \cap A_n = \emptyset$]. Now if $n < m$,

$$d(x_n, x_m) \leq \sum_{i=n}^{m-1} d(x_i, x_{i+1}) < \sum_{i=n}^{m-1} 2^{-i} \rightarrow 0 \quad \text{as } n, m \rightarrow \infty,$$

so by completeness, x_n converges to some $x \in \Omega$. Since $x_k \in B(x_n, \frac{1}{2}\delta_n)$ for $k > n$, we have $x \in B(x_n, \delta_n)$ for all n , hence x cannot belong to any A_n , a contradiction. ■

In A9.2, if for each $n = 1, 2, \dots$, U_n is an open dense subset of Ω , then $\bigcap_{n=1}^{\infty} U_n \neq \emptyset$, for if the intersection is empty, then Ω is the union of the nowhere dense sets $\Omega - U_n$. We prove a stronger result:

A9.3 Theorem. If U_n is an open dense subset of the complete metric space Ω ($n = 1, 2, \dots$), then $\bigcap_{n=1}^{\infty} U_n$ is dense.

PROOF. Consider any ball B . Then \bar{B} is a closed subset of the complete metric space Ω , hence \bar{B} is complete. Now $U_n \cap B$ is dense in B since B is open; hence $U_n \cap B$ (and therefore $U_n \cap \bar{B}$) is dense in \bar{B} . By the remark after A9.2, $(\bigcap_{n=1}^{\infty} U_n) \cap \bar{B} \neq \emptyset$. Thus $\bigcap_{n=1}^{\infty} U_n$ meets every closed ball, hence every open ball, hence every nonempty open set. ■

Other important spaces satisfy the conclusion of A9.3:

A9.4 Theorem. If U_n is an open dense subset of the locally compact Hausdorff space Ω ($n = 1, 2, \dots$), then $\bigcap_{n=1}^{\infty} U_n$ is dense.

PROOF. Let U be a nonempty open subset of Ω ; we must show that $U \cap \bigcap_{n=1}^{\infty} U_n \neq \emptyset$. Since U_1 is dense, $U \cap U_1 \neq \emptyset$, and by local compactness (see A5.12), there is a nonempty, relatively compact, open set V_1 with $\bar{V}_1 \subset U \cap U_1$.

Now $V_1 \cap U_2 \neq \emptyset$ so we may choose a nonempty, relatively compact, open set V_2 with $\bar{V}_2 \subset V_1 \cap U_2$. Inductively we select nonempty, relatively compact, open sets V_n with $\bar{V}_n \subset V_{n-1} \cap U_n$.

Since the \bar{V}_n decrease with n , they have the finite intersection property, and since all \bar{V}_n are subsets of the compact set \bar{V}_1 , $\bigcap_{n=1}^{\infty} \bar{V}_n \neq \emptyset$. If x belongs to this intersection, then $x \in U \cap \bigcap_{n=1}^{\infty} U_n$. ■

A9.5 Definition. The topological space Ω is said to be a *Baire space* iff the intersection of countably many open dense subsets of Ω is dense in Ω . By A9.3 and A9.4, complete metric spaces and locally compact Hausdorff spaces are Baire spaces.

A9.6 Theorem. (a) Ω is a Baire space iff for each A of category 1 in Ω , $A^0 = \emptyset$; that is, $\Omega - A$ is dense.

(b) Let Ω be a Baire space. If $\Omega = \bigcup_{n=1}^{\infty} A_n$ where the A_n are closed, then $A_n^0 \neq \emptyset$ for some n . Thus Ω is of category 2 in itself.

PROOF. (a) Assume Ω is a Baire space, and let $A = \bigcup_{n=1}^{\infty} A_n$, where the A_n are nowhere dense. Then $\Omega - A = \bigcap_{n=1}^{\infty} (\Omega - A_n) \supset \bigcap_{n=1}^{\infty} (\Omega - \bar{A}_n)$, which is dense by hypothesis.

Conversely, if every set of category 1 in Ω has empty interior, let U_n be dense in Ω , $n = 1, 2, \dots$. Then $\Omega - U_n$ is (closed and) nowhere dense, so

$\bigcup_{n=1}^{\infty} (\Omega - U_n)$ is of category 1 in Ω . By hypothesis, the complement of this set, namely $\bigcap_{n=1}^{\infty} U_n$, is dense.

(b) If $A_n^0 = \emptyset$ for all n , then Ω is of category 1 in itself, contradicting (a). ■

There are spaces that cannot be expressed as a countable union of nowhere dense sets, but are not Baire spaces. (As an example, consider the free union of a Baire space and a non-Baire space.)

A9.7 Definition. The topological space Ω is said to be *topologically complete* iff there is a complete metric d on Ω that induces the given topology.

The following sequence of results is aimed at characterizing topologically complete spaces.

A9.8 Theorem. Let (Ω, d) be a complete metric space. If U is an open subset of Ω , U is topologically complete.

PROOF. If $x, y \in U$, let $d'(x, y) = d(x, y) + |f(x) - f(y)|$, where $f(x) = [d(x, \Omega - U)]^{-1}$. Then d' is a metric on U , and d -convergence is equivalent to d' -convergence by continuity of f . Thus d' induces the original topology. If $\{x_n\}$ is d' -Cauchy, it is d -Cauchy; hence it converges (relative to d) to some $x \in \Omega$. But $x \in U$; for if not, then $f(x_n) \rightarrow f(x) = \infty$, contradicting $\{x_n\}$ d' -Cauchy. Since d and d' are equivalent on U , x_n converges to x relative to d' . ■

A9.9 Theorem. Let (Ω, d) be a complete metric space, and U_n , $n = 1, 2, \dots$, open subsets of Ω . If $A = \bigcap_{n=1}^{\infty} U_n$, that is, if A is a G_{δ} , then A , with the relative topology, is topologically complete.

PROOF. Let d_n be a complete metric for U_n with d_n equivalent to d [d_n exists by A9.8; we may assume that $d_n \leq 1$, for if not replace d_n by $d_n/(1 + d_n)$.] Define $g: A \rightarrow B = \prod_{i=1}^{\infty} U_i$ by $g(x) = (x, x, \dots)$, and define a metric d_0 on B by $d_0(x, y) = \sum_{i=1}^{\infty} 2^{-i} d_i(x_i, y_i)$. Then d_0 induces the product topology on B , and a sequence $\{x^{(n)}\}$ in B is d_0 -Cauchy iff $\{x_i^{(n)}\}$ is d_i -Cauchy for each i . It follows that d_0 is complete. (Thus we have shown that a countable product of topologically complete spaces is topologically complete.)

Now g is a homeomorphism of A and $g(A) \subset B$, and in fact $g(A)$ is closed in B . For if $(x_n, x_n, \dots) \rightarrow y \in B$, then for each i we have $d_i(x_n, y_i) \rightarrow 0$ as $n \rightarrow \infty$, so that $x_n \rightarrow y_i$. But then y_i is the same for all i , say $y_i = y_0$. Since

$y_i \in U_i$ for each i we have $y_0 \in \bigcap_{i=1}^{\infty} U_i = A$, hence $y = (y_0, y_0, \dots) \in g(A)$, proving $g(A)$ closed. Since B is complete, so is $g(A)$, and since g is a homeomorphism, A is topologically complete. ■

It follows from A9.9 that the irrational numbers are topologically complete. In fact, a complete metric for the set of irrationals in $(0, 1)$ is given by $d(a_1 a_2 \dots, b_1 b_2 \dots) = 1/n$, where n is the first integer for which $a_n \neq b_n$ (in the decimal expansion).

A9.10 Theorem. Let A be a topologically complete subspace of the Hausdorff space Ω . If A is dense in Ω , then A is a G_δ .

PROOF. Let $U_n = \{x \in \Omega : \text{for some neighborhood } V \text{ of } x, \text{ the diameter of } V \cap A \text{ is less than } 1/n\}$. (The diameter is calculated using a particular complete metric d for A .) If $x \in U_n$, then $x \in V \subset U_n$, so U_n is open in Ω . We are going to show that $A = \bigcap_{n=1}^{\infty} U_n$.

Let $x \in A$, and let $V_0 = \{y \in A : d(x, y) < 1/4n\}$. Then V_0 is open in A ; hence $V_0 = V \cap A$, where V is open in Ω . Now $x \in V_0 \subset V$, and if $y, z \in V_0$, we have $d(y, z) \leq d(y, x) + d(x, z) < 1/2n$, so that V_0 has diameter less than or equal to $1/2n < 1/n$. It follows that $x \in \bigcap_{n=1}^{\infty} U_n$.

Conversely, assume $x \in \bigcap_{n=1}^{\infty} U_n$. Since A is dense, there is a net $x_i \in A$ with $x_i \rightarrow x$. For each n , let V be a neighborhood of x such that $V \cap A$ has diameter less than $1/n$. Eventually $x_i \in V$, and it follows that for some index i_n we have $d(x_i, x_j) < 1/n$ for $i, j \geq i_n$. (In other words, the net is Cauchy.) It may be assumed that $i_n \leq i_{n+1}$ for all n ; then the x_{i_n} form a Cauchy sequence which converges to a limit $y \in A$ by completeness. Since Ω is Hausdorff, we have $y = x$; hence $x \in A$. ■

We now prove the main theorem on topological completeness.

A9.11 Theorem. Let Ω be a metrizable topological space. Then Ω is topologically complete iff it is an absolute G_δ ; in other words, Ω is a G_δ in every metric space in which it is topologically embedded.

PROOF. If Ω is topologically complete and is embedded in the metric space Ω_1 , then Ω is dense in $\bar{\Omega}$; so by A9.10, Ω is a G_δ in $\bar{\Omega}$. Thus Ω can be expressed as $\bigcap_{n=1}^{\infty} W_n$, where $W_n = U_n \cap \bar{\Omega}$, U_n open in Ω_1 . But $\bar{\Omega}$ is a closed subset of the metric space Ω_1 , so $\bar{\Omega}$ is a G_δ in Ω_1 ($\bar{\Omega} = \bigcap_{n=1}^{\infty} \{x \in \Omega_1 : d(x, \bar{\Omega}) < 1/n\}$). It follows that Ω is a G_δ in Ω_1 .

Conversely, let Ω be an absolute G_δ . Embed Ω in its completion (we assume as known the standard process of completing a metric space by forming equivalence classes of Cauchy sequences). By A9.9, Ω is topologically complete. ■

We now establish topological completeness for a wide class of spaces.

A9.12 Theorem. A locally compact metric space is topologically complete.

PROOF. If Ω is such a space and $\bar{\Omega}$ is its completion, then Ω is open in $\bar{\Omega}$. For if $x \in \Omega$, there is, by local compactness, an open (in $\bar{\Omega}$) set V with $x \in V \cap \Omega$ and $\bar{V} \cap \Omega$ compact. In fact $V \subset \Omega$, proving Ω open. For if $y \in V$ and $y \notin \Omega$, then for each $U \in \mathcal{U}(y)$, $U \cap V \cap \Omega \neq \emptyset$ since Ω is dense; consequently, $U \cap \bar{V} \cap \Omega \neq \emptyset$. The sets $U \cap \bar{V} \cap \Omega$, $U \in \mathcal{U}(y)$, form a filterbase \mathcal{B} in $\bar{V} \cap \Omega$ converging to y , and since $\bar{\Omega}$ is Hausdorff, y is the only possible accumulation point of \mathcal{B} . But $y \notin \bar{V} \cap \Omega$, contradicting compactness of $\bar{V} \cap \Omega$ [see A5.2(d)]. By A9.8, Ω is topologically complete. ■

A10 Uniform Spaces

We now give an alternative way of describing gauge spaces.

A10.1 Definitions and Comments. Let V be a relation on the set Ω , that is, a subset of $\Omega \times \Omega$. Then V is called a *connector* iff the diagonal

$$D = \{(x, x) : x \in \Omega\}$$

is a subset of V . A nonempty collection \mathcal{H} of connectors is called a *uniformity* iff

- (a) for all $V_1, V_2 \in \mathcal{H}$, there is a $W \in \mathcal{H}$ with $W \subset V_1 \cap V_2$ (in other words \mathcal{H} is a filterbase), and
- (b) for each $V \in \mathcal{H}$, there is a $W \in \mathcal{H}$ with $WW^{-1} \subset V$.

[The relation WW^{-1} is the composition of the two relations W and W^{-1} ; that is, $(x, z) \in WW^{-1}$ iff for some $y \in \Omega$ we have $(x, y) \in W^{-1}$ and $(y, z) \in W$; $(x, y) \in W^{-1}$ means $(y, x) \in W$.]

If in addition, \mathcal{H} is a filter, that is, if $V \in \mathcal{H}$ and $V \subset W$, then $W \in \mathcal{H}$, then \mathcal{H} is called a *uniform structure*. In particular, if \mathcal{H} is a uniformity, the filter generated by \mathcal{H} is a uniform structure.

As an example, let \mathcal{D} be a family of pseudometrics on Ω , and let \mathcal{H} consist of all sets of the form $V = \{(x, y) : d^+(x, y) < \delta\}, \delta > 0, d^+ \in \mathcal{D}^+$; then \mathcal{H} is a uniformity on Ω . [To see that condition (b) is satisfied, let

$$W = \{(x, y) : d^+(x, y) < \delta/2\};$$

then $WW^{-1} \subset V$ by the triangle inequality.]

By analogy with the pseudometric case, if V belongs to the uniformity \mathcal{H} , we say that x and y are V -close iff $(x, y) \in V$.

Two uniformities \mathcal{H}_1 and \mathcal{H}_2 are said to be *equivalent* iff they generate the same uniform structure; in other words, given $V_2 \in \mathcal{H}_2$, there is a $V_1 \in \mathcal{H}_1$ with $V_1 \subset V_2$ and given $W_1 \in \mathcal{H}_1$, there is a $W_2 \in \mathcal{H}_2$ with $W_2 \subset W_1$. Two uniform structures that are equivalent must be equal; hence *there is exactly one uniform structure in each equivalence class of uniformities*.

If \mathcal{H} is a uniform structure and $V, W \in \mathcal{H}$, $WW^{-1} \subset V$, then if D is the diagonal, we have

$$W^{-1} = DW^{-1} \subset WW^{-1} \subset V,$$

hence $W \subset V^{-1}$, so that $V^{-1} \in \mathcal{H}$. But then

$$V \cap V^{-1} \in \mathcal{H} \text{ (and } V \cap V^{-1} \subset V).$$

Now $U = V \cap V^{-1}$ is *symmetric*, that is, $U = U^{-1}$. Thus if \mathcal{H} is a uniform structure, the symmetric sets in \mathcal{H} form a uniformity that generates \mathcal{H} , so that *every uniformity is equivalent to a uniformity containing only symmetric sets*.

We are going to show later that every uniform structure is generated by a uniformity corresponding to a family of pseudometrics. As a preliminary, we establish the following result:

A10.2 Metrization Lemma. Let $U_n, n = 0, 1, 2, \dots$, be a sequence of subsets of $\Omega \times \Omega$ such that each U_n is a connector, $U_0 = \Omega \times \Omega$, and $U_{n+1}^3 (= U_{n+1} U_{n+1} U_{n+1}) \subset U_n$ for all n . Then there is a function d from $\Omega \times \Omega$ to the nonnegative reals, such that d satisfies the triangle inequality and $U_n \subset \{(x, y) : d(x, y) < 2^{-n}\} \subset U_{n-1}$ for all $n = 1, 2, \dots$. If each U_n is symmetric, there is a pseudometric satisfying this condition.

PROOF. Define $f(x, y) = 0$ if $(x, y) \in U_n$ for all n ; $f(x, y) = \frac{1}{2}$ if $(x, y) \in U_0 - U_1$; $f(x, y) = \frac{1}{4}$ if $(x, y) \in U_1 - U_2$, and in general, $f(x, y) = 2^{-n}$ if $(x, y) \in U_{n-1} - U_n$. Since $U_{n+1} \subset U_{n+1}^3 \subset U_n$, f is well-defined on $\Omega \times \Omega$; furthermore,

$$(x, y) \in U_n \text{ iff } f(x, y) < 2^{-n}. \quad (1)$$

If $x, y \in \Omega$, let $d(x, y) = \inf \sum_{i=0}^n f(x_i, x_{i+1})$, where the inf is taken over all finite sequences x_0, x_1, \dots, x_{n+1} with $x_0 = x, x_{n+1} = y$.

Since a chain from x to y followed by a chain from y to z yields a chain from x to z , d satisfies the triangle inequality, and since $d(x, y) \leq f(x, y)$, $U_n \subset \{(x, y) : d(x, y) < 2^{-n}\}$ by (1). If the U_n are symmetric, then so is d ; hence d is a pseudometric.

Now we claim that

$$f(x_0, x_{n+1}) \leq 2 \sum_{i=0}^n f(x_i, x_{i+1}) \quad \text{for any } x_0, \dots, x_{n+1}. \quad (2)$$

This is clear for $n = 0$, so assume the inequality for all integers up to $n - 1$. Let $a = \sum_{i=0}^n f(x_i, x_{i+1})$. If $a = 0$, then $(x_i, x_{i+1}) \in \bigcap_{n=0}^{\infty} U_n$, and since $U_i^3 \subset U_{i-1}$ we have $(x_0, x_{n+1}) \in U_i$ for all i , and there is nothing further to prove. Thus assume $a > 0$.

Let k be the largest integer such that $\sum_{i=0}^{k-1} f(x_i, x_{i+1}) \leq a/2$ [if $f(x_0, x_1) > a/2$, take $k = 0$]. Then $\sum_{i=0}^k f(x_i, x_{i+1}) > a/2$, so $\sum_{i=k+1}^n f(x_i, x_{i+1}) \leq a/2$. By induction hypothesis, $f(x_0, x_k) \leq 2 \sum_{i=0}^{k-1} f(x_i, x_{i+1}) \leq 2a/2 = a$ and $f(x_{k+1}, x_{n+1}) \leq 2 \sum_{i=k+1}^n f(x_i, x_{i+1}) \leq 2a/2 = a$. Also,

$$f(x_k, x_{k+1}) \leq \sum_{i=0}^n f(x_i, x_{i+1}) = a.$$

Let m be the smallest nonnegative integer such that $2^{-m} \leq a$. If $m \leq 2$, then (2) holds automatically since $f \leq \frac{1}{2}$; thus assume $m > 2$. Then $2^{-m} \leq a < 2^{-(m-1)}$, so $f(x_0, x_k) \leq a < 2^{-(m-1)}$, and therefore by (1), (x_0, x_k) [and similarly, $(x_k, x_{k+1}), (x_{k+1}, x_{n+1})$] belongs to U_{m-1} . Thus $(x_0, x_{n+1}) \in U_{m-1}^3 \subset U_{m-2}$, which implies that $f(x_0, x_{n+1}) < 2^{-(m-2)}$; that is,

$$f(x_0, x_{n+1}) \leq 2^{-(m-1)} \leq 2a,$$

proving (2).

Finally, if $d(x, y) < 2^{-n}$, then for any chain from x to y , (2) yields (with $x_0 = x, x_{n+1} = y$) $f(x, y) \leq 2d(x, y) < 2^{-(n-1)}$, so $(x, y) \in U_{n-1}$. ■

A10.3 Metrization Theorem. If \mathcal{H} is a uniformity, the following conditions are equivalent:

(a) \mathcal{H} is *pseudometrizable*; that is, there is an equivalent uniformity generated by a single pseudometric. (See the example in A10.1 for the construction of the uniformity corresponding to a family of pseudometrics.)

(b) \mathcal{H} has a *countable base*; that is, there are countably many sets V_n in the uniform structure generated by \mathcal{H} such that if $U \in \mathcal{H}$, then $V_n \subset U$ for some n .

PROOF. If \mathcal{H} is pseudometrizable with pseudometric d , the sets

$$\{(x, y) : d(x, y) < r\},$$

r rational, form a countable base. Conversely, assume \mathcal{H} has a countable base V_0, V_1, V_2, \dots . We may assume that the V_n are symmetric (if not, consider $V_n \cap V_n^{-1}$), $V_n \supseteq V_{n+1}$ for all n (if not, consider $\bigcap_{i=0}^n V_i$), and $V_0 = \Omega \times \Omega$. Take $U_0 = V_0$, U_1 a symmetric set in the uniform structure \mathcal{F} generated by \mathcal{H} such that $U_1^3 \subset V_1$ ($\subset U_0$); let U_2 be a symmetric set in \mathcal{F} such that $U_2^3 \subset V_2 \cap U_1$. In general, let U_n be a symmetric set in \mathcal{F} such that $U_n^3 \subset V_n \cap U_{n-1}$. (To see that the U_n exist, let W_n be a symmetric set in \mathcal{F} with $W_n^2 \subset V_n \cap U_{n-1}$. If U_n is a symmetric set in \mathcal{F} such that $U_n^2 \subset W_n$, then $U_n^3 = U_n^2 D \subset U_n^4 \subset W_n^2 \subset V_n \cap U_{n-1}$. Also, since $U_n \subset U_n^3 \subset V_n$, the U_n form a base for \mathcal{H} .)

By A10.2, there is a pseudometric d such that $U_n \subset \{(x, y) : d(x, y) < 2^{-n}\} \subset U_{n-1}$, $n = 1, 2, \dots$. It follows that \mathcal{H} is equivalent to the uniformity of d . ■

We now construct a topology arising from a uniformity.

A10.4 Definitions and Comments. Let \mathcal{H} be a uniformity. The *topology induced* by \mathcal{H} is defined by specifying the system $\mathcal{V}(x)$ of overneighborhoods of each point x . If $V \in \mathcal{H}$, let $V(x) = \{y : (x, y) \in V\}$. We take $\mathcal{V}(x)$ as the class of all oversets of the sets $V(x)$, $V \in \mathcal{H}$.

It is immediate that the first three axioms for an overneighborhood system are satisfied (see A1.1). We verify the last axiom [A1.1(d)]. If $V \in \mathcal{H}$, let U be a symmetric set in the uniform structure generated by \mathcal{H} such that $U^2 \subset V$. If $W \in \mathcal{H}$ and $W \subset U$, we claim that $W(y) \subset V(x)$ for each $y \in W(x)$. For if $z \in W(y)$, then $(y, z) \in W$; but also $(x, y) \in W$, and hence $(x, z) \in W^2 \subset U^2 \subset V$, so $z \in V(x)$, as asserted. Thus $V(x) \in \mathcal{V}(y)$ for all $y \in W(x) \subset V(x)$, as required in A1.1(d).

It follows that there is a unique topology having (for each x) the sets $V(x)$, $V \in \mathcal{H}$, as a base for the overneighborhoods of x . Furthermore, any uniformity equivalent to \mathcal{H} will induce the same topology. A topological space whose topology is determined in this way is called a *uniform space*.

In particular, let \mathcal{D} be a family of pseudometrics on Ω , and let \mathcal{H} consist of all sets $V = \{(x, y) : d^+(x, y) < \delta\}$, $\delta > 0$, $d^+ \in \mathcal{D}^+$; then

$$V(x) = \{y : d^+(x, y) < \delta\}.$$

Thus the topology induced by \mathcal{H} coincides with the gauge space topology determined by \mathcal{D} (see A8.1). In other words, *every gauge space is a uniform space*. Conversely, *every uniform space is a gauge space*, as we now prove.

A10.5 Theorem. Let \mathcal{F} be a uniform structure, and let \mathcal{H} be the uniformity generated by the pseudometrics d such that the sets $V_{d\delta} = \{(x, y) : d(x, y) < \delta\}$ belong to \mathcal{F} for all $\delta > 0$. Then \mathcal{F} and \mathcal{H} are equivalent, so that the topology induced by \mathcal{F} is identical to the gauge space topology determined by the pseudometrics d .

PROOF. It follows from the definition of \mathcal{H} that $\mathcal{H} \subset \mathcal{F}$. Now if U is a symmetric set in \mathcal{F} , set $U_0 = \Omega \times \Omega$, $U_1 = U$, U_2 a symmetric set in \mathcal{F} such that $U_2^3 \subset U_1$, and in general, let U_n be a symmetric set in \mathcal{F} such that $U_n^3 \subset U_{n-1}$. By A10.2, there is a pseudometric d on Ω with

$$U_n \subset V_{d, 2^{-n}} \subset U_{n-1}$$

for all n . Since \mathcal{F} is a uniform structure, $V_{d, 2^{-n}} \in \mathcal{F}$ for all n , hence $V_{d\delta} \in \mathcal{F}$ for all $\delta > 0$. But $U_2 \subset V_{d, 1/4} \subset U_1 = U$, proving the equivalence of \mathcal{F} and \mathcal{H} . ■

Since uniform spaces and gauge spaces coincide, the concept of uniform continuity, defined previously (see A8.4) for mappings of one gauge space to another, may now be translated into the language of uniform spaces. If $f: \Omega \rightarrow \Omega_1$, where Ω and Ω_1 are uniform spaces with uniformities \mathcal{U} and \mathcal{V} , f is uniformly continuous iff given $V \in \mathcal{V}$, there is a $U \in \mathcal{U}$ such that $(x, y) \in U$ implies $(f(x), f(y)) \in V$.

We now consider separation properties in uniform spaces.

A10.6 Theorem. Let Ω be a uniform space, with uniformity \mathcal{H} . The following are equivalent:

- (a) $\bigcap\{V: V \in \mathcal{H}\}$ is the diagonal D .
- (b) Ω is T_2 .
- (c) Ω is T_1 .
- (d) Ω is T_0 .

PROOF. If (a) holds and $x \neq y$, then $(x, y) \notin V$ for some $V \in \mathcal{H}$. If W is a symmetric set in the uniform structure generated by \mathcal{H} , and $W^2 \subset V$, then $W(x)$ and $W(y)$ are disjoint overneighborhoods of x and y , proving (b). If (d) holds and $x \neq y$, there is a set $V \in \mathcal{H}$ such that $y \notin V(x)$ [or a set $W \in \mathcal{H}$ with $x \notin W(y)$]. But then $(x, y) \notin V$ for some $V \in \mathcal{H}$, proving (a). ■

Finally, we discuss topological groups and topological vector spaces as uniform spaces.

A *topological group* is a group on which there is defined a topology which makes the group operations continuous ($x_n \rightarrow x$, $y_n \rightarrow y$ implies $x_n y_n \rightarrow xy$;

$x_n \rightarrow x$ implies $x_n^{-1} \rightarrow x^{-1}$). Familiar examples are the integers, with ordinary addition and the discrete topology; the unit circle $\{z: |z| = 1\}$ in the complex plane, with multiplication of complex numbers and the Euclidean topology; all nonsingular $n \times n$ matrices of complex numbers, with matrix multiplication and the Euclidean topology (on R^{n^2}). If $\Omega = \prod_i \Omega_i$, where the Ω_i are topological groups, then Ω is a topological group with the product topology if multiplication is defined by $xy = (x_i y_i, i \in I)$.

A10.7 Theorem. Let Ω be a topological group, and let \mathcal{H} consist of all sets $V_N = \{(x, y) \in \Omega \times \Omega: yx^{-1} \in N\}$, where N ranges over all overneighborhoods of the identity element e in Ω . Then \mathcal{H} is a uniformity, and the topology induced by \mathcal{H} coincides with the original topology. In particular, if Ω is Hausdorff, it is completely regular.

PROOF. The diagonal is a subset of each V_N , and $V_N \cap V_M = V_{N \cap M}$, so only the last condition [A10.1(b)] for a uniformity need be checked. Let $f(x, y) = xy^{-1}$, a continuous map of $\Omega \times \Omega$ into Ω (by definition of a topological group). If N is an overneighborhood of e , there is an overneighborhood M of e such that $f(M \times M) \subset N$.

We claim that $V_M V_M^{-1} \subset V_N$. For if $(x, y) \in V_M^{-1}$ and $(y, z) \in V_M$, then $(y, x) \in V_M$; hence $xy^{-1} \in M$ and $zy^{-1} \in M$. But $zx^{-1} = (zy^{-1})(yx^{-1}) = (zy^{-1})(xy^{-1})^{-1} \in f(M \times M) \subset N$. Consequently, $(x, z) \in V_N$, so \mathcal{H} is a uniformity.

Let U be an overneighborhood of the point x in the original topology. If $N = Ux^{-1} = \{yx^{-1}: y \in U\}$, then $V_N(x) = \{y: (x, y) \in V_N\} = \{y: yx^{-1} \in N\} = Nx = U$; therefore U is an overneighborhood of x in the topology induced by \mathcal{H} . Conversely, let U be an overneighborhood of x in the uniform space topology; then $V_N(x) \subset U$ for some overneighborhood N of e . But $V_N(x) = Nx$, and the map $y \rightarrow yx$, carrying N onto Nx , is a homeomorphism of Ω with itself. Thus Nx , hence U , is an overneighborhood of x in the original topology. ■

A10.8 Theorem. Let Ω be a topological group, with uniformity \mathcal{H} as defined in A10.7. The following are equivalent:

- (a) \mathcal{H} is pseudometrizable (see A10.3).
- (b) Ω is pseudometrizable; that is, there is a pseudometric that induces the given topology.
- (c) Ω has a countable base of neighborhoods at the identity e (hence at every x , since the map $N \rightarrow Nx$ sets up a one-to-one correspondence between neighborhoods of e and neighborhoods of x).

PROOF. We obtain (a) implies (b) by the definition of the topology induced by \mathcal{H} (see A10.4). Since every pseudometric space is first countable, it follows that (b) implies (c). Finally, if (c) holds and the sets N_1, N_2, \dots form a countable base at e , then by definition of \mathcal{H} , the sets V_{N_1}, V_{N_2}, \dots form a countable base for \mathcal{H} ; hence, by A10.3, \mathcal{H} is pseudometrizable. ■

A *topological vector space* is a vector space with a topology that makes addition and scalar multiplication continuous (see 3.5). In particular, a topological vector space is an abelian topological group under addition. The uniformity \mathcal{H} consists of sets of the form $V_N = \{(x, y) : y - x \in N\}$, where N is an overneighborhood of 0. Theorem A10.8 shows that a topological vector space is pseudometrizable iff there is a countable base at 0.

The following fact is needed in the proof of the open mapping theorem (see 3.5.9):

A10.9 Theorem. If L is a pseudometrizable topological vector space, there is an invariant pseudometric $[d(x, y) = d(x + z, y + z)$ for all z] that induces the topology of L .

PROOF. The pseudometric may be constructed by the method given in A10.2 and A10.3, and furthermore, the symmetric sets needed in A10.3 may be taken as sets in the uniformity \mathcal{H} itself rather than in the uniform structure generated by \mathcal{H} . For if N is an overneighborhood of 0, so is $-N = \{-x : x \in N\}$ since the map $x \rightarrow -x$ is a homeomorphism. Thus $M = N \cap (-N)$ is an overneighborhood of 0. But then V_M is symmetric and $V_M \subset V_N$.

Now by definition of V_M , we have $(x, y) \in V_M$ iff $(x + z, y + z) \in V_M$ for all z , and it follows from the proof of A10.2 that the pseudometric d is invariant. ■

Bibliography

- Apostol, T. M., "Mathematical Analysis." Addison-Wesley, Reading, Massachusetts, 1957.
- Ash, R. B., "Basic Probability Theory." Wiley, New York, 1970.
- Ash, R. B., "Complex Variables." Academic Press, New York, 1971.
- Bachman, G., and Narici, L., "Functional Analysis." Academic Press, New York, 1966.
- Billingsley, P., "Convergence of Probability Measures." Wiley, New York, 1968.
- Breiman, L., "Probability." Addison-Wesley, Reading, Massachusetts, 1968.
- Chung, K. L., "A Course in Probability Theory." Harcourt, New York, 1968.
- Doob, J. L., "Stochastic Processes." Wiley, New York, 1953.
- Dubins, L., and Savage, L., "How to Gamble If You Must." McGraw-Hill, New York, 1965.
- Dugundji, J., "Topology." Allyn and Bacon, Boston, 1966.
- Dunford, N., and Schwartz, J. T., "Linear Operators." Wiley (Interscience), New York, Part 1, 1958; Part 2, 1963; Part 3, 1970.
- Feller, W., "Introduction to Probability Theory." Wiley, New York, 1950.
- Gikhman, I. I., and Skorokhod, A. V., "Introduction to the Theory of Random Processes." Saunders, Philadelphia, 1969.
- Gnedenko, B. V., and Kolmogorov, A. N., "Limit Distributions for Sums of Independent Random Variables." Addison-Wesley, Reading, Massachusetts, 1954.
- Halmos, P. R., "Measure Theory." Van Nostrand-Reinhold, Princeton, New Jersey, 1950.
- Halmos, P. R., "Introduction to Hilbert Space." Chelsea, New York, 1951.
- Halmos, P. R., "Naive Set Theory." Van Nostrand-Reinhold, Princeton, New Jersey, 1960.
- Hausdorff, F., "Set Theory." Chelsea, New York, 1962.
- Karlin, S., "A First Course in Stochastic Processes." Academic Press, New York, 1966.
- Kelley, J. L., and Namioka, I., "Linear Topological Spaces." Van Nostrand-Reinhold, Princeton, New Jersey, 1963.
- Kemeny, J. G., Snell, J. L., and Knapp, A. W., "Denumerable Markov Chains." Van Nostrand-Rheinhold, Princeton, New Jersey, 1966.
- Liusternik, L., and Sobolev, V., "Elements of Functional Analysis." Ungar, New York, 1961.
- Loève, M., "Probability Theory." Van Nostrand-Reinhold, Princeton, New Jersey, 1955: 2nd ed., 1960; 3rd ed., 1963.
- Meyer, P. A., "Probability and Potentials." Blaisdell, Waltham, Massachusetts, 1966.
- Neveu, J., "Mathematical Foundations of the Calculus of Probability." Holden-Day, San Francisco, 1965.
- Parthasarathy, K., "Probability Measures on Metric Spaces." Academic Press, New York, 1967.
- Parzen, E., "Modern Probability Theory." Wiley, New York, 1960.
- Royden, H. L., "Real Analysis." Macmillan, New York, 1963; 2nd ed., 1968.
- Rudin, W., "Real and Complex Analysis." McGraw-Hill, New York, 1966.
- Schaefer, H., "Topological Vector Spaces." Macmillan, New York, 1966.
- Simmons, G., "Introduction to Topology and Modern Analysis." McGraw-Hill, New York, 1963.
- Taylor, A. E., "Introduction to Functional Analysis." Wiley, New York, 1958.
- Titchmarsh, E. C., "The Theory of Functions." Oxford Univ. Press, London and New York, 1939.
- Tucker, H. G., "A Graduate Course in Probability." Academic Press, New York, 1967.
- Yosida, K., "Functional Analysis." Springer-Verlag, Berlin and New York, 1968.

Solutions to Problems

Chapter 1

Section 1.1

2. We have $\limsup_n A_n = (-1, 1]$, $\liminf_n A_n = \{0\}$.
3. Using $\limsup_n A_n = \{\omega : \omega \in A_n \text{ for infinitely many } n\}$, $\liminf_n A_n = \{\omega : \omega \in A_n \text{ for all but finitely many } n\}$, we obtain

$$\liminf_n A_n = \{(x, y) : x^2 + y^2 < 1\},$$

$$\limsup_n A_n = \{(x, y) : x^2 + y^2 \leq 1\} - \{(0, 1), (0, -1)\}.$$

4. If $x = \limsup_{n \rightarrow \infty} x_n$, then $\limsup_n A_n$ is either $(-\infty, x)$ or $(-\infty, x]$. For if $y \in A_n$ for infinitely many n , then $x_n > y$ for infinitely many n ; hence $x \geq y$. Thus $\limsup_n A_n \subset (-\infty, x]$. But if $y < x$, then $x_n > y$ for infinitely many n , so $y \notin \limsup_n A_n$. Thus $(-\infty, x) \subset \limsup_n A_n$, and the result follows. The same result is valid for \liminf ; the above analysis applies, with “eventually” replacing “for infinitely many n .”

Section 1.2

4. (a) If $-\infty \leq a < b < c < \infty$, then $\mu(a, c] = \mu(a, b] + \mu(b, c]$, and $\mu(a, \infty) = \mu(a, b] + \mu(b, \infty)$; finite additivity follows quickly. If $A_n = (-\infty, n]$, then $A_n \uparrow R$, but $\mu(A_n) = n \rightarrow \mu(R) = 0$. Thus μ is not continuous from below, hence not countably additive.
- (b) Finiteness of μ follows from the definition; since $\mu(-\infty, n] \rightarrow \infty$, μ is unbounded.
5. We have $\mu(\bigcup_{i=1}^{\infty} A_i) \geq \mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$ for all n ; let $n \rightarrow \infty$ to obtain the desired result.
8. The minimal σ -field \mathcal{F} (which is also the minimal field) consists of the collection \mathcal{G} of all (finite) unions of sets of the form $B_1 \cap B_2 \cap \dots \cap B_n$, where B_i is either A_i or A_i^c . For any σ -field containing A_1, \dots, A_n must contain all sets in \mathcal{G} ; hence $\mathcal{G} \subset \mathcal{F}$. But \mathcal{G} is a σ -field; hence $\mathcal{F} \subset \mathcal{G}$. Since there are 2^n disjoint sets of the form $B_1 \cap \dots \cap B_n$, and each such set may or may not be included in a typical set in \mathcal{F} , \mathcal{F} has at most 2^{2^n} members. The upper bound is attained if all sets $B_1 \cap \dots \cap B_n$ are nonempty.
9. (a) As in Problem 8, any field over \mathcal{C} must contain all sets in \mathcal{G} ; hence $\mathcal{G} \subset \mathcal{F}$. But \mathcal{G} is a field; hence $\mathcal{F} \subset \mathcal{G}$. For if $A_i = \bigcap_{j=1}^r B_{ij}$, then $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n \bigcup_{j=1}^r B_{ij}^c$, which belongs to \mathcal{G} because of the distributive law $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- (b) Note that the complement of a finite intersection $\bigcap_{j=1}^r B_{ij}$ belongs to \mathcal{D} ; for example, if $B_1, B_2 \in \mathcal{C}$, then

$$\begin{aligned} & (B_1 \cap B_2^c)^c \\ &= B_1^c \cup B_2 \\ &= (B_1^c \cap B_2) \cup (B_1^c \cap B_2^c) \cup (B_2 \cap B_1) \cup (B_2 \cap B_1^c) \in \mathcal{D}. \end{aligned}$$

Now \mathcal{D} is closed under finite intersection by the distributive law, and it follows from this and the above remark that \mathcal{D} is closed under complementation and is therefore a field. Just as in the proof that $\mathcal{F} = \mathcal{G}$, we find that $\mathcal{F} = \mathcal{D}$.

- (c) This is immediate from (a) and (b).
11. (a) Let $A_n \in \mathcal{S}$, $n = 1, 2, \dots$. Then A_n belongs to some \mathcal{C}_{α_n} , and we may assume $\alpha_1 \leq \alpha_2 \leq \dots$, so $\mathcal{C}_{\alpha_1} \subset \mathcal{C}_{\alpha_2} \subset \dots$. Let $\alpha = \sup_n \alpha_n < \beta_1$. Then all $\mathcal{C}_{\alpha_n} \subset \mathcal{C}_\alpha$, hence all $A_n \in \mathcal{C}_\alpha$. Thus $\bigcup_n A_n \in \mathcal{C}_{\alpha+1} \subset \mathcal{S}$, so $\bigcup_n A_n \in \mathcal{S}$. If $A \in \mathcal{S}$, then A belongs to some \mathcal{C}_α ; hence $A^c \in \mathcal{C}_{\alpha+1} \subset \mathcal{S}$.
- (b) We have $\text{card } \mathcal{C}_\alpha \leq c$ for all α . For this is true for $\alpha = 0$, by hypothesis. If it is true for all $\beta < \alpha$, then $\bigcup_{\beta < \alpha} \mathcal{C}_\beta$ has cardinality at

most $(\text{card } \alpha)c = c$. Now if \mathcal{D} has cardinality c , then \mathcal{D}' has cardinality at most $c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0} = c$. Thus $\text{card } \mathcal{C}_\alpha \leq c$. It follows that $\bigcup_{\alpha < \beta_1} \mathcal{C}_\alpha$ has cardinality at most c .

Section 1.3

3. (a) Since $\lambda(\emptyset) = 0$ we have $\Omega \in \mathcal{M}$, and \mathcal{M} is clearly closed under complementation. If $E, F \in \mathcal{M}$ and $A \subset \Omega$, then

$$\begin{aligned}\lambda[A \cap (E \cup F)] &= \lambda[A \cap (E \cup F) \cap E] \\ &\quad + \lambda[A \cap (E \cup F) \cap E^c] \quad \text{since } E \in \mathcal{M} \\ &= \lambda(A \cap E) + \lambda(A \cap F \cap E^c).\end{aligned}$$

Thus

$$\begin{aligned}\lambda[A \cap (E \cup F)] + \lambda[A \cap (E \cup F)^c] &= \lambda(A \cap E) + \lambda(A \cap E^c \cap F) + \lambda(A \cap E^c \cap F^c) \\ &= \lambda(A \cap E) + \lambda(A \cap E^c) \quad \text{since } F \in \mathcal{M} \\ &= \lambda(A) \quad \text{since } E \in \mathcal{M}.\end{aligned}$$

This proves that \mathcal{M} is a field. Also, if E and F are disjoint we have

$$\begin{aligned}\lambda[A \cap (E \cup F)] &= \lambda[A \cap (E \cup F) \cap E] + \lambda[A \cap (E \cup F) \cap E^c] \\ &= \lambda(A \cap E) + \lambda(A \cap F \cap E^c) \\ &= \lambda(A \cap E) + \lambda(A \cap F) \quad \text{since } E \cap F = \emptyset.\end{aligned}$$

Now if the E_n are disjoint sets in \mathcal{M} and $F_n = \bigcup_{i=1}^n E_i \uparrow E$, then

$$\begin{aligned}\lambda(A) &= \lambda(A \cap F_n) + \lambda(A \cap F_n^c) \\ &\quad \text{since } F_n \text{ belongs to the field } \mathcal{M} \\ &\geq \lambda(A \cap F_n) + \lambda(A \cap E^c) \\ &\quad \text{since } E^c \subset F_n^c \text{ and } \lambda \text{ is monotone} \\ &= \sum_{i=1}^n \lambda(A \cap E_i) + \lambda(A \cap E^c) \\ &\quad \text{by what we have proved above.}\end{aligned}$$

Since n is arbitrary,

$$\begin{aligned}\lambda(A) &\geq \sum_{n=1}^{\infty} \lambda(A \cap E_n) + \lambda(A \cap E^c) \\ &\geq \lambda(A \cap E) + \lambda(A \cap E^c) \quad \text{by countable subadditivity of } \lambda.\end{aligned}$$

Thus $E \in \mathcal{M}$, proving that \mathcal{M} is a σ -field.

Now $\lambda(A \cap E) + \lambda(A \cap E^c) \geq \lambda(A)$ by subadditivity, hence

$$\lambda(A) = \sum_{n=1}^{\infty} \lambda(A \cap E_n) + \lambda(A \cap E^c).$$

Replace A by $A \cap E$ to obtain $\lambda(A \cap E) = \sum_{n=1}^{\infty} \lambda(A \cap E_n)$, as desired.

- (b) All properties are immediate except for countable subadditivity. If $A = \bigcup_{n=1}^{\infty} A_n$, we must show that $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$, and we may assume that $\mu^*(A_n) < \infty$ for all n . Given $\varepsilon > 0$, we may choose sets $E_{nk} \in \mathcal{F}_0$ with $A_n \subset \bigcup_k E_{nk}$ and $\sum_k \mu(E_{nk}) \leq \mu^*(A_n) + \varepsilon 2^{-n}$. Then $A \subset \bigcup_{n,k} E_{nk}$ and $\sum_{n,k} \mu(E_{nk}) \leq \sum_n \mu^*(A_n) + \varepsilon$. Thus $\mu^*(A) \leq \sum_n \mu^*(A_n) + \varepsilon$, ε arbitrary.

Now if $A \in \mathcal{F}_0$, then $\mu^*(A) \leq \mu(A)$ by definition of μ^* , and if $A \subset \bigcup_n E_n$, $E_n \in \mathcal{F}_0$, then $\mu(A) \leq \sum_n \mu(E_n)$ by 1.2.5 and 1.3.1. Take the infimum over all such coverings of A to obtain $\mu(A) \leq \mu^*(A)$; hence $\mu^* = \mu$ on \mathcal{F}_0 .

- (c) If $F \in \mathcal{F}_0$, $A \subset \Omega$, we must show that $\mu^*(A) \geq \mu^*(A \cap F) + \mu^*(A \cap F^c)$; we may assume $\mu^*(A) < \infty$. Given $\varepsilon > 0$, there are sets $E_n \in \mathcal{F}_0$ with $A \subset \bigcup_n E_n$ and $\sum_{n=1}^{\infty} \mu(E_n) \leq \mu^*(A) + \varepsilon$. Now

$$\begin{aligned} \mu^*(A \cap F) &\leq \mu^*\left(\bigcup_n (E_n \cap F)\right) \quad \text{by monotonicity} \\ &\leq \sum_n \mu(E_n \cap F) \\ &\quad \text{since } \mu^* \text{ is countably subadditive and } \mu^* = \mu \text{ on } \mathcal{F}_0. \end{aligned}$$

Similarly,

$$\mu^*(A \cap F^c) \leq \sum_n \mu(E_n \cap F^c).$$

Thus

$$\mu^*(A \cap F) + \mu^*(A \cap F^c) \leq \sum_n \mu(E_n) \leq \mu^*(A) + \varepsilon,$$

and the result follows.

- (d) If $A = B \cup N$, where $B \in \sigma(\mathcal{F}_0)$, $N \subset M \in \sigma(\mathcal{F}_0)$, $\mu^*(M) = 0$, then $B \in \mathcal{M}$ [note $\mathcal{F}_0 \subset \mathcal{M}$ and \mathcal{M} is a σ -field, so $\sigma(\mathcal{F}_0) \subset \mathcal{M}$]. Also, any set C with $\mu^*(C) = 0$ belongs to \mathcal{M} by definition of μ^* -measurability; hence $A \in \mathcal{M}$. Therefore the completion of $\sigma(\mathcal{F}_0)$ is included in \mathcal{M} .

Now assume μ σ -finite on \mathcal{F}_0 , and let $A \in \mathcal{M}$. If Ω is the disjoint union of sets $A_n \in \mathcal{F}_0$ with $\mu(A_n) < \infty$, then by definition of μ^* , there is a set $B_n \in \sigma(\mathcal{F}_0)$ such that $A \cap A_n \subset B_n$ and $\mu^*(B_n - (A \cap A_n)) = 0$.

[Note that if $A \notin \mathcal{M}$ we obtain only $\mu^*(B_n) = \mu^*(A \cap A_n)$; however if $A \in \mathcal{M}$ (so that $A \cap A_n$ also belongs to \mathcal{M}), we have

$$\mu^*(B_n - (A \cap A_n)) = \mu^*(B_n) - \mu^*(A \cap A_n) = 0.]$$

If $B = \bigcup_n B_n$, then $B \in \sigma(\mathcal{F}_0)$, $A \subset B$, and $\mu^*(B - A) = 0$. This argument applied to A^c yields a set $C \in \sigma(\mathcal{F}_0)$ with $C \subset A$ and $\mu^*(A - C) = 0$. Therefore $A = C \cup (A - C)$ with $C \in \sigma(\mathcal{F}_0)$, $A - C \subset B - C \in \sigma(\mathcal{F}_0)$, and $\mu^*(B - C) = \mu^*(B - A) + \mu^*(A - C) = 0$. Thus A belongs to the completion of $\sigma(\mathcal{F}_0)$ relative to μ^* .

Section 1.4

1. Using the formulas of 1.4.5, the following results are obtained:
 - (a) 3, (b) 8.5, (c) 5,
 - (d) 7.25, (e) $\mu(\frac{1}{2}, \infty) + \mu(-\infty, -\frac{1}{2}) = 7.25$.
4. Let $C_k = \{x \in R^n : -k < x_i \leq k, i = 1, \dots, n\}$. Then μ is finite on C_k ; hence the Borel subsets B of C_k such that $\mu(a + B) = \mu(B)$ form a monotone class including the field of finite disjoint unions of right-semiclosed intervals in C_k ; hence all Borel subsets of C_k belong to the class (see 1.2.2). If $B \in \mathcal{B}(R^n)$, then $B \cap C_k \uparrow B$; hence $a + (B \cap C_k) \uparrow a + B$, and it follows that $\mu(a + B) = \mu(B)$.

Now if $B \in \bar{\mathcal{B}}(R^n)$, then $B = A \cup C$, $A \in \mathcal{B}(R^n)$, $C \subset D \in \mathcal{B}(R^n)$, with $\mu(D) = 0$. Thus $a + B = (a + A) \cup (a + C)$, and, by Problem 3, $a + A \in \mathcal{B}(R^n)$, $a + C \subset a + D \in \mathcal{B}(R^n)$. By what we have proved above, $\mu(A + D) = \mu(D) = 0$; hence $a + B \in \mathcal{B}(R^n)$ and $\mu(a + B) = \mu(B)$.

5. Let A be the unit cube $\{x \in R^n : 0 < x_i \leq 1, i = 1, \dots, n\}$, and let $c = \mu(A)$. For any positive integer, r , we may divide each edge of A into r equal parts, so that A is decomposed into r^n subcubes A_1, \dots, A_{r^n} , each with volume r^{-n} . By translation-invariance, $\mu(A_i)$ is the same for all i , so if λ is Lebesgue measure, we have

$$\mu(A_i) = r^{-n} \mu(A) = r^{-n} c = r^{-n} c \lambda(A) = c \lambda(A_i), \quad i = 1, \dots, r^n.$$

Now any subinterval I of the unit cube can be expressed as the limit of an increasing sequence of sets B_k , where each B_k is a finite disjoint union of subcubes of the above type. Thus $\mu = c\lambda$ on subintervals of the unit cube, and hence on all Borel subsets of the unit cube by the Carathéodory extension theorem. Since R^n is a countable disjoint union of cubes, it follows that $\mu = c\lambda$ on $\mathcal{B}(R^n)$.

6. (a) If $r + x_1 = s + x_2$, $x_1, x_2 \in A$, then x_1 is equivalent to x_2 , so that $x_1 = x_2$ since A was constructed by taking one member from each distinct B_x . Thus $r = s$, a contradiction.

If $x \in R$, then $x \in B_x$; if y is the member of B_x that belongs to A , then $x - y$ is a rational number r , hence $x \in r + A$.

- (b) If $0 \leq r \leq 1$, then $r + A \subset [0, 2]$; thus

$$\begin{aligned} & \sum \{\mu(r + A) : 0 \leq r \leq 1, r \text{ rational}\} \\ &= \mu(\bigcup \{r + A : 0 \leq r \leq 1, r \text{ rational}\}) \quad \text{by (a)} \\ &\leq \mu[0, 2] < \infty. \end{aligned}$$

But $\mu(r + A) = \mu(A)$ by Problem 4; hence $\mu(r + A)$ must be 0 for all r . Since R is a countable union of sets $r + A$ by (a), $\mu(R) = 0$, a contradiction.

8. Let $F(x, y) = 1$ if $x + y \geq 0$; $F(x, y) = 0$ if $x + y < 0$. If $a_1 = 0, b_1 = 1, a_2 = -1, b_2 = 0$, then

$$\begin{aligned} \Delta_{b_1 a_1} \Delta_{b_2 a_2} F(x, y) &= F(b_1, b_2) - F(a_1, b_2) \\ &\quad - F(b_1, a_2) + F(a_1, a_2) \\ &= 1 - 1 - 1 + 0 = -1; \end{aligned}$$

hence F is not a distribution function. Other examples: $F(x, y) = \max(x, y)$, $F(x, y) = [x + y]$, the largest integer less than or equal to $x + y$.

12. (a) First assume that μ is finite. Let \mathcal{C} be the class of subsets of R^n having the desired property; we show that \mathcal{C} is a monotone class. For let $B_n \in \mathcal{C}$, $B_n \uparrow B$. Let K_n be a compact subset of B_n with $\mu(B_n) \leq \mu(K_n) + \varepsilon$, $\varepsilon > 0$ preassigned. By replacing K_n by $\bigcup_{i=1}^n K_i$, we may assume the K_n form an increasing sequence. Then $\mu(B) = \lim_{n \rightarrow \infty} \mu(B_n) \leq \lim_{n \rightarrow \infty} \mu(K_n) + \varepsilon$, so that

$$\mu(B) = \sup \{\mu(K) : K \text{ a compact subset of } B\},$$

and $B \in \mathcal{C}$. If $B_n \in \mathcal{C}$, $B_n \downarrow B$, let K_n be a compact subset of B_n such that $\mu(B_n) \leq \mu(K_n) + \varepsilon 2^{-n}$, and set $K = \bigcap_{n=1}^{\infty} K_n$. Then

$$\mu(B) - \mu(K) = \mu(B - K) \leq \mu \left(\bigcup_{n=1}^{\infty} (B_n - K_n) \right) \leq \sum_{n=1}^{\infty} \mu(B_n - K_n) \leq \varepsilon;$$

thus $B \in \mathcal{C}$. Therefore \mathcal{C} is a monotone class containing all finite disjoint unions of right-semiclosed intervals (a right-semiclosed interval is the limit of an increasing sequence of compact intervals). Hence \mathcal{C} contains all Borel sets.

If μ is σ -finite, each $B \in \mathcal{B}(R^n)$ is the limit of an increasing sequence of sets B_i of finite measure. Each B_i can be approximated from within by compact sets [apply the previous argument to the measure given by $\mu_i(A) = \mu(A \cap B_i)$, $A \in \mathcal{B}(R^n)$], and the above argument that \mathcal{C} is closed under limits of increasing sequences shows that $B \in \mathcal{C}$.

(b) We have $\mu(B) \leq \inf\{\mu(V): V \supset B, V \text{ open}\}$

$$\leq \inf\{\mu(W): W \supset B, W = K^c, K \text{ compact}\}.$$

If μ is finite, this equals $\mu(B)$ by (a) applied to B^c , and the result follows.

Now assume μ is an arbitrary Lebesgue-Stieltjes measure, and write $R^n = \bigcup_{k=1}^{\infty} B_k$, where the B_k are disjoint bounded sets; then $B_k \subset C_k$ for some bounded open set C_k . The measure $\mu_k(A) = \mu(A \cap C_k)$, $A \in \mathcal{B}(R^n)$, is finite; hence if B is a Borel subset of B_k and $\varepsilon > 0$, there is an open set $W_k \supset B$ such that $\mu_k(W_k) \leq \mu_k(B) + \varepsilon 2^{-k}$. Now $W_k \cap C_k$ is an open set V_k and $B \cap C_k = B$ since $B \subset B_k \subset C_k$; hence $\mu(V_k) \leq \mu(B) + \varepsilon 2^{-k}$. For any $A \in \mathcal{B}(R^n)$, let V_k be an open set with $V_k \supset A \cap B_k$ and $\mu(V_k) \leq \mu(A \cap B_k) + \varepsilon 2^{-k}$. Then $V = \bigcup_{k=1}^{\infty} V_k$ is open, $V \supset A$, and $\mu(V) \leq \sum_{k=1}^{\infty} \mu(V_k) \leq \mu(A) + \varepsilon$.

(c) Construct a measure μ on $\mathcal{B}(R)$ as follows. Let μ be concentrated on $S = \{1/n: n = 1, 2, \dots\}$ and take $\mu\{1/n\} = 1/n$ for all n . Since $R = \bigcup_{n=1}^{\infty} \{1/n\} \cup S^c$ and $\mu(S^c) = 0$, μ is σ -finite. Since

$$\mu[0, 1] = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

μ is not a Lebesgue-Stieltjes measure. Now $\mu\{0\} = 0$, but if V is an open set containing 0, we have

$$\begin{aligned} \mu(V) &\geq \mu(-\varepsilon, \varepsilon) \quad \text{for some } \varepsilon \\ &\geq \sum_{k=r}^{\infty} \frac{1}{k} \quad \text{for some } r \\ &= \infty. \end{aligned}$$

Thus (b) fails. (Another example: Let $\mu(A)$ be the number of rational points in A .)

Section 1.5

2. If $B \in \mathcal{B}(R)$,

$$\begin{aligned} \{\omega: h(\omega) \in B\} &= \{\omega \in A: h(\omega) \in B\} \cup \{\omega \in A^c: h(\omega) \in B\} \\ &= [A \cap f^{-1}(B)] \cup [A^c \cap g^{-1}(B)] \end{aligned}$$

which belongs to \mathcal{F} since f and g are Borel measurable.

5. (a) $\{x: f \text{ is discontinuous at } x\} = \bigcup_{n=1}^{\infty} D_n$, where $D_n = \{x \in R^k: \text{for all } \delta > 0, \text{there exist } x_1, x_2 \in R^k \text{ such that } |x_1 - x| < \delta \text{ and } |x_2 - x| < \delta, \text{ but } |f(x_1) - f(x_2)| \geq 1/n\}$. We show that the D_n are closed. Let $\{x_n\}$

be a sequence of points in D_n with $x_\alpha \rightarrow x$. If $\delta > 0$ and $N = \{y: |y - x| < \delta\}$, then $x_\alpha \in N$ for large α , and since $x_\alpha \in D_n$, there are points $x_{\alpha 1}$ and $x_{\alpha 2} \in N$ such that $|f(x_{\alpha 1}) - f(x_{\alpha 2})| \geq 1/n$. Thus $|x_{\alpha 1} - x| < \delta$, $|x_{\alpha 2} - x| < \delta$, but $|f(x_{\alpha 1}) - f(x_{\alpha 2})| \geq 1/n$, so that $x \in D_n$.

The result is true for a function from an arbitrary topological space S to a metric space (T, d) . Take $D_n = \{x \in S: \text{for every neighborhood } N \text{ of } x, \text{ there exist } x_1, x_2 \in N \text{ such that } d(f(x_1), f(x_2)) \geq 1/n\}$. (the above proof goes through with “sequence” replaced by “net.”)

In fact T may be a uniform space (see the appendix on general topology, Section A10). If \mathcal{D} is the associated family of pseudometrics, we take $D_n = \{x \in S: \text{for every neighborhood of } N \text{ of } x, \text{ there exist } x_1, x_2 \in N \text{ and } d \in \mathcal{D} \text{ such that } d(f(x_1), f(x_2)) \geq 1/n\}$ and proceed as above.

The result is false if no assumptions are made about the topology of the range space. For example, let $\Omega = \{1, 2, 3\}$, with open sets \emptyset , Ω , and $\{1\}$. Define $f: \Omega \rightarrow \Omega$ by $f(1) = f(3) = 1, f(2) = 2$. Then the set of discontinuities is $\{3\}$, which is not an F_σ .

- (b) This follows from part (a) because the irrationals I cannot be expressed as a countable union of closed sets C_n . For if this were possible, then each C_n would have empty interior since every nonempty open set contains rational points. But then I is of category 1 in R , and since $Q = R - I$ is of category 1 in R , it follows that R is of category 1 in itself, contradicting the Baire category theorem.
- 6. By Problem 11, Section 1.4, there are c Borel subsets of R^n ; hence there are only c simple functions on R^n . Since a Borel measurable function is the limit of a sequence of simple functions, there are $c^{\aleph_0} = c$ Borel measurable functions from R^n to R . By 1.5.8, there are only c Borel measurable functions from R^n to R^k .
- 7. (a) Since the P_n are measures, $\sum_k P_n(A_k) = P_n(\Omega) = 1$, and it follows quickly that the a_{nk} satisfy the hypotheses of Steinhaus' lemma. If $\{x_n\}$ is the sequence given by the lemma, let $S = \{k: x_k = 1\}$ and let B be the union of the sets $A_k, k \in S$. Then

$$t_n = \frac{1}{1-\alpha} \sum_{k \in S} [P_n(A_k) - P(A_k)] = \frac{1}{1-\alpha} \left[P_n(B) - \sum_{k \in S} P(A_k) \right],$$

and it follows that t_n converges, a contradiction. Thus P is a probability measure. If $B_k \in \mathcal{F}, B_k \downarrow \emptyset$, then given $\varepsilon > 0$, we have $P(A_k) < \varepsilon$ for large k , say for $k \geq k_0$. Thus $P_n(A_{k_0}) < \varepsilon$ for large n , say for $n \geq n_0$. Since the A_k decrease, we have $\sup_{n \geq n_0} P_n(A_k) \leq \varepsilon$ for $k \geq k_0$, and since $A_k \downarrow \emptyset$, there is a k_1 such that for $n = 1, 2, \dots, n_0 - 1, P_n(A_k) < \varepsilon$ for $k \geq k_1$. Thus $\sup_n P_n(A_k) \leq \varepsilon, k \geq \max(k_0, k_1)$.

- (b) Without loss of generality, assume $P_n(\Omega) \leq 1$ for all n . Add a point (call it ∞) to the space and set $P_n\{\infty\} = 1 - P_n(\Omega) \rightarrow 1 - P(\Omega) = P\{\infty\}$. The P_n are now probability measures, and the result follows from part (a).

Section 1.6

2. $\int_{\Omega} \sum_{n=1}^{\infty} |f_n| d\mu = \sum_{n=1}^{\infty} \int_{\Omega} |f_n| d\mu < \infty$; hence $\sum_{n=1}^{\infty} |f_n|$ is integrable and therefore finite a.e. Thus $\sum_{n=1}^{\infty} f_n$ converges a.e. to a finite-valued function g .

Let $g_n = \sum_{k=1}^n f_k$. Then $|g_n| \leq \sum_{k=1}^{\infty} |f_k|$, an integrable function. By the dominated convergence theorem, $\int_{\Omega} g_n d\mu \rightarrow \int_{\Omega} g d\mu$, that is,

$$\sum_{n=1}^{\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} \sum_{n=1}^{\infty} f_n d\mu.$$

3. Let $x_0 \in (c, d)$, and let $x_n \rightarrow x_0$, $x_n \neq x_0$. Then

$$\begin{aligned} & \frac{1}{(x_n - x_0)} \left[\int_a^b f(x_n, y) dy - \int_a^b f(x_0, y) dy \right] \\ &= \int_a^b \left[\frac{f(x_n, y) - f(x_0, y)}{x_n - x_0} \right] dy. \end{aligned}$$

By the mean value theorem,

$$\frac{f(x_n, y) - f(x_0, y)}{x_n - x_0} = f_1(\lambda_n, y)$$

for some $\lambda_n = \lambda_n(y)$ between x_n and x_0 . By hypothesis, $|f_1(\lambda_n, y)| \leq h(y)$, where h is integrable, and the result now follows from the dominated convergence theorem (since $[f(x_n, y) - f(x_0, y)]/[x_n - x_0] \rightarrow f_1(x_0, y)$, $f_1(x, \cdot)$ is Borel measurable for each x).

8. Let μ be Lebesgue measure. If f is an indicator I_B , $B \in \mathcal{B}(R)$, the result to be proved states that $\mu(B) = \mu(a + B)$, which holds by translation-invariance of μ (Problem 4, Section 1.4). The passage to nonnegative simple functions, nonnegative measurable functions, and arbitrary measurable functions is done as in 1.6.12.

Section 1.7

2. (a) If f is Riemann-Stieltjes integrable, $\alpha = f = \beta$ a.e. [μ] as in 1.7.1(a). Thus the set of discontinuities of f is a subset of a set of μ -measure 0, together with the endpoints of the subintervals of the P_k . Take a different sequence of partitions having the original endpoints as

interior points to conclude that f is continuous a.e. [μ]. Conversely, if f is continuous a.e. [μ], then $\alpha = f = \beta$ a.e. [μ]. [The result that f continuous at x implies $\alpha(x) = f(x) = \beta(x)$ is true even if x is an endpoint.] As in 1.7.1(a), f is Riemann-Stieltjes integrable.

- (b) This is done exactly as in 1.7.1(b).
3. (a) By definition of the improper Riemann integral, f must be Riemann integrable (hence continuous a.e.) on each bounded interval, and the result follows. For the counterexample to the converse, take $f(x) = 1$, $n \leq x < n + 1$, n an even integer; $f(x) = -1$, $n \leq x < n + 1$, n an odd integer. Then the limit of $r_{ab}(f)$ does not exist. (Alternatively, take $f(x)$ identically 1; then $r_{ab}(f) \rightarrow +\infty$ as $a \rightarrow -\infty$, $b \rightarrow \infty$.)
- (b) Define

$$f_n(x) = \begin{cases} f(x) & \text{if } -n \leq x \leq n, \\ 0 & \text{elsewhere.} \end{cases}$$

Then $f_n \uparrow f$; hence f is measurable relative to the completed σ -field; also, $\int_{\Omega} f_n d\mu \uparrow \int_{\Omega} f d\mu$ by the monotone convergence theorem. But $\int_{\Omega} f_n d\mu = r_{-n, n}(f)$ by 1.7.1(b), and $r_{-n, n}(f) \rightarrow r(f)$ by hypothesis; the result follows.

For the counterexample, take

$$f(x) = \begin{cases} \frac{(-1)^n}{n+1}, & n \leq x < n+1, \quad n = 0, 1, \dots, \\ 0, & x < 0. \end{cases}$$

We have $r(f) = 1 - \frac{1}{2} + \frac{1}{3} - \dots$, but

$$\int_R |f| d\mu = 1 + \frac{1}{2} + \frac{1}{3} + \dots = \infty,$$

so that f is not Lebesgue integrable on R .

Chapter 2

Section 2.1

2. Let $D = \{\omega: f(\omega) < 0\}$; then $\lambda(A \cap D) \leq 0$, $\lambda(A \cap D^c) \geq 0$ for all $A \in \mathcal{F}$. By 2.1.3(d),

$$\lambda^+(A) = \lambda(A \cap D^c) = \int_A f^+ d\mu$$

since $f^+ = f$ on D^c and $f^+ = 0$ on D . Similarly,

$$\lambda^-(A) = -\lambda(A \cap D) = \int_A f^- d\mu$$

since $f^- = -f$ on D , and $f^- = 0$ on D^c . The result follows.

4. If E_1, \dots, E_n are disjoint sets in \mathcal{F} , with all $E_i \subset A$,

$$\begin{aligned} \sum_{i=1}^n |\lambda(E_i)| &= \sum_{i=1}^n |\lambda^+(E_i) - \lambda^-(E_i)| \leq \sum_{i=1}^n [\lambda^+(E_i) + \lambda^-(E_i)] \\ &= |\lambda| \left(\bigcup_{i=1}^n E_i \right) \leq |\lambda|(A). \end{aligned}$$

Thus the sup of the terms $\sum_{i=1}^n |\lambda(E_i)|$ is at most $|\lambda|(A)$. But

$$\begin{aligned} |\lambda|(A) &= \lambda^+(A) + \lambda^-(A) \\ &= \lambda(A \cap D^c) - \lambda(A \cap D) \\ &= |\lambda(A \cap D^c)| + |\lambda(A \cap D)| \\ &\quad \text{since } \lambda(A \cap D^c) \geq 0 \quad \text{and} \quad \lambda(A \cap D) \leq 0 \\ &= |\lambda(E_1)| + |\lambda(E_2)| \end{aligned}$$

and the result follows.

Section 2.2

2. Let $A_n = \{\omega : |g(\omega)| \geq 1/n\}$, $n = 1, 2, \dots$, so that $A = \bigcup_{n=1}^{\infty} A_n$. Now

$$\infty > \int_{A_n} |g| d\mu \geq \frac{1}{n} \mu(A_n);$$

hence $\mu(A_n) < \infty$. For the example, let μ be Lebesgue measure on $\mathcal{B}(R)$, and let $g(x)$ be any strictly positive integrable function, such as $g(x) = e^{-|x|}$. In this case, $A = R$, so that $\mu(A) = \infty$.

4. If f is an indicator I_A , the result is true by hypothesis. If f is a non-negative simple function $\sum_{j=1}^n x_j I_{A_j}$, the A_j disjoint sets in \mathcal{F} , then

$$\begin{aligned} \int_{\Omega} f d\lambda &= \sum_{j=1}^n x_j \lambda(A_j) = \sum_{j=1}^n x_j \int_{A_j} g d\mu = \sum_{j=1}^n x_j \int_{\Omega} I_{A_j} g d\mu \\ &= \int_{\Omega} fg d\mu \quad \text{by the additivity theorem.} \end{aligned}$$

If f is a nonnegative Borel measurable function, let f_1, f_2, \dots be non-negative simple functions increasing to f . By what we have just proved

$\int_{\Omega} f_n d\lambda = \int_{\Omega} f_n g d\mu$; hence $\int_{\Omega} f d\lambda = \int_{\Omega} fg d\mu$ by the monotone convergence theorem. Finally, if f is an arbitrary Borel measurable function, write $f = f^+ - f^-$. By what we have just proved,

$$\int_{\Omega} f^+ d\lambda = \int_{\Omega} f^+ g d\mu, \quad \int_{\Omega} f^- d\lambda = \int_{\Omega} f^- g d\mu$$

and the result follows from the additivity theorem.

6. (a) In the definition of $|\lambda|$, we may assume without loss of generality that the E_i partition A . If A is the disjoint union of sets $A_1, A_2, \dots \in \mathcal{F}$, then

$$\begin{aligned} \sum_{j=1}^n |\lambda(E_j)| &= \sum_{j=1}^n \left| \sum_{i=1}^{\infty} \lambda(E_j \cap A_i) \right| \leq \sum_{j=1}^n \sum_{i=1}^{\infty} |\lambda(E_j \cap A_i)| \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^n |\lambda(E_j \cap A_i)| \leq \sum_{i=1}^{\infty} |\lambda|(A_i). \end{aligned}$$

Thus $|\lambda|(A) \leq \sum_{i=1}^{\infty} |\lambda|(A_i)$. Now to show the reverse inequality, we may assume $|\lambda|(A) < \infty$; hence $|\lambda|(A_i) \leq |\lambda|(A) < \infty$. For each i , there is a partition $\{E_{i1}, \dots, E_{in_i}\}$ of A_i such that

$$\sum_{j=1}^{n_i} |\lambda(E_{ij})| > |\lambda|(A_i) - \frac{\varepsilon}{2^i}, \quad \varepsilon > 0 \quad \text{preassigned.}$$

Then for any n ,

$$|\lambda|(A) \geq \sum_{i=1}^n \sum_{j=1}^{n_i} |\lambda(E_{ij})| \geq \sum_{i=1}^n |\lambda|(A_i) - \varepsilon.$$

Since n and ε are arbitrary, the result follows.

- (b) If E_1, \dots, E_n are disjoint measurable subsets of A ,

$$\begin{aligned} \sum_{i=1}^n |(\lambda_1 + \lambda_2)(E_i)| &\leq \sum_{i=1}^n |\lambda_1(E_i)| + \sum_{i=1}^n |\lambda_2(E_i)| \\ &\leq |\lambda_1|(A) + |\lambda_2|(A), \end{aligned}$$

proving $|\lambda_1 + \lambda_2| \leq |\lambda_1| + |\lambda_2|$; $|a\lambda| = |a||\lambda|$ is immediate from the definition of total variation.

- (c) If $\mu(A_i) = 0$ and $|\lambda_i|(A_i^c) = 0$, $i = 1, 2$, then $\mu(A_1 \cup A_2) = 0$ and by (b), $|\lambda_1 + \lambda_2|(A_1^c \cap A_2^c) \leq |\lambda_1|(A_1^c) + |\lambda_2|(A_2^c) = 0$.
- (d) This has been established when λ is real (see 2.2.5), so assume λ complex, say, $\lambda = \lambda_1 + i\lambda_2$. If $\mu(A) = 0$, then $\lambda_1(A) = \lambda_2(A) = 0$; hence $\lambda \ll \mu$ implies $\lambda_1 \ll \mu$ and $\lambda_2 \ll \mu$. By 2.2.5(b), $|\lambda_1| \ll \mu$, $|\lambda_2| \ll \mu$; hence by (b), $|\lambda| \ll \mu$. The converse is clear since $|\lambda(A)| \leq |\lambda|(A)$.
- (e) The proof is the same as in 2.2.5(c).
- (f) See 2.2.5(d).

- (g) The “if” part is done as in 2.2.5(e); for the “only if” part, let $\mu(A_n) \rightarrow 0$. Since $|\lambda| \ll \mu$ by (d), $|\lambda|(A_n) \rightarrow 0$ by 2.2.5(e); hence $\lambda(A_n) \rightarrow 0$.

Section 2.3

2. We have

$$\begin{aligned} \int_a^b f'(x) dx &= \int_a^b \lim_{h \rightarrow 0} \left[\frac{f(x+h) - f(x)}{h} \right] dx \\ &\quad \text{where we may assume } h > 0 \\ &\leq \liminf_{h \rightarrow 0} \int_a^b \left[\frac{f(x+h) - f(x)}{h} \right] dx \\ &= \liminf_{h \rightarrow 0} \frac{1}{h} \left[\int_b^{b+h} f(x) dx - \int_a^{a+h} f(x) dx \right] \\ &\quad [\text{define } f(x) = f(b), \quad x > b; \quad f(x) = f(a), \quad x < a] \\ &\leq \liminf_{h \rightarrow 0} \frac{1}{h} [hf(b+h) - hf(a)] \quad \text{since } f \text{ is increasing} \\ &= f(b) - f(a). \end{aligned}$$

Alternatively, let μ be the Lebesgue–Stieltjes measure corresponding to f (adjusted so as to be right continuous), and let m be Lebesgue measure. Write $\mu = \mu_1 + \mu_2$, where $\mu_1 \ll m$ and $\mu_2 \perp m$. By 2.3.8 and 2.3.9,

$$\begin{aligned} \int_a^b f'(x) dx &= \int_a^b f' dm \\ &= \int_a^b D\mu dm \\ &= \int_a^b \frac{d\mu_1}{dm} dm = \mu_1(a, b] \leq \mu(a, b] = f(b) - f(a). \end{aligned}$$

6. (a) Since A is linear and m is translation-invariant, so is λ ; hence by Problem 5, Section 1.4, $\lambda = c(A)m$ for some constant $c(A)$. Now if A_1 and A_2 are linear transformations on R^k , then

$$m(A_1 A_2 E) = c(A_1)m(A_2 E) = c(A_1)c(A_2)m(E)$$

and

$$m(A_1 A_2 E) = c(A_1 A_2)m(E);$$

hence

$$c(A_1 A_2) = c(A_1)c(A_2).$$

Since $\det(A_1 A_2) = \det A_1 \det A_2$, it suffices to assume that A corresponds to an elementary row operation. Now if Q is the unit cube $\{x: 0 < x_i \leq 1, i = 1, \dots, k\}$, then $m(Q) = 1$; hence $c(A) = m(A(Q))$. If e_1, \dots, e_k is the standard basis for \mathbb{R}^k , then A falls into one of the following three categories:

- (1) $Ae_i = e_j, Ae_j = e_i, Ae_k = e_k, k \neq i \text{ or } j$. Then $c(A) = 1 = |\det A|$.
- (2) $Ae_i = ke_i, Ae_j = e_j, j \neq i$. Then $c(A) = |k| = |\det A|$.
- (3) $Ae_i = e_i + e_j, Ae_k = e_k, k \neq i$. Then $\det A = 1$ and $c(A)$ is 1 also, by the following argument. We may assume $i = 1, j = 2$. Then

$$\begin{aligned} A(Q) &= \left\{ A \sum_{i=1}^k a_i e_i : 0 < a_i \leq 1, \quad i = 1, \dots, k \right\} \\ &= \left\{ y = \sum_{i=1}^k b_i e_i : 0 < b_i \leq 1, \quad i \neq 2, \quad b_1 < b_2 \leq b_1 + 1 \right\}. \end{aligned}$$

If $B_1 = \{y \in A(Q) : b_2 \leq 1\}, B_2 = \{y \in A(Q) : b_2 > 1\}$, and $B_2 - e_2 = \{y - e_2 : y \in B_2\}$, then $B_1 \cap (B_2 - e_2) = \emptyset$; for if $y \in B_1$, then $b_1 < b_2$ and if $y + e_2 \in B_2 \subset A(Q)$, then $b_2 + 1 \leq b_1 + 1$. Therefore,

$$\begin{aligned} c(A) &= m(A(Q)) = m(B_1) + m(B_2) = m(B_1) + m(B_2 - e_2) \\ &= m(B_1 \cup (B_2 - e_2)) = m(Q) = 1. \end{aligned}$$

- (b) Fix $x \in V$ and define $S(y) = T(x + y) - T(x)$, $y \in \{z - x : z \in V\}$. Then if C is a cube containing 0, we have

$$\begin{aligned} T(x + C) &= \{T(x + y) : y \in C\} = T(x) + \{S(y) : y \in C\} \\ &= T(x) + S(C). \end{aligned}$$

Therefore, $m(T(x + C)) = m(S(C))$, so that differentiability of T at x is equivalent to differentiability of S at 0, and the derivatives, if they exist, are equal. Also, the Jacobian matrix of S at 0 coincides with the Jacobian matrix of T at x , and the result follows.

- (c) Let $A = A(0)$, and define $S(x) = A^{-1}(T(x))$, $x \in V$; the Jacobian matrix of S at 0 is the identity matrix, and $S(0) = A^{-1}(0) = 0$. If we show that the measure given by $m(S(E))$, $E \in \mathcal{B}(V)$, is differentiable

at 0 with derivative 1, then

$$\left| \frac{m(S(C))}{m(C)} - 1 \right| < \frac{\varepsilon}{|\det A|}$$

for sufficiently small cubes C containing 0. Thus by (a), $m(T(C)) = m(AS(C)) = |\det A|m(S(C))$; hence

$$\left| \frac{m(T(C))}{m(C)} - |\det A| \right| = \left| \frac{m(S(C))}{m(C)} - 1 \right| |\det A| < \varepsilon,$$

so that μ is differentiable at 0 with derivative $|\det A|$.

- (d) (i) If $x \in \bar{C}$, then $|x| \leq \sqrt{k} \beta < \delta$; hence $|T(x) - x| \leq \alpha\beta = \frac{1}{2}(\beta_2 - \beta)$. Therefore $T(x) \in C_2$.
- (ii) If $x \in \partial C$, then $|T(x) - x| \leq \alpha\beta$ as above, and $\alpha\beta = \frac{1}{2}(\beta - \beta_1)$. Therefore $T(x) \notin C_1$.
- (iii) We have $|T(x) - x| \leq \alpha\beta$, and

$$\alpha\beta = \frac{\alpha\beta_1}{1 - 2\alpha} \leq \frac{\frac{1}{2}\beta_1}{1 - \frac{1}{2}} \beta_1 = \frac{1}{2} \beta_1,$$

and the result follows.

- (iv) If $y \in C_1 - T(C)$ but $y \in T(\bar{C})$, then $y \in T(\partial C)$, contradicting (ii).

Now C_1 is the disjoint union of the sets $C_1 \cap T(C)$ and $C_1 - T(C)$; the first set is open since T is an open map, and the second set is open by (iv). By (iii), $C_1 \cap T(C) \neq \emptyset$, so by connectedness of C_1 , we have $C_1 = C_1 \cap T(C)$, that is, $C_1 \subset T(C)$. Also, $T(C) \subset C_2$ by (i).

Therefore $m(C_1) \leq m(T(C)) \leq m(C_2)$, that is,

$$(1 - 2\alpha)^k \beta^k \leq m(T(C)) \leq (1 + 2\alpha)^k \beta^k.$$

Thus $1 - \varepsilon < (1 - 2\alpha)^k < m(T(C))/m(C) \leq (1 + 2\alpha)^k < 1 + \varepsilon$ as desired.

- (e) Assume $m(E) = 0$ and $\lambda(E) > 0$. By Problem 12, Section 1.4, and the fact that

$$E = \bigcup_{n,j=1}^{\infty} E \cap \left\{ \sup_{C_r, r < 1/j} \frac{\lambda(C_r)}{m(C_r)} < n \right\},$$

we can find a compact set K and positive integers n and j such that $m(K) = 0$, $\lambda(K) > 0$, and $\lambda(C) < nm(C)$ for all open cubes C containing a point of K and having diameter less than $1/j$. If $\varepsilon > 0$, choose an open set $D \subset K$ such that $m(D) < \varepsilon$.

Now partition R^k into disjoint (partially closed) cubes B of diameter less than $1/j$ and small enough so that if $B \cap K \neq \emptyset$, then $B \subset D$. If the cubes that meet K are B_1, \dots, B_t , we may find open cubes $C_i \supset B_i$, $1 \leq i \leq t$, such that $m(C_i) < 2m(B_i)$ and $\text{diam } C_i < 1/j$. Then

$$\begin{aligned}\lambda(K) &\leq \sum_{i=1}^t \lambda(B_i) \leq \sum_{i=1}^t \lambda(C_i) < n \sum_{i=1}^t m(C_i) < 2n \sum_{i=1}^t m(B_i) \\ &\leq 2nm(D) < 2n\epsilon.\end{aligned}$$

Since ϵ is arbitrary, $\lambda(K) = 0$, a contradiction.

- (f) If f is an indicator I_B , let $E = T^{-1}(B)$, so that $B = T(E)$. Then

$$\int_W f(y) dy = m(T(E)) \quad \text{and} \quad \int_V f(T(x)) |J(x)| dx = \int_E |J(x)| dx,$$

and the result follows from part (e). The usual passage to simple functions, nonnegative measurable functions and arbitrary measurable functions completes the proof.

Section 2.4

1. (a) If $f = \{a_1, \dots, a_n, 0, 0, \dots\}$, then $\int_\Omega f d\mu = \sum_{k=1}^n a_k$ by definition of the integral of a simple function. If $f = \{a_n, n = 1, 2, \dots\}$, with all $a_n \geq 0$, $\int_\Omega f d\mu = \sum_{n=1}^\infty a_n$ by the result for simple functions and the monotone convergence theorem. If the a_n are real numbers, then $\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu = \sum_{n=1}^\infty a_n^+ - \sum_{n=1}^\infty a_n^-$ if this is not of the form $+\infty - \infty$. Finally, if the a_n are complex,

$$\int_\Omega f d\mu = \sum_{n=1}^\infty \operatorname{Re} a_n + i \sum_{n=1}^\infty \operatorname{Im} a_n$$

provided $\operatorname{Re} f$ and $\operatorname{Im} f$ are integrable; since $|\operatorname{Re} a_n|, |\operatorname{Im} a_n| \leq |a_n| \leq |\operatorname{Re} a_n| + |\operatorname{Im} a_n|$, this is equivalent to $\sum_{n=1}^\infty |a_n| < \infty$.

- (b) If $f(\alpha) = 0$ except for $\alpha \in F$, F finite, then $\int_\Omega f d\mu = \sum_\alpha f(\alpha)$ by definition of the integral of a simple function. If $f \geq 0$, then $\int_\Omega f d\mu \geq \int_F f d\mu$ for all finite F ; hence $\int_\Omega f d\mu \geq \sum_\alpha f(\alpha)$. If $f(\alpha) > 0$ for uncountably many α , then $\sum_\alpha f(\alpha) = \infty$; hence $\int_\Omega f d\mu = \infty$ also. If $f(\alpha) > 0$ for only countably many α , then $\int_\Omega f d\mu = \sum_\alpha f(\alpha)$ by the monotone convergence theorem. The remainder of the proof is as in part (a).

8. Apply Hölder's inequality with $g = 1$, f replaced by $|f|^r$, $p = s/r$, $1/q = 1 - r/s$, to obtain

$$\int_{\Omega} |f|^r d\mu \leq \left(\int_{\Omega} |f|^{rp} d\mu \right)^{1/p} \left(\int_{\Omega} 1 d\mu \right)^{1/q}.$$

Therefore $\|f\|_r \leq \|f\|_s [\mu(\Omega)]^{1/rq}$, as desired.

9. We have $\int_{\Omega} |f|^p d\mu \leq \|f\|_{\infty}^p \mu(\Omega)$, so $\lim \sup_{p \rightarrow \infty} \|f\|_p \leq \|f\|_{\infty}$. Now let $\varepsilon > 0$, $A = \{\omega: |f(\omega)| \geq \|f\|_{\infty} - \varepsilon\}$ (assuming $\|f\|_{\infty} < \infty$). Then

$$\int_{\Omega} |f|^p d\mu \geq \int_A |f|^p d\mu \geq (\|f\|_{\infty} - \varepsilon)^p \mu(A).$$

Since $\mu(A) > 0$ by definition of $\|f\|_{\infty}$, we have $\liminf_{p \rightarrow \infty} \|f\|_p \geq \|f\|_{\infty}$. If $\|f\|_{\infty} = \infty$, let $A = \{\omega: |f(\omega)| \geq M\}$ and show that

$$\liminf_{p \rightarrow \infty} \|f\|_p \geq M;$$

since M can be arbitrarily large, the result follows.

If $\mu(\Omega) = \infty$, it is still true that $\liminf_{p \rightarrow \infty} \|f\|_p \geq \|f\|_{\infty}$; for if $\mu(A) = \infty$ in the above argument, then $\|f\|_p = \infty$ for all $p < \infty$. However, if μ is Lebesgue measure on $\mathcal{B}(R)$ and $f(x) = 1$ for $n \leq x \leq n + (1/n)$, $n = 1, 2, \dots$, and $f(x) = 0$ elsewhere, then $\|f\|_p = \infty$ for $p < \infty$, but $\|f\|_{\infty} = 1$.

11. (a) We have $|\int_E (f - z) d\mu| \leq \int_E |f - z| d\mu$. But $\omega \in E$ implies $f(\omega) \in D$; hence $|f(\omega) - z| \leq r$; and thus $\int_E |f - z| d\mu \leq r\mu(E)$. If $\mu(E) > 0$, then

$$\left| \frac{1}{\mu(E)} \int_E f d\mu - z \right| = \left| \frac{1}{\mu(E)} \int_E (f - z) d\mu \right| \leq r;$$

hence $[1/\mu(E)] \int_E f d\mu \in D \subset S^c$, a contradiction. Therefore $\mu(E) = 0$, that is, $\mu\{\omega: f(\omega) \in D\} = 0$. Since $\{\omega: f(\omega) \notin S\}$ is a countable union of sets $f^{-1}(D)$, the result follows.

- (b) Let $\mu = |\lambda|$; if E_1, \dots, E_n are disjoint measurable subsets of A_r ,

$$\begin{aligned} \sum_{j=1}^n |\lambda(E_j)| &= \sum_{j=1}^n \left| \int_{E_j} h d\mu \right| \leq \sum_{j=1}^n \int_{E_j} |h| d\mu \\ &\leq r \sum_{j=1}^n \mu(E_j) \leq r\mu(A_r). \end{aligned}$$

Thus $\mu(A_r) \leq r\mu(A_r)$, and since $0 < r < 1$, we have $\mu(A_r) = 0$. If $A = \{\omega: |h(\omega)| < 1\} = \bigcup \{A_r: 0 < r < 1, r \text{ rational}\}$, then $\mu(A) = 0$, so that $|h| \geq 1$ a.e.

Now if $\mu(E) > 0$, then $[1/\mu(E)] \int_E h d\mu = \lambda(E)/\mu(E) \in S$, where $S = \{z \in C: |z| \leq 1\}$. By (a), $h(\omega) \in S$ for almost every ω , so $|h| \leq 1$ a.e. $[\lambda]$.

- (c) If $E \in \mathcal{F}$, $\int_{\Omega} I_E h d|\lambda| = \int_E h d|\lambda| = \lambda(E)$ by (b); also, $\int_{\Omega} I_E g d\mu = \int_E g d\mu = \lambda(E)$ by definition of λ . It follows immediately that $\int_{\Omega} fh d|\lambda| = \int_{\Omega} fg d\mu$ when f is a complex-valued simple function. If f is a bounded, complex-valued Borel measurable function, by 1.5.5(b), there are simple functions $f_n \rightarrow f$ with $|f_n| \leq |f|$. By the dominated convergence theorem, $\int_{\Omega} fh d|\lambda| = \int_{\Omega} fg d\mu$. If $f = hI_E$, we obtain $|\lambda|(E) = \int_E hg d\mu$.
- (d) In (c), $|\lambda|(E) \geq 0$ for all E ; hence $\bar{h}g \geq 0$ a.e. [μ] by 1.6.11. But if $g(\omega) = |g(\omega)|e^{i\theta(\omega)}$ and $h(\omega) = e^{i\varphi(\omega)}$, then $e^{i(\theta-\varphi)} = 1$ a.e. on $\{g \neq 0\}$, so that $\bar{h}g = |g|$ a.e., as desired.
12. If $I_{(a, b)}$ can be approximated in L^∞ by continuous functions, let $0 < \varepsilon < \frac{1}{2}$ and let f be a continuous function such that

$$\|I_{(a, b)} - f\|_\infty \leq \varepsilon;$$

hence $|I_{(a, b)} - f| \leq \varepsilon$ a.e. For every $\delta > 0$, there are points $x \in (a, a + \delta)$ and $y \in (a - \delta, a)$ such that $|1 - f(x)| \leq \varepsilon$ and $|f(y)| \leq \varepsilon$. Consequently, $\limsup_{x \rightarrow a^+} f(x) \geq 1 - \varepsilon$ and $\liminf_{x \rightarrow a^-} f(x) \leq \varepsilon$, contradicting continuity of f .

Section 2.5

4. Let $B_{jk\delta} = \{|f_j - f_k| \geq \delta\}$, $B_\delta = \bigcap_{n=1}^{\infty} \bigcup_{j, k=n}^{\infty} B_{jk\delta}$. Then

$$\bigcup_{j, k=n}^{\infty} B_{jk\delta} \downarrow B_\delta,$$

and the proof proceeds just as in 2.5.4.

5. Let $\{f_{n_k}\}$ be a subsequence converging a.e., necessarily to f by Problem 1. By 1.6.9, f is μ -integrable. Now if $\int_{\Omega} f_n d\mu \rightarrow \int_{\Omega} f d\mu$, then for some $\varepsilon > 0$, we have $|\int_{\Omega} f_n d\mu - \int_{\Omega} f d\mu| \geq \varepsilon$ for n in some subsequence $\{m_k\}$. But we may then extract a subsequence $\{f_{r_j}\}$ of $\{f_{m_k}\}$ converging to f a.e., so that $\int_{\Omega} f_{r_j} d\mu \rightarrow \int_{\Omega} f d\mu$ by 1.6.9, a contradiction.

Section 2.6

4. By Fubini's theorem,

$$\mu(C) = \iint_{\Omega} I_C d\mu = \int_{\Omega_1} \left[\int_{\Omega_2} I_C d\mu_2 \right] d\mu_1 = \int_{\Omega_1} \mu_2(C(\omega_1)) d\mu_1(\omega_1).$$

Similarly, $\mu(C) = \int_{\Omega_2} \mu_1(C(\omega_2)) d\mu_2(\omega_2)$. The result follows since $f \geq 0$, $\int_{\Omega} f = 0$ implies $f = 0$ a.e.

7. (a) Let

$$A_{nk} = \left\{ x \in \Omega_1 : \frac{k-1}{n} \leq f(x) < \frac{k}{n} \right\},$$

$$B_{nk} = \left\{ y \in \Omega_2 : \frac{k-1}{n} \leq y < \frac{k}{n} \right\}$$

($n = 1, 2, \dots, k = 1, 2, \dots, n$; where $k = n$, include the right endpoint as well). Then

$$G = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^n (A_{nk} \times B_{nk}) \in \mathcal{F}.$$

If f is only defined on a subset of Ω_1 , replace Ω_1 by the domain of f in the definition of A_{nk} .

- (b) Assume $B \subset C_1$. Each $x \in \Omega_1$ is countable and $(x, y) \in B$ implies $y \leq x$; hence there are only countably many points $y_{x1}, y_{x2}, \dots \in \Omega_2$ such that $(x, y_{xn}) \in B$. (If there are only finitely many points y_{x1}, \dots, y_{xn} , take $y_{xk} = y_{xn}$ for $k \geq n$.) Thus $B = \bigcup_{n=1}^{\infty} G_n$, where G_n is the graph of the function f_n defined by $f_n(x) = y_{xn}$. By part (a), $B \in \mathcal{F}$. [Note that $y_{xn} \in \Omega_2$, which may be identified with $[0, 1]$; thus (a) applies.] If $B \subset C_2$, each $y \in \Omega_2$ is countable, and $(x, y) \in B$ implies $x < y$, so there are only countably many points $x_{yn} \in \Omega_1$ such that $(x_{yn}, y) \in B$. The result follows as above.
- (c) If $F \subset \Omega$, then $F = (F \cap C_1) \cup (F \cap C_2)$ and the result follows from part (b).
9. Assuming the continuum hypothesis, we may replace $[0, 1]$ by the first uncountable ordinal β_1 . Thus we may take $\Omega_1 = \Omega_2 = \beta_1$, $\mathcal{F}_1 = \mathcal{F}_2 =$ the image of the Borel sets under the correspondence of $[0, 1]$ with β_1 , and $\mu_1 = \mu_2 =$ Lebesgue measure. Let $f = I_C$, where $C = \{(x, y) \in \Omega_1 \times \Omega_2 : y \leq x\}$ and the ordering " \leq " is taken in β_1 , not $[0, 1]$. For each x , $\{y : f(x, y) = 1\}$ is countable, and for each y , $\{x : f(x, y) = 0\}$ is countable; it follows that f is measurable in each coordinate separately. Now $\int_0^1 f(x, y) dy = 0$ for all x ; hence $\int_0^1 [\int_0^1 f(x, y) dy] dx = 0$. But $\int_0^1 [1 - f(x, y)] dx = 0$ for all y ; hence $\int_0^1 [\int_0^1 [1 - f(x, y)] dx] dy = 1$. It follows that f is not jointly measurable, for if so, the iterated integrals would be equal by Fubini's theorem.

Section 2.7

1. Let \mathcal{G} be the smallest σ -field containing the measurable rectangles. Then $\mathcal{G} \subset \prod_{j=1}^{\infty} \mathcal{F}_j$ since a measurable rectangle is a measurable cylinder. But the class of sets $A \subset \prod_{j=1}^n \Omega_j$ such that $\{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in A\} \in \mathcal{G}$ is a

σ -field that contains the measurable rectangles of $\prod_{j=1}^n \Omega_j$, and hence contains all sets in $\prod_{j=1}^n \mathcal{F}_j$. Thus all measurable cylinders belong to \mathcal{G} , so $\prod_{j=1}^\infty \mathcal{F}_j \subset \mathcal{G}$.

5.

$$\{x \in R^\infty : f(x) = n\} = \left\{ x : \sum_{i=1}^k x_i < 1, \quad k = 1, 2, \dots, n-1, \quad \sum_{i=1}^n x_i \geq 1 \right\}$$

if $n = 1, 2, \dots,$

$$\{x \in R^\infty : f(x) = \infty\} = \left\{ x : \sum_{i=1}^n x_i < 1, \quad n = 1, 2, \dots \right\}.$$

In each case we have a finite or countable intersection of measurable cylinders.

Chapter 3

Section 3.2

6. (a) Let $z \in K$, a compact subset of U . If $r < d_0$, a standard application of the Cauchy integral formula yields

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} f(z + re^{i\theta}) d\theta.$$

Thus if $0 < d < d_0$,

$$f(z) \frac{d^2}{2} = \int_0^d r f(z) dr = \frac{1}{2\pi} \int_0^d r dr \int_0^{2\pi} f(z + re^{i\theta}) d\theta,$$

or

$$\begin{aligned} f(z) &= \frac{1}{\pi d^2} \int_{\theta=0}^{2\pi} \int_{r=0}^d f(z + re^{i\theta}) r dr d\theta \\ &= \frac{1}{\pi d^2} \iint_D f(x + iy) dx dy \end{aligned}$$

where D is the disk with center at z and radius r . An application of the Cauchy-Schwarz inequality to the functions 1 and f shows that

$$|f(z)| \leq \frac{1}{\pi d^2} (\pi d^2)^{1/2} \|f\|,$$

as desired.

- (b) If f_1, f_2, \dots , is a Cauchy sequence in $H(U)$, part (a) shows that f_n converges uniformly on compact subsets to a function f analytic on U . But $H(U) \subset L^2(\Omega, \mathcal{F}, \mu)$ where $\Omega = U$, \mathcal{F} is the class of Borel sets, and μ is Lebesgue measure; hence f_n converges in L^2 to a function $g \in H(U)$. Since a subsequence of $\{f_n\}$ converges to g a.e., we have $f = g$ a.e. Therefore $f \in H(U)$ and $f_n \rightarrow f$ in L^2 , that is, in the $H(U)$ norm.

7. (a) If $0 \leq r < 1$,

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta &= \frac{1}{2\pi} \int_0^{2\pi} \sum_n a_n r^n e^{in\theta} \sum_m \bar{a}_m r^m e^{-im\theta} d\theta \\ &= \frac{1}{2\pi} \sum_{n,m} a_n \bar{a}_m r^{n+m} \int_0^{2\pi} e^{i(n-m)\theta} d\theta \end{aligned}$$

since the Taylor series
converges uniformly on compact
subsets of D

$$= \sum_{n=0}^{\infty} |a_n|^2 r^{2n},$$

which increases to $\sum_{n=0}^{\infty} |a_n|^2$ as r increases to 1.

$$\begin{aligned} \text{(b)} \quad \iint_D |f(x+iy)|^2 dx dy &= \int_0^1 r dr \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta \\ &\leq 2\pi N^2(f) \int_0^1 r dr = \pi N^2(f). \end{aligned}$$

$$\text{(c)} \quad \iint_D |f_n(x+iy)|^2 dx dy = \int_0^{2\pi} \int_0^1 r^{2n} r dr d\theta = \frac{2\pi}{(2n+2)} \rightarrow 0,$$

but

$$\frac{1}{2\pi} \int_0^{2\pi} |f_n(re^{i\theta})|^2 d\theta = \frac{1}{2\pi} \int_0^{2\pi} r^{2n} d\theta = r^{2n};$$

hence $N(f_n) = 1$ for all n .

- (d) If $\{f_n\}$ is a Cauchy sequence in H^2 , part (b) shows that $\{f_n\}$ is Cauchy in $H(D)$. By Problem 6, f_n converges uniformly on compact subsets and in $H(D)$ to a function f analytic on D . Now if $0 < r_0 < 1$, $\varepsilon > 0$, the Cauchy property in H^2 gives

$$\frac{1}{2\pi} \int_0^{2\pi} |f_n(re^{i\theta}) - f_m(re^{i\theta})|^2 d\theta \leq \varepsilon$$

for $r \leq r_0$ and n, m exceeding some integer $N(\varepsilon)$. Let $m \rightarrow \infty$; since $f_m \rightarrow f$ uniformly for $|z| \leq r_0$,

$$\frac{1}{2\pi} \int_0^{2\pi} |f_n(re^{i\theta}) - f(re^{i\theta})|^2 d\theta \leq \varepsilon, \quad r \leq r_0, \quad n \geq N(\varepsilon).$$

Since r_0 may be chosen arbitrarily close to 1, $N^2(f_n - f) \leq \varepsilon$ for $n \geq N(\varepsilon)$, proving completeness.

- (e) Since e_n corresponds to $(0, \dots, 0, 1, 0, \dots)$, with the 1 in position n , in the isometric isomorphism between H^2 and a subspace of l^2 , the e_n are orthonormal. Now if $f \in H^2$, with Taylor coefficients a_n , $n = 0, 1, \dots$, then $\langle f, e_n \rangle = a_n$, again by the isometric isomorphism. Thus

$$N^2(f) = \sum_{n=0}^{\infty} |a_n|^2 = \sum_{n=0}^{\infty} |\langle f, e_n \rangle|^2$$

and the result follows from 3.2.13(f).

$$\begin{aligned} (f) \quad \langle e_n, e_m \rangle &= \iint_D e_n(x+iy) \bar{e}_m(x+iy) dx dy = \iint_D e_n(re^{i\theta}) \bar{e}_m(re^{i\theta}) r dr d\theta \\ &= [(2n+2)(2m+2)]^{1/2} \left(\frac{1}{2\pi} \right) \int_0^{2\pi} \int_0^1 r^{n+m} e^{i(n-m)\theta} r dr d\theta \\ &= \begin{cases} 0, & n \neq m, \\ 1, & n = m. \end{cases} \end{aligned}$$

Thus the e_n are orthonormal. Now if $f \in H(D)$ with

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

then

$$\begin{aligned} &\int_0^{2\pi} \int_0^{r_0} f(re^{i\theta}) \bar{e}_m(re^{i\theta}) r dr d\theta \\ &= \sum_{n=0}^{\infty} a_n \int_0^{2\pi} \int_0^{r_0} r^n e^{in\theta} \left(\frac{2m+2}{2\pi} \right)^{1/2} r^m e^{-im\theta} r dr d\theta \end{aligned}$$

since the Taylor series converges uniformly on compact subsets of D

$$= a_m \left(\frac{2m+2}{2\pi} \right)^{1/2} \frac{r_0^{2m+2}}{2m+2} 2\pi = a_m \left(\frac{2\pi}{2m+2} \right)^{1/2} r_0^{2m+2}.$$

Now $f\bar{e}_m$ is integrable on D (by the Cauchy–Schwarz inequality), so we may let $r_0 \rightarrow 1$ and invoke the dominated convergence theorem to obtain

$$\langle f, e_m \rangle = a_m \left(\frac{2\pi}{2m+2} \right)^{1/2}.$$

But the same argument with e_m replaced by f shows that

$$\begin{aligned} \|f\|_{H(D)}^2 &= \lim_{r_0 \rightarrow 1} \sum_{n,m=0}^{\infty} a_n \bar{a}_m \int_0^{2\pi} \int_0^{r_0} r^n e^{in\theta} r^m e^{-im\theta} r dr d\theta \\ &= \sum_{n=0}^{\infty} \frac{|a_n|^2 2\pi}{2n+2}. \end{aligned}$$

The result now follows from 3.2.13(f).

9. (a) Let g be a continuous complex-valued function on $[0, 2\pi]$ with $g(0) = g(2\pi)$. Then $g(t) = h(e^{it})$, where $h(z)$ is continuous on $\{z \in C : |z| = 1\}$. By the Stone–Weierstrass theorem, h can be uniformly approximated by functions of the form $\sum_{k=-n}^n c_k z^k$. For the algebra generated by z^n , $n = 0, \pm 1, \pm 2, \dots$, separates points, contains the constant functions, and contains the complex conjugate of each of its members since $\bar{z} = 1/z$ for $|z| = 1$.

Thus $g(t)$ can be uniformly approximated (hence approximated in L^2) by trigonometric polynomials $\sum_{k=-n}^n c_k e^{ikt}$. Since any continuous function on $[0, 2\pi]$ can be approximated in L^2 by a continuous function with $g(0) = g(2\pi)$, and the continuous functions are dense in L^2 , the trigonometric polynomials are dense in L^2 .

- (b) By 3.2.6, $\int_0^{2\pi} |f(t) - \sum_{k=-n}^n c_k e^{ikt}|^2 dt$ is minimized when $c_k = a_k = (1/2\pi) \int_0^{2\pi} f(t) e^{-ikt} dt$. Furthermore, some sequence of trigonometric polynomials converges to f in L^2 since the trigonometric polynomials are dense. The result follows.
(c) This follows from part (a) and 3.2.13(c), or, equally well, from part (b) and 3.2.13(d).
10. (a) Let $\{e_n\}$ be an infinite orthonormal subset of H . Take $M = \{x_1, x_2, \dots\}$, where $x_n = (1 + 1/n)e_n$, $n = 1, 2, \dots$. To show that M is closed, we compute, for $n \neq m$,

$$\begin{aligned} \|x_n - x_m\|^2 &= \left\| \left(1 + \frac{1}{n}\right) e_n - \left(1 + \frac{1}{m}\right) e_m \right\|^2 \\ &= \left(1 + \frac{1}{n}\right)^2 + \left(1 + \frac{1}{m}\right)^2 \geq 2. \end{aligned}$$

Thus if $y_n \in M$, $y_n \rightarrow y$, then $y_n = y$ eventually, so $y \in M$. Since $\|x_n\|^2 = 1 + (1/n)$, M has no element of minimum norm.

- (b) Let M be a nonempty closed subset of the finite-dimensional space H . If $x \in H$ and $N = M \cap \{y: \|x - y\| \leq n\}$, then $N \neq \emptyset$ for some n . Since $y \rightarrow \|x - y\|$, $y \in N$, is continuous and N is compact, $\inf\{\|x - y\|: y \in N\} = \|x - y_0\|$ for some $y_0 \in N \subset M$. But the inf over N is the same as the inf over M ; for if $y \in M$, $y \notin N$, then $\|x - y_0\| \leq n < \|x - y\|$. Note that y_0 need not be unique; for example, let $H = \mathbb{R}$, $M = \{-1, 1\}$, $x = 0$.

Section 3.3

3. Since $\int_a^b |K(s, t)| dt$ is continuous in s , it assumes a maximum at some point $u \in [a, b]$. If $K(u, t) = r(t)e^{i\theta(t)}$, $r(t) \geq 0$, let $z(t) = e^{-i\theta(t)}$. Let x_1, x_2, \dots be a sequence in $C[a, b]$ such that $\int_a^b |x_n(t) - z(t)| dt \rightarrow 0$; we may assume that $|x_n(t)| \leq 1$ for all n and t (see 2.4.14). Since K is bounded,

$$\left| \int_a^b K(s, t)z(t) dt \right| = \lim_{n \rightarrow \infty} \left| \int_a^b K(s, t)x_n(t) dt \right| = \lim_{n \rightarrow \infty} |(Ax_n)(s)| \leq \|A\|.$$

Set $s = u$ to obtain

$$\int_a^b |K(u, t)| dt \leq \|A\|,$$

as desired.

7. (a) If $x \in L$, then

$$\|x\|_1 = \left\| \sum_{i=1}^n x_i e_i \right\|_1 \leq \sum_{i=1}^n |x_i| \|e_i\|_1 \leq \left(\max_i \|e_i\|_1 \right) \sum_{i=1}^n |x_i|.$$

But $\sum_{i=1}^n |x_i| \leq \sqrt{n} \|x\|_2$, so we may take $k = \sqrt{n} \max_i \|e_i\|_1$.

- (b) Let $S = \{x: \|x\|_2 = 1\}$. Since $(L, \|\cdot\|_2)$ is isometrically isomorphic to C^n , S is compact in the norm $\|\cdot\|_2$. Now the map $x \rightarrow \|x\|_1$ is a continuous real-valued function on $(L, \|\cdot\|_1)$, and by part (a), the topology induced by $\|\cdot\|_1$ is weaker than the topology induced by $\|\cdot\|_2$. Thus the map is continuous on $(L, \|\cdot\|_2)$; hence it attains a minimum on S , necessarily positive since $x \in S$ implies $x \neq 0$.
- (c) If $x \in L$, $x \neq 0$, let $y = x/\|x\|_2$; then $\|y\|_1 \geq m\|y\|_2$ by (b); hence $\|x\|_1 \geq m\|x\|_2$. By (a) and Problem 6(b), $\|\cdot\|_1$ and $\|\cdot\|_2$ induce the same topology.
- (d) By the above results, the map $T: \sum_{i=1}^n x_i e_i \rightarrow (x_1, \dots, x_n)$ is a one-to-one onto, linear, bicontinuous map of L and C^n [note that $\|\sum_{i=1}^n x_i\|_2$ is the Euclidean norm of (x_1, \dots, x_n) in C^n]. If $y_j \in L$,

$y_j \rightarrow y \in M$, then $y_j - y_k \rightarrow 0$ as $j, k \rightarrow \infty$; hence $T(y_j - y_k) = Ty_j - Ty_k \rightarrow 0$. Thus $\{Ty_j\}$ is a Cauchy sequence in C^n . If $Ty_j \rightarrow z \in C^n$, then $y_j \rightarrow T^{-1}z \in L$.

9. For (a) implies (b), see Problem 7; if (c) holds, then $\{x : \|x\| \leq \varepsilon\}$ is compact for small enough $\varepsilon > 0$; hence every closed ball is compact (note that the map $x \rightarrow kx$ is a homeomorphism). But any closed bounded set is a subset of a closed ball, and hence is compact.

To prove that (f) implies (a), choose $x_1 \in L$ such that $\|x_1\| = 1$. Suppose we have chosen $x_1, \dots, x_k \in L$ such that $\|x_i\| = 1$ and $\|x_i - x_j\| \geq \frac{1}{2}$ for $i, j = 1, \dots, k$, $i \neq j$. If L is not finite-dimensional, then $S\{x_1, \dots, x_k\}$ is a proper subspace of L , necessarily closed, by Problem 7(d). By Problem 8, we can find $x_{k+1} \in L$ with $\|x_{k+1}\| = 1$ and $\|x_i - x_{k+1}\| \geq \frac{1}{2}$, $i = 1, \dots, k$. The sequence x_1, x_2, \dots satisfies $\|x_n\| = 1$ for all n , but $\|x_n - x_m\| \geq \frac{1}{2}$ for $n \neq m$; hence the unit sphere cannot possibly be covered by a finite number of balls of radius less than $\frac{1}{4}$.

11. (a) Define $\lambda(A) = f(I_A)$, $A \in \mathcal{F}$. If A_1, A_2, \dots are disjoint sets in \mathcal{F} whose union is A , then $\lambda(A) = \sum_{i=1}^{\infty} f(I_{A_i})$ since f is continuous and $\sum_{i=1}^n I_{A_i} \xrightarrow{L^p} I_A$. [Note that

$$\int_{\Omega} \left| \sum_{i=1}^n I_{A_i} - I_A \right|^p d\mu = \sum_{i=n+1}^{\infty} \mu(A_i) \rightarrow 0$$

by finiteness of μ .] Thus λ is a complex measure on \mathcal{F} . If $\mu(A) = 0$, then $I_A = 0$ a.e. [μ], so we may write $I_A \xrightarrow{L^p} 0$ and use the continuity of f to obtain $\lambda(A) = 0$. By the Radon-Nikodym theorem, we have $\lambda(A) = \int_A y d\mu$ for some μ -integrable y . Thus $f(x) = \int_{\Omega} xy d\mu$ when x is an indicator; hence when x is a simple function. Since f is continuous, y is μ -integrable, and the finite-valued simple functions are dense in L^p , the result holds when x is a bounded Borel measurable function.

Now let y_1, y_2, \dots be nonnegative, finite-valued, simple functions increasing to $|y|$. Then

$$\begin{aligned} \|y_n\|_q^q &= \int_{\Omega} y_n^q d\mu \leq \int_{\Omega} y_n^{q-1} |y| d\mu = \int_{\Omega} y_n^{q-1} e^{-i \arg y} y d\mu \\ &= f(y_n^{q-1} e^{-i \arg y}) \quad \text{since } y_n^{q-1} e^{-i \arg y} \text{ is bounded} \\ &\leq \|f\| \|y_n^{q-1}\|_p = \|f\| \|y_n\|_q^{q/p} \quad \text{since } (q-1)p = q. \end{aligned}$$

Thus $\|y_n\|_q \leq \|f\|$; hence by the monotone convergence theorem, $\|y\|_q \leq \|f\|$; in particular, $y \in L^q$. But now Hölder's inequality and the fact that finite-valued simple functions are dense in L^p yield $f(x) = \int_{\Omega} xy d\mu$ for all $x \in L^p$. Hölder's inequality also gives

$\|f\| \leq \|y\|_q$; hence $\|f\| = \|y\|_q$. If y_1 is another such function, then $g(x) = \int_{\Omega} x(y - y_1) d\mu = 0$ for all $x \in L^p$. By the above argument, $\|y - y_1\|_q = 0$, so $y = y_1$ a.e. $[\mu]$.

- (b) (i) If, say, $y_A - y_B > 0$ on the set $F \subset A \cap B$, let $x = I_F$; then $xI_A = xI_B$; hence $\int_F x(y_A - y_B) d\mu = 0$, that is,

$$\int_F (y_A - y_B) d\mu = 0.$$

But then $\mu(F) = 0$.

- (ii) Since $y_{A_n \cup A_m} = y_{A_n}$ a.e. on A_n we have

$$\|y_{A_n}\|_q^q \leq \|y_{A_n \cup A_m}\|_q^q = \|y_{A_n}\|_q^q + \int_{A_m - A_n} |y_{A_m}|^q d\mu.$$

Since $\|y_{A_n}\|_q^q$ approaches k^q as $n \rightarrow \infty$, so does $\|y_{A_n \cup A_m}\|_q^q$, and it follows that $\int_{A_m - A_n} |y_{A_m}|^q d\mu \rightarrow 0$ as $n \rightarrow \infty$. By symmetry, we may interchange m and n to obtain

$$\int_{\Omega} |y_{A_m} - y_{A_n}|^q d\mu = \int_{A_m - A_n} |y_{A_m}|^q d\mu + \int_{A_n - A_m} |y_{A_n}|^q d\mu \rightarrow 0$$

as $n, m \rightarrow \infty$. Thus y_{A_n} converges in L^q to a limit y , and since $\|y_{A_n}\|_q \leq \|f\|$ for all n , $\|y\|_q \leq \|f\|$. If $\{B_n\}$ is another sequence of sets with $\|y_{B_n}\|_q \rightarrow k$, the above argument with A_n replaced by B_n shows that $\|y_{A_n} - y_{B_n}\|_q \rightarrow 0$; hence $y_{B_n} \rightarrow y$ also.

- (iii) Let $A \in \mathcal{F}$, $\mu(A) < \infty$. In (ii) we may take all $A_n \supset A$, so that $y_{A_n} = y_A$ a.e. on A ; hence $y = y_A$ a.e. on A . Thus if $x = I_A$, then $f(x) = f(xI_A) = \int_{\Omega} xy_A d\mu = \int_{\Omega} xy d\mu$. It follows that $f(x) = \int_{\Omega} xy d\mu$ if x is simple. [If $\mu(B) = \infty$, then x must be 0 on B since $x \in L^p$.] Since $y \in L^q$, the continuity of f and Hölder's inequality extend this result to all $x \in L^p$.
- (c) The argument of (a) yields a μ -integrable y such that $f(x) = \int_{\Omega} xy d\mu$ for all bounded Borel measurable x . Let $B = \{\omega: |y(\omega)| \geq k\}$; then

$$\begin{aligned} k\mu(B) &\leq \int_B |y| d\mu = \int_{\Omega} I_B e^{-i \arg y} y d\mu \\ &= f(I_B e^{-i \arg y}) \leq \|f\| \|I_B\|_1 = \|f\| \mu(B). \end{aligned}$$

Thus if $k > \|f\|$, we have $\mu(B) = 0$, proving that $y \in L^\infty$ and $\|y\|_\infty \leq \|f\|$. As in (a), we obtain $f(x) = \int_{\Omega} xy d\mu$ for all $x \in L^1$, $\|f\| = \|y\|_\infty$, and y is essentially unique.

- (d) Part (i) of (b) holds, with the same proof. Now if Ω is the union of disjoint sets A_n , with $\mu(A_n) < \infty$, define y on Ω by taking $y = y_{A_n}$ on A_n . Since $\|y_{A_n}\|_\infty \leq \|f\|$ for all n , we have $y \in L^\infty$ and $\|y\|_\infty \leq$

$\|f\|$. If $x \in L^1$, then $\sum_{i=1}^n x I_{A_i} \xrightarrow{L^1} x$; hence

$$f(x) = \sum_{n=1}^{\infty} f(x I_{A_n}) = \sum_{n=1}^{\infty} \int_{\Omega} x y_{A_n} d\mu = \sum_{n=1}^{\infty} \int_{A_n} x y d\mu = \int_{\Omega} x y d\mu.$$

Since $\|f\| \leq \|y\|_{\infty}$ by Hölder's inequality (with $p = 1, q = \infty$), the result follows.

Section 3.4

3. Let $\{y_n\}$ be a Cauchy sequence in M , and let x_0 be any element of L with norm 1. By 3.4.5 (c), there is an $f \in L^*$ with $\|f\| = 1$ and $f(x_0) = \|x_0\| = 1$; we define $A_n \in [L, M]$ by $A_n x = f(x)y_n$. Then $\|(A_n - A_m)x\| = |f(x)| \|y_n - y_m\| \leq \|y_n - y_m\| \|x\|$, so that $\|A_n - A_m\| \leq \|y_n - y_m\| \rightarrow 0$. By hypothesis, the A_n converge uniformly to some $A \in [L, M]$; therefore $\|y_n - Ax_0\| = \|A_n x_0 - Ax_0\| \leq \|A_n - A\| \rightarrow 0$.
7. Let L be the set of all scalar-valued functions on Ω , with sup norm, and let M be the subspace of L consisting of simple functions

$$x = \sum_j x_j I_{A_j},$$

where the A_j are disjoint sets in \mathcal{F}_0 . Define g on M by

$$g(x) = \sum_j x_j \mu_0(A_j).$$

Now $|g(x)| \leq \max_j |x_j| \sum_j \mu_0(A_j) \leq \max_j |x_j| \mu_0(\Omega) = \mu_0(\Omega) \|x\|$; hence $\|g\| \leq \mu_0(\Omega) < \infty$. By the Hahn–Banach theorem, g has an extension to a continuous linear functional f on L , with $\|f\| = \|g\|$. Define $\mu(A) = f(I_A)$, $A \subset \Omega$. Since f is linear, μ is finitely additive, and since f is an extension of g , μ is an extension of μ_0 . Now if $\mu(A) < 0$, then

$$f(I_{A^c}) = \mu(A^c) = \mu(\Omega) - \mu(A) > \mu(\Omega).$$

But $\|I_{A^c}\| = 1$, so that $\|f\| > \mu(\Omega)$, a contradiction.

8. Since $A_n x \rightarrow Ax$ for each x , $\sup_n \|A_n x\| < \infty$. By the uniform boundedness principle, $\sup_n \|A_n\| = M < \infty$; hence

$$\begin{aligned} \|Ax\| &= \lim_{n \rightarrow \infty} \|A_n x\| = \liminf_{n \rightarrow \infty} \|A_n x\| \\ &\leq \|x\| \liminf_{n \rightarrow \infty} \|A_n\| \leq M \|x\|. \end{aligned}$$

12. Let L be the set of all complex-valued functions x on $[0, 1]$ with a continuous derivative x' , M the set of all continuous complex-valued functions on $[0, 1]$, with the sup norm on L and M . If $Ax = x'$, $x \in L$, then A is a linear map of L onto M , and A is closed. For if $x_n \rightarrow x$ and $x_n' \rightarrow y$, then since convergence relative to the sup norm is uniform convergence,

$\int_0^t x_n'(s) ds \rightarrow \int_0^t y(s) ds = z(t)$. Thus $x_n(t) - x_n(0) \rightarrow z(t)$; hence $x(t) - x(0) = z(t)$. Therefore $x' = z' = y$. But A is unbounded, for if $x_n(t) = \sin nt$, then $\|x_n\| = 1$, $\|Ax_n\| \rightarrow \infty$.

13. By the open mapping theorem, A is open; hence $A\{x \in L: \|x\| < 1\}$ is a neighborhood of 0 in M ; say $\{y \in M: \|y\| < \delta\} \subset A\{x \in L: \|x\| < 1\}$. If $y \in M$, $y \neq 0$, then $\delta y/2\|y\|$ has norm less than δ , hence equals Az for some $z \in L$, $\|z\| < 1$. If $x = (2\|y\|/\delta)z$, then $Ax = y$ and $\|x\| < 2\|y\|/\delta$. Thus we may take $k = 2/\delta$.

Section 3.5

3. Let x_1, \dots, x_n be a basis for L , and define $A: C^n \rightarrow L$ by $A(a_1, \dots, a_n) = a_1x_1 + \dots + a_nx_n$. It is immediate that A is one-to-one onto, linear, and continuous; we must also show that A^{-1} is continuous. If L is one-dimensional, then $A^{-1}(ax) = a$, so the null space of A^{-1} is $\{0\}$, which is closed by the Hausdorff hypothesis. By Problem 2, A^{-1} is continuous. If $\dim L = n + 1$ and the result holds up to dimension n , let f be a non-trivial linear functional on L . Then [see 3.3.4(a)] $N = f^{-1}\{0\}$ is a maximal proper subspace of L ; hence by the induction hypothesis, it is topologically isomorphic to C^n . It follows that N is closed in L ; the argument is the same as in Problem 7(d), Section 3.3.

By Problem 2, every linear functional on L is continuous, so if p_i is the projection of C^n on the i th coordinate space, $p_i \circ A^{-1}$ is continuous for each i , hence A^{-1} is continuous.

5. If L were metrizable, it would have a countable base at 0, which can be assumed to be of the form

$$U_n = \{f \in L: |f(x)| < \delta_n \text{ for all } x \in A_n\},$$

where $\delta_n > 0$ and A_n is a finite subset of $[0, 1]$, $n = 1, 2, \dots$. Let y be a point in $[0, 1]$ not belonging to $\bigcup_{n=1}^{\infty} A_n$, and let $U = \{f \in L: |f(y)| < 1\}$. Then U is a neighborhood of 0 in L but no $U_n \subset U$, for there are (continuous) functions f with $f = 0$ on A_n but $f(y) = 1$.

6. (a) implies (b): The sets $\{x: d(x, 0) < r\}$, r rational, form a base at 0.
 (b) implies (c): The base at 0 may be assumed to consist of circled convex sets, and the proof of 3.5.7 shows that the corresponding Minkowski functionals generate the topology of L .
 (c) implies (a): If p_1, p_2, \dots are seminorms generating the topology of L , define

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(x - y)}{1 + p_n(x - y)}.$$

It is easily checked that d is a pseudometric, and $d(x_i, x) \rightarrow 0$ iff $p_n(x_i - x) \rightarrow 0$ for each n , in other words, iff $x_i \rightarrow x$ in the topology of L .

7. Let $K_n = \{z: |z| \leq 1 - 1/n\}$, $n = 1, 2, \dots$. If K is any compact subset of D , $K \subset K_n$ for large n ; hence $\sup\{|f(z)|: z \in K\} \leq \sup\{|f(z)|: z \in K_n\} = p_n(f)$. Thus the sets $U_n = \{f \in L: p_n(f) \leq 1/n\}$, $n = 1, 2, \dots$, form a countable base at 0, so by Problem 6, L is metrizable.

Suppose a single norm $\|\cdot\|$ induces the topology of L . By continuity of p_n , for each n there is a $\delta_n > 0$ such that $\|f\| < \delta_n$ implies $p_n(f) < 1$; hence $\|f\| < 1$ implies $p_n(f) < 1/\delta_n$. Therefore $W = \{f \in L: p_n(f) < 1/\delta_n \text{ for all } n\}$ is an overneighborhood of 0.

For each n let z_n be a point in K_{n+1} but not in K_n , and let $f_n \in U_n$ such that $|f_n(z_n)| > 1/\delta_{n+1}$; a choice of f_n is possible because, for example, if

$$1 - \frac{1}{n} < |z_0| < |z_n| \leq 1 - \frac{1}{n+1},$$

then $(z/z_0)^k$ approaches 0 uniformly on K_n as $k \rightarrow \infty$, but approaches ∞ at z_n .

Now $p_{n+1}(f_n) \geq |f_n(z_n)| > 1/\delta_{n+1}$, and since the U_n decrease with n and $f_n \in U_n$, we have, for any k , $f_n \in U_k$ for all $n \geq k$; consequently $f_n \rightarrow 0$. But then $f_n \in W$ for large n ; hence $p_{n+1}(f_n) < 1/\delta_{n+1}$, a contradiction.

8. (a) We may write $x = xI_{[0, r)} + xI_{[r, 1]} = y + z$, and

$$d(x, 0) = \int_0^1 |x(t)|^p dt = d(y, 0) + d(z, 0)$$

since $[0, r) \cap [r, 1] = \emptyset$. Now $d(y, 0) = \int_0^r |x(t)|^p dt$, which is continuous in r , approaches 0 as $r \rightarrow 0$, and approaches $d(x, 0)$ as $r \rightarrow 1$. By the intermediate value theorem, $d(y, 0) = d(z, 0) = \frac{1}{2}d(x, 0)$ for some choice of r .

- (b) By part (a) we can find y_1 with $d(y_1, 0) = \frac{1}{2}d(x, 0)$ and $|f(y_1)| \geq \frac{1}{2}$.

Let $x_1 = 2y_1$; then $|f(x_1)| \geq 1$ and $d(x_1, 0) = 2^p d(y_1, 0) = 2^{p-1} d(x, 0)$. Having chosen x_1, \dots, x_n with $|f(x_i)| \geq 1$ and $d(x_i, 0) = 2^{i(p-1)} d(x, 0)$, $i = 1, \dots, n$, apply part (a) to x_n to obtain x_{n+1} with $|f(x_{n+1})| \geq 1$ and $d(x_{n+1}, 0) = 2^{p-1} d(x_n, 0) = 2^{(n+1)(p-1)} d(x, 0)$. Since $p < 1$, $d(x_n, 0) \rightarrow 0$ as $n \rightarrow \infty$, as desired.

13. (a) Let i be the identity map from (L, \mathcal{T}_2) to (L, \mathcal{T}_1) . Since $\mathcal{T}_1 \subset \mathcal{T}_2$, i is continuous, and by the open mapping theorem, $\mathcal{T}_2 \subset \mathcal{T}_1$.
(b) If \mathcal{T}_j is the topology induced by $\|\cdot\|_j$, $j = 1, 2$, then $\mathcal{T}_1 \subset \mathcal{T}_2$ by hypothesis, and the result follows from part (a).
15. Define $T: L \rightarrow C^n$ by $T(x) = (f_1(x), \dots, f_n(x))$ and define $h: T(L) \rightarrow C^n$ by $h(Tx) = g(x)$. Then h is well-defined; for if $Tx_1 = Tx_2$, then $f_i(x_1) = f_i(x_2)$ for all i , so $g(x_1) = g(x_2)$ by hypothesis. Since h is linear on

$T(L)$, it may be extended to a linear functional on C^n , necessarily of the form $h(y_1, \dots, y_n) = c_1y_1 + \dots + c_ny_n$. Thus $g(x) = h(Tx) = c_1f_1(x) + \dots + c_nf_n(x)$ for all $x \in L$.

16. (a) Assume $f_i(x) = 0$ for all i . If k is any real number, $f_i(kx) = 0$ for all i , hence $\|kx\| < 1$. Since k is arbitrary, x must be 0.
- (b) Since the y_j are linearly independent and T^{-1} is one-to-one, the x_j are linearly independent. If $x \in L$, then $Tx = \sum_{j=1}^k c_jy_j$ for some c_1, \dots, c_k ; hence $x = \sum_{j=1}^k c_jx_j$.

Chapter 4

Section 4.2

2. Let Ω be the first uncountable ordinal, with \mathcal{F} the class of all subsets of Ω , and $\mu(A) = 0$ if A is countable, $\mu(A) = \infty$ if A is uncountable. Define, for each $\alpha \in \Omega$, $f_\alpha(\omega) = 1$ if $\omega \leq \alpha$; $f_\alpha(\omega) = 0$ if $\omega > \alpha$. Then $f_\alpha \uparrow f$, where $f \equiv 1$, and $\int_\Omega f_\alpha d\mu = 0$ for all α since f_α is the indicator of a countable set. But $\int_\Omega f d\mu = \infty$, so the monotone convergence theorem fails.

Section 4.3

4. (a) First note that H is a vector space. For if g is continuous, $\{f \in H: f + g \in H\}$ contains the continuous functions and is closed under pointwise limits of monotone sequences, and hence coincides with H . Thus if $f \in H$ and g is continuous, then $f + g \in H$. A repetition of this argument (notice the bootstrapping technique) shows that if $g \in H$, then $\{f \in H: f + g \in H\} = H$; hence H is closed under addition. Now if a is real, then $\{f \in H: af \in H\} = H$ by the same argument; hence H is closed under scalar multiplication.

Let \mathcal{S} be the open F_σ sets. Then \mathcal{S} is closed under finite intersection, and $I_A \in H$ for all $A \in \mathcal{S}$. (By 4.3.5, I_A is the limit of an increasing sequence of continuous functions.) By 4.1.4, $I_A \in H$ for all $A \in \sigma(\mathcal{S})$ [$= \mathcal{A}(\Omega)$ by 4.3.2]. The usual passage to nonnegative simple functions, nonnegative measurable functions, and arbitrary measurable functions shows that all Baire measurable functions belong to H . But the class of Baire measurable functions contains the continuous functions and is closed under pointwise limits of monotone sequences; hence H is the class of Baire measurable functions.

- (b) All functions in H are Baire measurable by part (a), hence $\sigma(H) \subset \mathcal{A}$. But if $A \in \mathcal{A}$, then I_A is Baire measurable, so $I_A \in H$ by part (a).

But then I_A is $\sigma(H)$ -measurable by definition of $\sigma(H)$, hence $A \in \sigma(H)$.

5. Let H be the class of Baire measurable functions. Then $K_0 \subset H$, and if $K_\beta \subset H$ for all $\beta < \alpha$, then $K_\alpha \subset H$ since H is closed under monotone limits. Thus $K \subset H$. But K contains the continuous functions and is closed under pointwise limits of monotone sequences. For if $f_1, f_2, \dots \in K$ and $f_n \rightarrow f$, say $f_n \in K_{\alpha_n}$, $n = 1, 2, \dots$, where $\alpha_n < \beta_1$. If $\alpha = \sup_n \alpha_n < \beta_1$, then $f_n \in K_\alpha$ for all n ; hence $f \in K_{\alpha+1} \subset K$. By minimality of H (Problem 4), $H \subset K$.
6. Existence of μ and λ follows from 4.3.13 applied to the real and imaginary parts of E . Now if $\int_\Omega f d\mu = 0$ for all $f \in C(\Omega)$, then $\int_\Omega g d\mu_1 = \int_\Omega g d\mu_2 = 0$ for all real-valued continuous functions on Ω ; hence $\mu_1 = \mu_2 = 0$ by 4.3.13. This proves uniqueness of μ (and similarly, uniqueness of λ).

Now if $f \in C(\Omega)$, then $\int_\Omega f d\lambda = \int_\Omega fh d|\lambda|$ for some complex-valued Borel measurable h of absolute value 1; hence $|E(f)| \leq \int_\Omega |f| d|\lambda|$; therefore $\|E\| \leq |\lambda|(\Omega)$. On the other hand, let E_1, \dots, E_n be disjoint Borel subsets of Ω , with $\lambda(E_j) = r_j e^{i\theta_j}$, $j = 1, \dots, n$. Define $f = \sum_{j=1}^n e^{-i\theta_j} I_{E_j}$ and, for a given $\delta > 0$, let g be a continuous complex-valued function on Ω such that $|g| \leq 1$ and $\int_\Omega |f - g| d|\lambda| < \delta$ (see 4.3.14). Now

$$\int_\Omega f d\lambda = \sum_{j=1}^n r_j = \sum_{j=1}^n |\lambda(E_j)|,$$

hence

$$|E(g)| = \left| \int_\Omega g d\lambda \right| \geq \left| \int_\Omega f d\lambda \right| - \delta = \sum_{j=1}^n |\lambda(E_j)| - \delta.$$

Since δ is arbitrary, it follows that $\|E\| \geq |\lambda|(\Omega)$. But just as in 4.3.13, λ is the unique extension of μ , so $|\lambda|(\Omega) = |\mu|(\Omega)$, completing the proof.

Section 4.4

2. Since there are c Borel subsets of R^n , there are only c measurable rectangles in \mathcal{F} . Thus (Problem 11, Section 1.2) $\text{card } \mathcal{F} \leq c$. But $\text{card } \mathcal{F} \geq c$ (consider $\{\omega : \omega_i = a\}$, $a \in R$); hence $\text{card } \mathcal{F} = c$.
3. Let \mathcal{C} be the class of sets in \mathcal{F} for which the conclusion holds. If $A \in \mathcal{C}$, then $\omega \in A^c$ iff $\omega_{T_0} \notin B_0$, that is, iff $\omega_{T_0} \in B_0^c$; thus $A^c \in \mathcal{C}$. If $A_n \in \mathcal{C}$, $n = 1, 2, \dots$, with $\omega \in A_n$ iff $\omega_{T_n} \in B_{T_n}$, let $T_0 = \bigcup_{n=1}^\infty T_n$, a countable subset of T . Then $\omega \in \bigcup_{n=1}^\infty A_n$ iff $\omega_{T_0} \in B_{T_0}$, where $B_{T_0} = \{y \in \Omega_{T_0} : y_{T_n} \in B_{T_n} \text{ for some } n\}$. Now $\{y \in \Omega_{T_0} : y_{T_n} \in B_{T_n}\} = p_n^{-1}(B_{T_n})$, where p_n is the projection of Ω_{T_0} onto Ω_{T_n} . Thus $p_n : (\Omega_{T_0}, \mathcal{F}_{T_0}) \rightarrow (\Omega_{T_n}, \mathcal{F}_{T_n})$, so

that $B_{T_0} \in \mathcal{F}_{T_0}$, and consequently $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$. Therefore \mathcal{C} is a σ -field containing the measurable cylinders, so $\mathcal{C} = \mathcal{F}$.

4. (a) If the set A in question belongs to \mathcal{F} let T_0 and B_0 be as in Problem 3; assume without loss of generality that $t_0 \in T_0$. Since T_0 is countable, there is a sequence $t_n \rightarrow t_0$ with $t_n \notin T_0$. Define $x(t) = 0$, $t \in T_0$; $x(t) = 1$, $t \notin T_0$. Now if $\omega(t) = 0$ for all t , then $\omega \in A$, hence $\omega_{T_0} \in B_0$. But $\omega_{T_0} = x_{T_0}$, so that $x \in A$. Since x is discontinuous at t_0 , this is a contradiction.
- (b) Again assume the set A belongs to \mathcal{F} and let T_0 and B_0 be as in Problem 3. Since T_0 is countable, we can find $s \in T$ with $s \notin T_0$. Define $x(t) = c - 1$, $t \neq s$; $x(s) = c + 1$, and let $\omega(t) = c - 1$ for all t . Then $\omega \in A$ and $\omega(t) = x(t)$ for $t \in T_0$; hence $x \in A$, a contradiction.
5. (a) Let $F = \bigcap_{n=1}^{\infty} V_n$, V_n open. Let g_n be a continuous function from Ω_t to $[0, 1]$ with $g_n = 0$ on F , $g_n = 1$ on V_n^c . Let $g = \sum_{n=1}^{\infty} 2^{-n} g_n$. Then g is continuous and $g^{-1}\{0\} = F$. Define $f: \Omega \rightarrow R$ by $f(\omega) = g(\omega(t))$. Then $f \in C(\Omega)$ and $\{\omega: f(\omega) = 0\} = A$; therefore $A \in \mathcal{A}(\Omega)$. But if \mathcal{C} is the class of subsets B of Ω_t such that $\{\omega \in \Omega: \omega(t) \in B\} \in \mathcal{A}(\Omega)$, then \mathcal{C} is a σ -field containing the closed sets; hence $\mathcal{C} = \mathcal{F}_t$. It follows that all measurable rectangles belong to $\mathcal{A}(\Omega)$; hence $\mathcal{F} \subset \mathcal{A}(\Omega)$.
- (b) If $x, y \in \Omega$, $x \neq y$, then for some t , we have $x(t) \neq y(t)$. Let g be a continuous function from Ω_t to R with $g(x(t)) \neq g(y(t))$. Define $f(\omega) = g(\omega(t))$, $\omega \in \Omega$. Then $f \in C(\Omega)$, f depends on only one coordinate, and $f(x) \neq f(y)$. Since constant functions trivially depend on only one coordinate, the Stone-Weierstrass theorem implies that the algebra A generated by the functions in $C(\Omega)$ depending on one coordinate is dense in $C(\Omega)$. [The algebra A may be described explicitly as the collection of finite sums of finite products of functions depending on one coordinate.]

Now if $f(\omega) = g(\omega(t))$, $g \in C(\Omega_t)$, then $f = g \circ p_t$, where p_t is the projection of Ω onto Ω_t . Thus every function in A is measurable relative to \mathcal{F} and $\mathcal{B}(R)$. But each function in $C(\Omega)$ is a uniform (hence pointwise) limit of a sequence of functions in A , so that \mathcal{F} makes every function in $C(\Omega)$ measurable. Therefore $\mathcal{A}(\Omega) \subset \mathcal{F}$.

8. (a) Since d is continuous, so is f . If $\omega \neq \omega'$, let $\{\omega_n^*\}$ be a sequence in D with $\omega_n^* \rightarrow \omega$. Eventually $d(\omega, \omega_n^*) \neq d(\omega', \omega_n^*)$; hence f is one-to-one. If $f(\omega^n) \rightarrow f(\omega)$, but $d(\omega^n, \omega) \geq \delta > 0$ for all n , let ω^* be an element of D such that $d(\omega, \omega^*) < \delta/2$. Since $f(\omega^n) \rightarrow f(\omega)$, we have $d(\omega^n, \omega^*) \rightarrow d(\omega, \omega^*)$ as $n \rightarrow \infty$, and therefore $d(\omega^n, \omega) \leq d(\omega^n, \omega^*) + d(\omega^*, \omega) < \delta$ for large n , a contradiction. Thus f has a continuous inverse.

- (b) By part (a), f is a homeomorphism of Ω and $f(\Omega) \subset [0, 1]^\infty$. Since Ω is complete, it is a G_δ in any space in which it is topologically embedded (see the appendix on general topology, Theorem A9.11). Thus $f(\Omega)$ is a G_δ in $[0, 1]^\infty$, in particular a Borel set.
- (c) Let $r \in [0, 1]$ with binary expansion $0.a_1a_2\cdots$ (to avoid ambiguity, do not use expansions that terminate in all ones). Define $g(r) = (a_1, a_2, \dots)$; g is then one-to-one. If $k/2^n$ is a dyadic rational number with binary expansion $0.a_1\cdots a_n 0 0 \cdots$, then

$$g\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \{y \in S^\infty : y_1 = a_1, \dots, y_n = a_n\}.$$

Thus g maps finite disjoint unions of dyadic rational intervals onto measurable cylinders, and since $g[0, 1] = S^\infty - \{y \in S^\infty : y_n \text{ is eventually } 1\} = S^\infty - \text{a countable set}$, we have $g[0, 1] \in \mathcal{S}^\infty$ and g and g^{-1} are measurable.

- (d) Let $A = \{y \in S^\infty : y_n \text{ is eventually } 1\}$, $B = \{y \in S^\infty : y \in g[0, 1] \text{ and } g^{-1}(y) \text{ is rational}\}$, where g is the map of part (c). Let q be a one-to-one correspondence of the rationals in $[0, 1]$ and $A \cup B$. Define $h: [0, 1] \rightarrow S^\infty$ by

$$h(x) = \begin{cases} g(x) & \text{if } x \text{ is irrational,} \\ q(x) & \text{if } x \text{ is rational.} \end{cases}$$

Then h has the desired properties.

- (e) Define $h: \prod_n \Omega_n \rightarrow \prod_n S_n$ by $h(\omega_1, \omega_2, \dots) = (h_1(\omega_1), h_2(\omega_2), \dots)$. The mapping h yields the desired equivalence.

Section 4.5

3. Assume $\mu_n \xrightarrow{w} \mu$, and let A be a bounded Borel set whose boundary has μ -measure 0. Let V be a bounded open set with $V \supset \bar{A}$, and let $G_j = \{x \in V : d(x, \bar{A}) < 1/j\}$. If f_j is a continuous map from Ω to $[0, 1]$ such that $f_j = 1$ on \bar{A} and $f_j = 0$ off G_j [see A5.13(b)], then

$$\begin{aligned} |\mu_n(A) - \mu(A)| &\leq \left| \int_{\Omega} (I_A - f_j) d\mu_n \right| + \left| \int_{\Omega} f_j d\mu_n - \int_{\Omega} f_j d\mu \right| \\ &\quad + \left| \int_{\Omega} (f_j - I_A) d\mu \right|. \end{aligned} \tag{1}$$

Since $f_j - I_A$ is 0 on A and also on G_j^c , the third term on the right side of (1) is bounded by $\mu(G_j - A) \leq \mu(\bar{G}_j - A^0)$. Similarly, the first term is bounded by $\mu_n(\bar{G}_j - A^0)$. The second term approaches 0 as $n \rightarrow \infty$ since the support of f_j is a subset of the compact set \bar{G}_j .

Now let g_{jk} be a continuous function from Ω to $[0, 1]$, with $\text{supp } g_{jk} \subset \bar{V}$, such that as $k \rightarrow \infty$,

$$g_{jk} \downarrow I_{\{\bar{G}_j - A^0\}}.$$

To verify that the g_{jk} exist, note that $\bar{G}_j - A^0 = \bigcap_{n=1}^{\infty} U_n$, where $U_n = \{x \in \bar{V}: d(x, \bar{G}_j - A^0) < 1/n\}$. Let g'_{jn} be a continuous function from Ω to $[0, 1]$ such that $g'_{jn} = 1$ on $\bar{G}_j - A^0$, $g'_{jn} = 0$ off U_n , and set $g_{jk} = \min(g'_{j1}, \dots, g'_{jk})$. Now

$$\mu_n(\bar{G}_j - A^0) = \int_{\Omega} I_{\bar{G}_j - A^0} d\mu_n \leq \int_{\Omega} g_{jk} d\mu_n \rightarrow \int_{\Omega} g_{jk} d\mu \quad \text{as } n \rightarrow \infty.$$

Thus

$$\limsup_{n \rightarrow \infty} \mu_n(\bar{G}_j - A^0) \leq \int_{\Omega} g_{jk} d\mu.$$

Since $\text{supp } g_{jk} \subset \bar{V}$ and $\mu(\bar{V}) < \infty$, we may let $k \rightarrow \infty$ and invoke the monotone convergence theorem to obtain

$$\limsup_{n \rightarrow \infty} \mu_n(\bar{G}_j - A^0) \leq \mu(\bar{G}_j - A^0).$$

It follows from (1) and the accompanying remarks that

$$\limsup_{n \rightarrow \infty} |\mu_n(A) - \mu(A)| \leq 2\mu(\bar{G}_j - A^0).$$

As $j \rightarrow \infty$ we have $\bar{G}_j - A^0 \downarrow \bar{A} - A^0 = \partial A$, and since $\mu(\partial A) = 0$ we conclude that $\mu_n(A) \rightarrow \mu(A)$, proving that (a) implies (b).

Assume $\mu_n(A) \rightarrow \mu(A)$ for all bounded Borel sets with $\mu(\partial A) = 0$, and let f be a bounded function from Ω to R , continuous a.e. [μ] with $\text{supp } f \subset K$, K compact. Let V and W be bounded open sets such that $K \subset V \subset \bar{V} \subset W$ [see A5.12(c)]. Now

$$\bar{V} = \bigcap_{\delta > 0} \{x \in W: d(x, \bar{V}) < \delta\} = \bigcap_{\delta > 0} W_{\delta};$$

the W_{δ} are open and

$$\partial W_{\delta} \subset \{x \in \bar{W}: d(x, \bar{V}) = \delta\}.$$

Thus the W_{δ} have disjoint boundaries and $\mu(\bar{W}) < \infty$, so $\mu(\partial W_{\delta}) = 0$ for some δ . Therefore we may assume without loss of generality that we have $K \subset V$, with V a bounded open set and $\mu(\partial V) = 0$.

Now if $A \subset V$, the interior of A is the same relative to V as to the entire space Ω since V is open. The closure of A relative to V is given by $\bar{A}_V = \bar{A} \cap V$; hence the boundary of A relative to V is $\partial_V A = (\partial A) \cap V$.

If A is a Borel subset of V and $\mu(\partial_V A) = 0$, then

$$\mu[(\partial A) \cap V] = 0;$$

also

$$(\partial A) \cap V^c \subset \bar{A} \cap V^c \subset \bar{V} - V = \bar{V} - V^0 = \partial V;$$

hence $\mu[(\partial A) \cap V^c] = 0$, so that $\mu(\partial A) = 0$. Thus by hypothesis, $\mu_n(A) \rightarrow \mu(A)$. By 4.5.1, if μ'_n and μ' denote the restrictions of μ_n and μ to V , we have $\mu'_n \xrightarrow{w} \mu'$. Since f restricted to V is still bounded and continuous a.e. [μ], we have $\int_V f d\mu'_n \rightarrow \int_V f d\mu'$, that is, $\int_\Omega f d\mu_n \rightarrow \int_\Omega f d\mu$. This proves that (b) implies (c); (c) implies (a) is immediate.

Chapter 5

Section 5.8

4. To prove (a), let $A, B \in \mathcal{C}_{i_1}$, $B \subset A$. If $C_{i_k} \in \mathcal{C}_{i_k}$, $k = 2, \dots, n$, then (using a product notation for intersection)

$$\begin{aligned} P[(A - B)C_{i_2} \cdots C_{i_n}] &= P(AC_{i_2} \cdots C_{i_n}) - P(BC_{i_2} \cdots C_{i_n}) \\ &= [P(A) - P(B)] \prod_{k=2}^n P(C_{i_k}) \\ &= P(A - B)P(C_{i_2}) \cdots P(C_{i_n}). \end{aligned}$$

Thus $A - B$ can be added to C_{i_1} while preserving independence. Since i_1 is arbitrary, (a) follows. The proofs of (b) and (c) are quite similar, and (d) follows from (a), (b), and (c).

Now let $\mathcal{C}_1 = \{A, B\}$, $\mathcal{C}_2 = \{C\}$, where A and C are independent, and B and C are independent. Since $P[(A \cap B) \cap C]$ need not equal $P(A \cap B)P(C)$ [see 5.3.3(b)], $A \cap B$ cannot be added to \mathcal{C}_1 .

Finally, we show that the $\sigma(\mathcal{C}_i)$ are independent iff each \mathcal{C}_i is closed under finite intersection. Fix i , and consider the collection of classes \mathcal{A}_i such that \mathcal{A}_i is closed under finite intersection, $\mathcal{C}_i \subset \mathcal{A}_i$, and \mathcal{A}_i and the \mathcal{C}_j , $j \neq i$, are independent. Partially order the \mathcal{A}_i by inclusion. Each chain has an upper bound (the union of the chain) so there is a maximal

class \mathcal{D}_i . Since \mathcal{D}_i is closed under finite intersection, it is closed under arbitrary differences by (a) ($A - B = A - (A \cap B)$); hence by (a), (b), and (c), \mathcal{D}_i is a σ -field. Thus $\sigma(\mathcal{C}_i) \subset \mathcal{D}_i$, and consequently \mathcal{C}_i can be replaced by $\sigma(\mathcal{C}_i)$ while preserving independence. Since i is arbitrary, the result follows.

Section 5.9

4. $P\{Y \in B\} = P\{X \in g^{-1}(B)\} = \int_{g^{-1}(B)} f_1(x) dx$. Under the transformation $x = h(y)$, $y = g(x)$, this becomes $\int_B f_1(h(y))|J_h(y)|dy$ (see Problem 6, Section 2.3). The result follows.
6. Let $A_j = \{X_j > 2 + T_1\}$, $T_1 = \min_i X_i$. Then $X_0 = \sum_{j=1}^n I_{A_j}$, where I_{A_j} is the indicator of A_j . Therefore $E(X_0) = \sum_{j=1}^n E(I_{A_j}) = \sum_{j=1}^n P(A_j)$, where $E(X_0) = \int_{\Omega} X_0 dp$. By symmetry,

$$\begin{aligned} E(X_0) &= nP\{X_2 > 2 + T_1\} \\ &= n \sum_{\substack{j=1 \\ j \neq 2}}^n P\{T_1 = X_j, X_2 > 2 + X_j\} \\ &= n(n-1)P\{T_1 = X_1, X_2 > 2 + X_1\} \\ &= n(n-1)P\{X_2 > 2 + X_1, X_3 > X_1, \dots, X_n > X_1\} \\ &= n(n-1) \int_{-\infty}^{\infty} f(x_1) dx_1 \int_{2+x_1}^{\infty} f(x_2) dx_2 \int_{x_1}^{\infty} f(x_3) dx_3 \cdots \int_{x_1}^{\infty} f(x_n) dx_n \\ &= n(n-1) \int_{-\infty}^{\infty} f(x_1)[1 - F(x_1)]^{n-2}[1 - F(2+x_1)] dx_1. \end{aligned}$$

7. Define F^{-1} and X as suggested. Then if $0 < y < 1$, $F^{-1}(y) \leq x$ iff $y \leq F(x)$. For if $F^{-1}(y) > x$, we can find $x_0 > x$ such that $F(x_0) < y$, and thus $F(x) \leq F(x_0) < y$; if $F(x) < y$, by right continuity we can find $x_0 > x$ such that $F(x_0) < y$; therefore $F^{-1}(y) \geq x_0 > x$. Now

$$P\{\omega: X(\omega) \leq x\} = P\{\omega: F^{-1}(\omega) \leq x\} = P\{\omega: \omega \leq F(x)\} = F(x).$$

(This also shows that X is measurable, and that

$$X(\omega) = \min\{x: F(x) \geq \omega\}, \quad 0 < \omega < 1.$$

Section 5.10

2. Separate F into discrete and absolutely continuous parts: $F = F_1 + F_2$, where

$$F_1(x) = \begin{cases} 0, & x < 2, \\ \frac{1}{3}, & x \geq 2, \end{cases}$$

$$F_2(x) = \int_{-\infty}^x f_2(t) dt$$

where

$$f_2(x) = \begin{cases} 0, & x < 2 \text{ or } x > 3, \\ \frac{2}{3}, & 2 \leq x \leq 3. \end{cases}$$

Thus

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 dF(x) = \int_{-\infty}^{\infty} x^2 dF_1(x) + \int_{-\infty}^{\infty} x^2 dF_2(x) \\ &= 2^2 \left(\frac{1}{3} \right) + \int_{-\infty}^{\infty} x^2 f_2(x) dx \\ &= \frac{4}{3} + \frac{2}{3} \int_2^3 x^2 dx = \frac{50}{9}. \end{aligned}$$

Section 5.11

1. (a) Assume $Y_i: (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$, $i = 1, 2, \dots$. By hypothesis,
- $$\begin{aligned} P\{(Y_1, \dots, Y_n) \in A, (Y_{n+1}, Y_{n+2}, \dots) \in B\} \\ = P\{(Y_1, \dots, Y_n) \in A\}P\{(Y_{n+1}, Y_{n+2}, \dots) \in B\} \end{aligned}$$

if A is a measurable rectangle in $\prod_{i=1}^n \mathcal{F}_i$ and B is a measurable rectangle in $\prod_{i=n+1}^{\infty} \mathcal{F}_i$, the formula is still valid if A and B are finite disjoint unions of measurable rectangles. Two applications of the monotone class theorem establish this result for all $A \in \prod_{i=1}^n \mathcal{F}_i$ and $B \in \prod_{i=n+1}^{\infty} \mathcal{F}_i$.

- (b) Let \mathcal{C} be the class of sets $B \in \mathcal{S}^{\infty}$ such that

$$P\{(Y_1, Y_2, \dots) \in B\} = P\{(Y_n, Y_{n+1}, \dots) \in B\}.$$

Since the Y_i are independent and P_{Y_i} is the same for all i , all measurable rectangles belong to \mathcal{C} ; hence \mathcal{C} contains all finite disjoint unions of measurable rectangles. But \mathcal{C} is a monotone class; hence $\mathcal{C} = \mathcal{S}^{\infty}$, and the result follows.

Chapter 6

Section 6.3

1. (a) The conditional density of X given Y is $h(x|y) = f(x, y)/f_2(y)$, where $f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$. Thus

$$E(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x)h(x|y) dx,$$

assuming $E[g(X)]$ exists [cf. 6.3.5(c), Eq. (5)].

(b) $E(Y|A) = E(YI_A)/P(A)$ [see 6.3.5(b)]. Now

$$P(A) = \int_{x \in B} \int_{y=-\infty}^{\infty} f(x, y) dx dy$$

and

$$E(YI_A) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yI_A(x, y)f(x, y) dx dy = \int_{x \in B} \int_{y=-\infty}^{\infty} yf(x, y) dx dy.$$

(c) $E(X|A) = E(XI_A)/P(A)$, where

$$P(A) = \iint_{x+y \in B} f(x, y) dx dy$$

and

$$E(XI_A) = \iint_{x+y \in B} xf(x, y) dx dy.$$

3. By 6.3.1, $P\{X \in A, Y \in B\} = \sum_{x \in A} P\{X = x\} \int_B h(y|x) dy$. Thus (take $A = R$) Y has density $f(y) = \sum_x P\{X = x\}h(y|x)$. Now define

$$P\{X \in A | Y = y\} = \sum_{x \in A} P\{X = x | Y = y\},$$

where $P\{X = x | Y = y\}$ is as specified in the problem. Then

$$\begin{aligned} \int_B P\{X \in A | Y = y\} dP_Y(y) &= \int_B P\{X \in A | Y = y\} f(y) dy \\ &= \int_B \sum_{x \in A} P\{X = x\} h(y|x) dy \\ &= \sum_{x \in A} P\{X = x\} \int_B h(y|x) dy \\ &= P\{X \in A, Y \in B\}. \end{aligned}$$

The result follows from 6.3.1.

Section 6.4

1. (a) We show that $\{(X(\omega), Z(\omega)) : \omega \in \Omega\}$ is a function; f may then be defined arbitrarily off $X(\Omega)$. If we do not have a function, then there are points $\omega_1, \omega_2 \in \Omega$ with $X(\omega_1) = X(\omega_2)$ but $Z(\omega_1) \neq Z(\omega_2)$. Let $C_1, C_2 \in \mathcal{F}''$, $C_1 \cap C_2 = \emptyset$, with $Z(\omega_1) \in C_1$, $Z(\omega_2) \in C_2$. Now

$Z^{-1}(C_1) \cap Z^{-1}(C_2) = \emptyset$, and since $C_1, C_2 \in \mathcal{F}''$, we have $Z^{-1}(C_j) = X^{-1}(B_j)$ for some $B_j \in \mathcal{F}', j = 1, 2$. Now $Z(\omega_1) \in C_1$; hence $\omega_1 \in Z^{-1}(C_1) = X^{-1}(B_1)$; therefore $X(\omega_1) \in B_1$. But $X(\omega_1) \notin B_2$, for if so, $Z(\omega_1) \in C_2$ as well as C_1 . Similarly, $X(\omega_2) \in B_2, X(\omega_2) \notin B_1$. But $X(\omega_1) = X(\omega_2)$, and this contradicts the fact that $(B_1 - B_2) \cap (B_2 - B_1)$ is always empty.

- (b) Let $\Omega_0 = X(\Omega)$. If $C \in \mathcal{F}''$, then

$$Z^{-1}(C) = X^{-1}(f^{-1}(C)) = X^{-1}(f^{-1}(C) \cap \Omega_0).$$

But

$$Z^{-1}(C) = X^{-1}(B) = X^{-1}(B \cap \Omega_0) \quad \text{for some } B \in \mathcal{F}'.$$

Since X maps onto Ω_0 , $X[X^{-1}(A)] = A$ for any $A \subset \Omega_0$; hence $f^{-1}(C) \cap \Omega_0 = B \cap \Omega_0 \in \mathcal{F}'$. But if $f(\Omega' - \Omega_0) = \{a\}$, then

$$f^{-1}(C) \cap (\Omega' - \Omega_0) = \begin{cases} \emptyset & \text{if } a \notin C, \\ \Omega' - \Omega_0 & \text{if } a \in C. \end{cases}$$

Therefore $f^{-1}(C) \cap (\Omega' - \Omega_0) \in \mathcal{F}'$; hence $f^{-1}(C) \in \mathcal{F}'$.

Section 6.5

$$\begin{aligned} 2. \quad \int_{\{X \in A, Z \in B\}} Y dP &= E[Y(I_A \circ X)(I_B \circ Z)] \\ &= E[Y(I_A \circ X)]E[I_B \circ Z] \quad \text{by independence} \\ &= E[E(Y(I_A \circ X) | X)]E[I_B \circ Z] \\ &= E[(I_A \circ X)E(Y | X)]E[I_B \circ Z] \\ &= E[(I_A \circ X)E(Y | X)E(I_B \circ Z)] \quad \text{by independence} \\ &= \int_{\{X \in A, Z \in B\}} E(Y | X) dP. \end{aligned}$$

Thus

$$\int_{\{(X, Z) \in C\}} Y dP = \int_{\{(X, Z) \in C\}} E(Y | X) dP \tag{1}$$

for C a measurable rectangle $A \times B, A \in \mathcal{F}', B \in \mathcal{F}''$ (where $X: (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$, $Z: (\Omega, \mathcal{F}) \rightarrow (\Omega'', \mathcal{F}'')$). By the monotone class theorem, (1) holds for all $C \in \mathcal{F}' \times \mathcal{F}''$. [Integrability of Y is used in showing that if (1) holds for C_1, C_2, \dots and $C_n \downarrow C$, then (1) holds for C .]

Section 6.6

1. (a) Let $g(x, y) = I_{A_0}(x)I_{B_0}(y)$ be the indicator of a measurable rectangle $A_0 \times B_0 \in \mathcal{F}' \times \mathcal{F}''$. Then

$$\begin{aligned}\int_{\{X \in A\}} g(X, Y) dP &= P\{X \in A \cap A_0, Y \in B_0\} \\ &= \int_{A \cap A_0} P\{Y \in B_0 | X = x\} dP_x(x) \\ &= \int_A I_{A_0}(x)P_x(B_0) dP_x(x).\end{aligned}$$

Thus

$$\begin{aligned}E[g(X, Y) | X = x] &= I_{A_0}(x)P_x(B_0) \\ &= \int_{\Omega''} I_{A_0}(x)I_{B_0}(y) dP_x(y) \\ &= \int_{\Omega''} g(x, y) dP_x(y) \quad (\text{a.e. } [P_x]).\end{aligned}$$

Thus the result holds for g of this type. We proceed to indicators of arbitrary sets in $\mathcal{F}' \times \mathcal{F}''$ using the monotone class theorem, and then to arbitrary g in the usual way.

- (b) By (a),

$$\begin{aligned}P\{(X, Y) \in C | X = x\} &= E[I_C(X, Y) | X = x] \\ &= \int_{\Omega''} I_C(x, y) dP_x(y) \\ &= P_x(C(x)) \quad (\text{a.e. } [P_x]).\end{aligned}$$

- (c) $P\{X \in \Omega', (X, Y) \in C\} = \int_{\Omega'} P\{(X, Y) \in C | X = x\} dP_X(x)$ by definition of conditional probability. The result follows from (b).
3. (a) Since $\mu(E) > 0$ and $\delta < 1$, there is an open set $V \supset E$ such that $\mu(V) \leq \delta^{-1}\mu(E)$; V is a disjoint union of open intervals I_n . Then

$$\delta \sum_n \mu(I_n) = \delta \mu(V) \leq \mu(E) = \mu(E \cap V) = \sum_n \mu(E \cap I_n). \quad (1)$$

Therefore $\delta \mu(I_n) \leq \mu(E \cap I_n)$ for some n . [Note that

$$\sum_n \mu(I_n) = \mu(V) \leq \delta^{-1}\mu(E) < \infty,$$

so it is not possible to have both sums infinite in (1).]

- (b) By (a) there is an open interval I such that $\mu(E \cap I) \geq \frac{3}{4}\mu(I)$. We show that $(-\frac{1}{2}\mu(I), \frac{1}{2}\mu(I)) \subset D(E)$. Let $|x| < \frac{1}{2}\mu(I)$. If $E \cap I$ and $(E \cap I)$

$+ x$ are disjoint, the measure of their union is $2\mu(E \cap I) \geq \frac{3}{2}\mu(I)$. But $(E \cap I) \cup [(E \cap I) + x] \subset I \cup (I + x)$, an open interval of length less than $\mu(I) + \frac{1}{2}\mu(I) = \frac{3}{2}\mu(I)$, a contradiction. Thus there is an element $y \in (E \cap I) \cup [(E \cap I) + x]$. But then $y \in E$ and $y = z + x$ for some $z \in E$; hence $x \in D(E)$, as desired.

- (c) Since the circle is compact, there is a subsequence converging to a point v on the circle. Given any positive integer N , choose z_n such that $n \geq N$ and $|z_n - v| < \varepsilon/2$; then pick z_{n+k} ($k > 0$) such that $|z_{n+k} - v| < \varepsilon/2$. Then $0 < |z_n - z_{n+k}| < \varepsilon$. (Note that $z_n \neq z_{n+k}$ since $\alpha/2\pi$ is irrational.) Thus $z_n, z_{n+k}, z_{n+2k}, \dots$ form a chain that eventually goes entirely around the circle, with the distance between successive points less than ε . Thus, given N , we can find z_r , $r \geq N$, such that $|z_r - z| < \varepsilon$. The result follows.
- (d) Since $C = \{1 + x: x \in B\}$, it suffices to consider B . But B is dense iff the set of numbers $n\zeta$, n an integer, reduced modulo 2, is dense in $[0, 2)$. Equivalently (consider $\theta \rightarrow e^{i\pi\theta}$, $0 \leq \theta < 2$), $\{e^{inx}: n \text{ an integer}\}$ is dense in the circle if α/π is irrational. But in this case $\alpha/2\pi$ is also irrational, and the result follows from (c).
- (e) Let $F \in \mathcal{B}(R)$, $F \subset E_0$. We claim that $D(F) \cap A \subset \{0\}$. For if $x, y \in F$ and $x - y \in A$, then $x \sim y$; but $x, y \in E_0$; hence $x = y$ by definition of E_0 . Now assume $\mu(F) > 0$. $D(F)$ includes a neighborhood of 0 by (b), so that $(0, a) \subset D(F)$ for some $a > 0$. Since A is dense by (c), we have $(0, a) \cap A \neq \emptyset$, contradicting $D(F) \cap A \subset \{0\}$. Thus $\mu(F)$ must be 0, so that if E_0 is Lebesgue measurable, then $\mu(E_0) = 0$.

Now if $x \in R$, then x is equivalent to some $y \in E_0$; hence $x - y \in A$. Therefore $R = \bigcup\{E_0 + a: a \in A\}$. But if $y + a_1 = z + a_2$, where $y, z \in E_0$, $a_1, a_2 \in A$, then $y - z = a_2 - a_1 \in A$. (Note that A is a group under addition.) Thus $y \sim z$; but since $y, z \in E_0$, $y = z$ and therefore $a_1 = a_2$. Thus the sets $E_0 + a$, $a \in A$, are disjoint.

Finally, assume E_0 Lebesgue measurable. Then $\mu(E_0 + a) = \mu(E_0)$ by translation-invariance of Lebesgue measure. Since A is countable, the preceding paragraph implies that $\mu(R) = 0$, a contradiction.

- (f) If $x \in R$, then $x = y + a$ for some $y \in E_0$, $a \in A$ [see the argument of (e)]. Since $A = B \cup C$, it follows that $R = M \cup M'$. Let F be a Borel subset of M . We claim that $D(F) \cap C \subset \{0\}$. For let $x, y \in F$ with $x - y \in C \subset A$. Then $x \sim y$, and $x = z_1 + b_1$, $y = z_2 + b_2$, where $z_1, z_2 \in E_0$, $b_1, b_2 \in B$. It follows that $z_1 - z_2 = x - y + b_2 - b_1 \in A$; hence $z_1 = z_2$. But then $x - y = b_1 - b_2 \in B \cap C$, so $x = y$. Since C is dense by (d), the same argument as in (e) shows that $\mu(F) = 0$. Finally, since $M' = \{x + 1: x \in M\}$, any Borel subset of M' has Lebesgue measure 0, by translation-invariance.

- (g) The first statement follows from (f). If $E \cap M \subset G \subset E$, then $E - G \subset E - M \subset E \cap M'$, so the second statement follows from (f) also.
4. (a) If $(B_1 \cap H) \cup (B_2 \cap H^c) = (B_1' \cap H) \cup (B_2' \cap H^c)$, then $B_1 \cap H = B_1' \cap H$, $B_2 \cap H^c = B_2' \cap H^c$. If, say, $\mu(B_1 - B_1') > 0$, then $B_1 - B_1'$ is not a subset of H^c since H^c has inner Lebesgue measure 0, so there is an $x \in (B_1 - B_1') \cap H$, contradicting $B_1 \cap H = B_1' \cap H$. Thus $\mu(B_1 - B_1') = 0$; a symmetrical argument shows that $\mu(B_1' - B_1) = 0$.
- (b) If $B \in \mathcal{G}$,

$$P(H \cap B) = \frac{1}{2}\mu(B) = \frac{1}{2}P(B) = \int_B \frac{1}{2} dP.$$

Thus $P(H|\mathcal{G}) = \frac{1}{2}$ a.e. But $P(H|\mathcal{G}) = Q(\cdot, H)$ a.e.; hence $Q(\omega, H) = \frac{1}{2}$ a.e. Similarly $Q(\omega, H^c) = \frac{1}{2}$ a.e.

- (c) If $B, B_1 \in \mathcal{G}$, $P(B \cap B_1) = \int_{B_1} I_B dP$, so $P(B|\mathcal{G}) = I_B$ a.e.
- (d) By (b) and (c), there is a set $N \in \mathcal{F}$ with $P(N) = 0$ such that for $\omega \notin N$, $Q(\omega, H) = Q(\omega, H^c) = \frac{1}{2}$ and $Q(\omega, B) = I_B(\omega)$ for all intervals B with rational endpoints. For any such ω , $Q(\omega, \{\omega\}) = 1$. [Consider a sequence of rational intervals decreasing to $\{\omega\}$ and use the fact that $Q(\omega, \cdot)$ is a probability measure]. But if $\omega \in H$, then $Q(\omega, \{\omega\}) \leq Q(\omega, H) = \frac{1}{2}$, and if $\omega \notin H$, then $Q(\omega, \{\omega\}) \leq Q(\omega, H^c) = \frac{1}{2}$, a contradiction.

Chapter 7

Section 7.1

$$\begin{aligned} 1. \quad (a) \quad & \frac{\sum_{j,k=1}^n P(A_j \cap A_k)}{[\sum_{k=1}^n P(A_k)]^2} \\ &= \frac{\sum_{j \neq k} P(A_j)P(A_k) + \sum_{k=1}^n P(A_k)}{\sum_{j,k=1}^n P(A_j)P(A_k)} \\ &= \frac{\sum_{j,k=1}^n P(A_j)P(A_k) + \sum_{k=1}^n P(A_k) - \sum_{k=1}^n (P(A_k))^2}{\sum_{j,k=1}^n P(A_j)P(A_k)}, \end{aligned}$$

Now

$$0 \leq \frac{\sum_{k=1}^n P(A_k)(1 - P(A_k))}{[\sum_{k=1}^n P(A_k)]^2} \leq \frac{\sum_{k=1}^n P(A_k)}{[\sum_{k=1}^n P(A_k)]^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty;$$

hence the \liminf condition is satisfied.

$$(b) \quad P\left(\left|\sum_{k=1}^n I_k - \sum_{k=1}^n P(A_k)\right| > \varepsilon \sum_{k=1}^n P(A_k)\right) \leq \frac{\text{Var}(\sum_{k=1}^n I_k)}{\varepsilon^2 [\sum_{k=1}^n P(A_k)]^2}.$$

Now

$$\begin{aligned}\text{Var}\left(\sum_{k=1}^n I_k\right) &= E\left[\left(\sum_{k=1}^n I_k\right)^2\right] - \left[\sum_{k=1}^n E(I_k)\right]^2 \\ &= \sum_{j, k=1}^n P(A_j \cap A_k) - \left[\sum_{k=1}^n P(A_k)\right]^2.\end{aligned}$$

The “ \liminf ” hypothesis implies the desired result.

(c) Let

$$\begin{aligned}d_n &= P\left\{\sum_{k=1}^n I_k < \frac{1}{2} \sum_{k=1}^n P(A_k)\right\} \\ &\leq P\left\{\left|\sum_{k=1}^n I_k - \sum_{k=1}^n P(A_k)\right| > \frac{1}{2} \sum_{k=1}^n P(A_k)\right\}.\end{aligned}$$

Then $\liminf_{n \rightarrow \infty} d_n = 0$ by (b). Thus we can find integers $n_1 < n_2 < \dots$ such that $\sum_{j=1}^{\infty} d_{n_j} < \infty$. By the Borel–Cantelli lemma, with probability 1 we have

$$\sum_{k=1}^{n_j} I_k < \frac{1}{2} \sum_{k=1}^{n_j} P(A_k)$$

for only finitely many j , that is,

$$\sum_{k=1}^{n_j} I_k \geq \frac{1}{2} \sum_{k=1}^{n_j} P(A_k)$$

for large enough j .

(d) By (c), $\sum_{k=1}^{\infty} I_k$ diverges a.e.; hence with probability 1, infinitely many A_n occur; thus $P(\limsup_n A_n) = 1$.

Section 7.2

1. Let $S_n = \sum_{k=1}^n X_k$. If S_n/n converges a.e. to a finite limit, then

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} \rightarrow 0 \quad \text{a.e.}$$

By the second Borel–Cantelli lemma, $\sum_{n=1}^{\infty} P\{|X_n| \geq n\} < \infty$. Since all X_n have the same distribution, $\sum_{n=1}^{\infty} P\{|X_1| \geq n\} < \infty$, so by 7.2.4, $E(X_1)$ is finite. But then $S_n/n \rightarrow E(X_1)$ a.e. by 7.2.5.

4. Let $X = (X_1, X_2, \dots)$. Then $g(X) = g(X(T))$ whenever T permutes finitely many coordinates. Thus $\{g(X) \leq a\}$ is symmetric, and hence by the Hewitt–Savage zero–one law has probability 0 or 1. If $c = \inf\{a : P\{g(X) \leq a\} = 1\}$, then $g(X) = c$ a.e.
7. (a) Let x be an r -adic rational with r -adic expansion $i_1 i_2 \cdots i_n 0 0 \cdots$. Then $P\{x < X \leq x + r^{-n}\} = P\{X_1 = i_1, \dots, X_n = i_n\} = r^{-n}$. (Note that $P\{X = x\} = \prod_{k=1}^{\infty} a_k$, with $a_k = r^{-n}$ for all k ; thus $P\{X = x\} = 0$.) It follows that if λ is Lebesgue measure, $P\{X \in I\} = \lambda(I)$ for every r -adic interval $I \subset [0, 1]$, and hence for every interval $I \subset [0, 1]$ by continuity. Thus $P\{X \leq y\} = y$, $0 \leq y \leq 1$, as desired.
- (b) Fix r and i , and let $A_{ri} = \{x \in [0, 1] : \text{the relative frequency of } i \text{ in the first } n \text{ digits of the } r\text{-adic expansion of } x \text{ converges to } 1/r\}$. Then, if Y_k is the indicator of $\{X_k = i\}$,

$$P\{X \in A_{ri}\} = P\left\{\frac{1}{n}(Y_1 + \cdots + Y_n) \rightarrow \frac{1}{r}\right\} = 1$$

by the strong law of large numbers. If

$$A = \bigcap_{r=2}^{\infty} \bigcap_{i=0}^{r-1} A_{ri},$$

it follows that $P_X(A) = P\{X \in A\} = 1$. But P_X is Lebesgue measure by part (a), and the result follows.

- (c) We may write

$$\int_0^1 R_n(x) dx = \int_0^1 R_n(x) dP_X(x) = E[R_n(X)] = E[2X_n - 1] = 0,$$

and similarly,

$$\int_0^1 R_n(x)R_m(x) dx = E[R_n(X)R_m(X)].$$

Since $R_n(X)$ and $R_m(X)$ are independent for $n \neq m$, and $R_n^2(X) \equiv 1$, the result follows.

Section 7.3

3. Each A_{nj} is a countable union of sets $A_{n+1,k}$; hence

$$\begin{aligned} \int_{A_{nj}} X_{n+1} dP &= \sum_k \int_{A_{n+1,k}} X_{n+1} dP \\ &= \sum_k \lambda(A_{n+1,k}) \quad \text{by definition of } X_{n+1} \\ &= \lambda(A_{nj}) \quad \text{since } \lambda \text{ is countably additive} \\ &= \int_{A_{nj}} X_n dP \quad \text{by definition of } X_n. \end{aligned}$$

$$5. \quad E(Y_n) = \int_{\{p_n(x) > 0\}} \frac{q_n(x)}{p_n(x)} p_n(x) dx \leq \int_{R^n} q_n(x) dx = 1.$$

Let $A = \{(X_1, \dots, X_n) \in B\}$, $B \in \mathcal{B}(R^n)$. Then $\int_A Y_{n+1} dP = E(Y_{n+1} I_A)$; hence

$$\int_A Y_{n+1} dP = \int_{\{x \in B, p_{n+1}(x') > 0\}} \frac{q_{n+1}(x')}{p_{n+1}(x')} p_{n+1}(x') dx', \quad (1)$$

where $x' = (x_1, \dots, x_{n+1}) \in R^{n+1}$, $x = (x_1, \dots, x_n)$, the first n coordinates of x' . Now $p_n(x) = \int_{-\infty}^{\infty} p_{n+1}(x') dx_{n+1}$; thus if $x \in B$ and $p_n(x) = 0$, then $p_{n+1}(x') = 0$ except for x_{n+1} in a set of Lebesgue measure 0, so the integration of $q_{n+1}(x')$ with respect to x_{n+1} in (1) will be 0. In other words,

$$\int_{\{x \in B, p_n(x) = 0, p_{n+1}(x') > 0\}} q_{n+1}(x') dx' = 0.$$

Therefore the right side of (1) becomes

$$\begin{aligned} & \int_{\{x \in B, p_n(x) > 0, p_{n+1}(x') > 0\}} q_{n+1}(x') dx' \\ & \leq \int_{\{x \in B, p_n(x) > 0\}} q_{n+1}(x') dx' \\ & = \int_{\{x \in B, p_n(x) > 0\}} q_n(x) dx \\ & = \int_{\{x \in B, p_n(x) > 0\}} \frac{q_n(x)}{p_n(x)} p_n(x) dx \\ & = E(Y_n I_A) = \int_A Y_n dP, \end{aligned}$$

proving the supermartingale property.

$$6. \quad (a) \quad A_n = \sum_{i=0}^{n-1} X_i - \sum_{i=1}^n E(X_i | \mathcal{F}_{i-1})$$

$$Y_n = \sum_{i=0}^n X_i - \sum_{i=1}^n E(X_i | \mathcal{F}_{i-1}),$$

so $Y_n - A_n = X_n$.

$$\begin{aligned} (b) \quad E(Y_n | \mathcal{F}_{n-1}) &= \sum_{i=0}^{n-1} X_i + E(X_n | \mathcal{F}_{n-1}) - \sum_{i=1}^n E[E(X_i | \mathcal{F}_{i-1}) | \mathcal{F}_{n-1}] \\ &= Y_{n-1} + E(X_n | \mathcal{F}_{n-1}) - E(X_n | \mathcal{F}_{n-1}) \\ &= Y_{n-1}. \end{aligned}$$

$$(c) \quad A_{n+1} - A_n = X_n - E(X_{n+1} | \mathcal{F}_n) \geq 0 \quad \text{a.e.}$$

Section 7.4

1. (a) $E(X_{n+1} | X_n = 0) = p_{n+1}(a_{n+1} - a_{n+1}) + (1 - 2p_{n+1})0 = 0,$
 $E(X_{n+1} | X_n = a_n) = a_n, \quad E(X_{n+1} | X_n = -a_n) = -a_n,$

proving the martingale property. Since for all ω , either $X_n(\omega) = 0$ for all n or for some j , $X_n(\omega) = a_j$ for $n \geq j$, X_n converges everywhere.

- (b) $E(|X_2|) = 2p_2 a_2, \quad E(|X_3|) = 2p_2 a_2 + (1 - 2p_2)2p_3 a_3,$
 $E(|X_4|) = 2p_2 a_2 + (1 - 2p_2)2p_3 a_3 + (1 - 2p_2)(1 - 2p_3)2p_4 a_4,$

and so on. Thus

$$\lim_{k \rightarrow \infty} E(|X_k|) \geq \left[\prod_{k=1}^{\infty} (1 - 2p_k) \right] 2 \sum_{k=2}^{\infty} p_k a_k.$$

The infinite product is greater than 0 since $\sum p_k < \infty$; hence $E(|X_k|) \rightarrow \infty$.

4. By definition of the problem, $E(X_{n+1} | X_1, \dots, X_n) = E(X_{n+1} | X_n)$. If before the n th drawing there are r balls in the urn, c of them white,

$$E\left(X_{n+1} \middle| X_n = \frac{c}{r}\right) = \left(\frac{c}{r}\right)\left(\frac{c+1}{r+1}\right) + \left(1 - \frac{c}{r}\right)\frac{c}{r+1} = \frac{c}{r}.$$

Thus $E(X_{n+1} | X_n) = X_n$. Since $|X_n| \leq 1$ for all n , $X_n \rightarrow X_\infty$ a.e. By the dominated convergence theorem, $E(X_\infty) = \lim_n E(X_n) = E(X_1)$.

Section 7.5

1. Since $|f_n|^p \leq 2^{p-1}(|f_n - f|^p + |f|^p)$, $p \geq 1$, and $|f_n|^p \leq |f_n - f|^p + |f|^p$, $p \leq 1$, by 7.5.3 it suffices to show that the $|f_n - f|^p$ are uniformly integrable. Now

$$\int_A |f_n - f|^p d\mu \rightarrow 0 \quad \text{as } \mu(A) \rightarrow 0$$

for any fixed n , and

$$\begin{aligned} \int_A |f_n - f|^p d\mu &\leq \int_\Omega |f_n - f|^p d\mu \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

by the L^p -convergence. It follows that the integrals of $|f_n - f|^p$ are uniformly continuous and uniformly bounded; the result follows from 7.5.3.

Section 7.6

2. (a) If the X_n are uniformly integrable, $E(X_n) \rightarrow E(X_\infty)$ by 7.5.5. Conversely assume $E(X_n) \rightarrow E(X_\infty)$. Then $|X_n - X_\infty| = (X_n \vee X_\infty) - (X_n \wedge X_\infty)$ and $X_n + X_\infty = (X_n \vee X_\infty) + (X_n \wedge X_\infty)$. By hypothesis, $E(X_n + X_\infty) \rightarrow 2E(X_\infty)$, and by the dominated convergence theorem, $E(X_n \wedge X_\infty) \rightarrow E(X_\infty)$. Hence $E(X_n \vee X_\infty) \rightarrow E(X_\infty)$, so $E(|X_n - X_\infty|) \rightarrow E(X_\infty) - E(X_\infty) = 0$. Thus $X_n \rightarrow X_\infty$ in L^1 , and it follows that the X_n are uniformly integrable. (See Problem 1, Section 7.5; in general, L^p convergence of $\{f_n\}$ implies uniform integrability of $\{|f_n|^p\}$.)
- (b) If $A \in \mathcal{F}_n$, $n \leq m$, then $\int_A X_n \geq \int_A X_m$. Let $m \rightarrow \infty$; by Fatou's lemma,

$$\int_A X_\infty = \int_A \lim_m X_m \leq \liminf_m \int_A X_m \leq \int_A X_n.$$

- (c) By Fatou's lemma, $E(X_\infty) = E(\lim_n X_n) \leq \liminf_n E(X_n) = 0$.

Section 7.7

1. $|X_T| = X_T^+ + X_T^- = 2X_T^+ - X_T$; hence $E(|X_T|) \leq 2E(X_T^+) - E(X_T)$ by 7.7.6. But $\{X_1^+, \dots, X_n^+\}$ is a submartingale by 7.3.6(a); hence $E(X_T^+) \leq E(X_n^+)$ by 7.7.6, as desired.
3. (a) Define T as indicated. By 7.7.6, $\{X_T, X_n\}$ is a submartingale; hence

$$\begin{aligned} E(X_n) &\geq E(X_T) = \int_{\{\max X_i \geq \lambda\}} X_T dP + \int_{\{\max X_i < \lambda\}} X_T dP \\ &\geq \lambda P\{\max X_i \geq \lambda\} + \int_{\{\max X_i < \lambda\}} X_n dP \end{aligned}$$

and the result follows.

- (b) By 7.7.6, $\{X_1, X_T\}$ is a supermartingale; hence

$$\begin{aligned} E(X_1) &\geq E(X_T) = \int_{\{\max X_i \geq \lambda\}} X_T dP + \int_{\{\max X_i < \lambda\}} X_T dP \\ &\geq \lambda P\{\max X_i \geq \lambda\} + \int_{\{\max X_i < \lambda\}} X_n dP. \end{aligned}$$

Since $-X_n \leq (-X_n)^+ = X_n^-$, the result follows.

- (c) Since

$$\left\{ \max_{1 \leq i \leq n} X_i \geq \lambda + \frac{1}{k} \right\} \uparrow \left\{ \sup_n X_n > \lambda \right\} \quad \text{as } n, k \rightarrow \infty,$$

the result follows from (a) and (b). Note also that the same inequalities hold with $\{\sup_n X_n > \lambda\}$ replaced by $\{\sup_n X_n \geq \lambda\}$; this follows because

$$\left\{ \sup_n X_n > \lambda - \frac{1}{k} \right\} \downarrow \left\{ \sup_n X_n \geq \lambda \right\}.$$

5. (a) Since $X_n - nm = \sum_{k=1}^n (Y_k - m)$, and $E(Y_k - m) = 0$, $\{X_n - nm\}$ is a martingale. By 7.7.4(a), $E(X_1 - m) = E(X_{T_n} - T_n m)$; hence $E(X_{T_n}) = mE(T_n)$. Since $Y_j \geq 0$, $X_{T_n} \uparrow X_T$ as $n \rightarrow \infty$, so by the monotone convergence theorem, $E(X_T) = mE(T)$.
- (b) We write $X_n = \sum_{j=1}^n Y_j^+ - \sum_{j=1}^n Y_j^- = X_n' - X_n''$. By (a), $E(X_T') = E(Y_1^+)E(T)$, $E(X_T'') = E(Y_1^-)E(T)$. Since $E(T)$ is finite, so are $E(X_T')$ and $E(X_T'')$; hence $E(|X_T|) < \infty$ and

$$E(X_T) = E(X_T') - E(X_T'') = [E(Y_1^+) - E(Y_1^-)]E(T) = mE(T).$$

- (c) To prove (a), observe that if all $Y_j \geq 0$, then

$$\begin{aligned} E(X_T) &= \sum_{n=1}^{\infty} E(X_n I_{\{T=n\}}) \\ &= \sum_{n=1}^{\infty} E(X_n) P\{T=n\} \quad \text{by independence} \\ &= m \sum_{n=1}^{\infty} n P\{T=n\} \\ &= mE(T). \end{aligned}$$

Part (b) is proved just as above.

Section 7.8

3. (a) $\{X_n\}$ converges a.e. to a finite limit; hence

$$P\{|X_{n+1} - X_n| \geq b \text{ for infinitely many } n\} = 0,$$

so, a.e., $X_{n+1} = X_n$ eventually.

- (b) Since $X_n \geq 0$, $X_T I_{\{T \leq n\}} \uparrow X_T$; hence by the monotone convergence theorem,

$$E(X_T) = \lim_{n \rightarrow \infty} \int_{\{T \leq n\}} X_T = \lim_n \int_{\{T \leq n\}} X_n$$

since on $\{T=k\}$, $k \leq n$, we have $X_T = X_k = X_{k+1} = \cdots = X_n$. But

$$\lim_n \int_{\{T \leq n\}} X_n \leq \limsup_n \int_{\Omega} X_n \leq E(X_0)$$

by the supermartingale property.

- (c) T is the time at which the betting stops. In this case, T is also the time of going broke. By (a), T is a.e. finite, and the result follows.
- (d) Realistically, there is a limit on what we can lose. In practice, what we are doing is starting with a capital of $x > 0$, and stopping when we reach $x + 1$, provided we have not been wiped out (reduced to zero) beforehand. The probability of reaching $x + 1$ before 0 is $x/(x + 1) < 1$, and

$$E(X_T) = \frac{x}{x+1} (x+1) + \frac{1}{x+1} (0) = x = E(X_0)$$

[See Ash (1970, Section 6.2) for details.]

- 5. (i) If $\sum_{j=1}^{\infty} I_{A_j} = \infty$ and $\sum_{j=1}^{\infty} q_j < \infty$, then $X_n \rightarrow \infty$.
 - (ii) If $\sum_{j=1}^{\infty} I_{A_j} < \infty$ and $\sum_{j=1}^{\infty} q_j = \infty$, then $X_n \rightarrow -\infty$.
- But $\{X_n\}$ is a martingale by Problem 4; hence by 7.8.4, X_n converges a.e. to a finite limit on $\{\sup X_n < \infty \text{ or } \inf X_n > -\infty\}$. In case (i), $\inf X_n > -\infty$ and in case (ii), $\sup X_n < \infty$, so we have a contradiction unless the sets

$$\{\sum I_{A_j} = \infty, \quad \sum q_j < \infty\} \quad \text{and} \quad \{\sum I_{A_j} < \infty, \quad \sum q_j = \infty\}$$

have probability 0. (Note that $|I_{A_j} - q_j| \leq 2$ so 7.8.4 actually applies.)

Chapter 8

Section 8.1

1. Reproduce the proof of Theorem 8.1.3, with the following changes. Replace R by R^n , $[-T, T]$ by $\{x \in R^n : -T \leq x_i \leq T, i = 1, \dots, n\}$, and the unit circle by $\{z \in C^n : |z_i| = 1 \text{ for all } i\}$. The relation $g(x) = h(\exp i\pi x/T)$ becomes $g(x) = h(\exp i\pi x_1/T, \dots, \exp i\pi x_n/T)$. The form of $h_n(z)$ is now

$$\sum_{k_1, \dots, k_n = -m_n}^{m_n} a_{k_1 \dots k_n} z_1^{k_1} \cdots z_n^{k_n}.$$

The remainder of the proof is just as before.

4. (a) If $|h(u)| = 1$, then $h(u) = e^{iu a}$ for some a ; hence $e^{iux} = \int_R e^{iux} dF(x)$, or $1 = \int_R e^{iux(x-a)} dF(x)$. Take real parts to obtain $\int_R [1 - \cos u(x-a)] dF(x) = 0$. Since the integrand is nonnegative, we have $\cos u(x-a) = 1$ a.e. [P_X]. But $\cos u(x-a) = 1$ iff $x = a + 2\pi n u^{-1}$, n an integer, so X has a lattice distribution. The converse is proved by reversing the argument.

(b) By part (a),

$$\begin{aligned} P\{X = a + 2\pi n u^{-1} \text{ for some integer } n\} \\ = P\{X = b + 2\pi m(\alpha u)^{-1} \text{ for some integer } m\} = 1 \end{aligned}$$

for appropriate real numbers a and b . If X is nondegenerate, the lattices $\{a + 2\pi n u^{-1} : n \text{ an integer}\}$ and $\{b + 2\pi m(\alpha u)^{-1} : m \text{ an integer}\}$ must have at least two points in common, and this implies that $2\pi u^{-1}$ and $2\pi(\alpha u)^{-1}$ are rationally related. Thus α is a rational number, a contradiction.

5. (a) By 8.1.5(e), h has n continuous derivatives on R and $h^{(k)}(0) = i^k E(X^k)$, $k = 0, 1, \dots, n$. Now if $h: I \rightarrow C$, where I is an interval of R containing 0, and h has n continuous derivatives on I , then for $u \in I$,

$$h(u) = \sum_{k=0}^{n-1} \frac{h^{(k)}(0)}{k!} u^k + u^n \int_0^1 h^{(n)}(ut) \frac{(1-t)^{n-1}}{(n-1)!} dt.$$

[This is an exercise in calculus; see Ash (1970, p.172) for details.] Add and subtract

$$\frac{h^{(n)}(0)u^n}{n!} = u^n \int_0^1 h^{(n)}(0) \frac{(1-t)^{n-1}}{(n-1)!} dt$$

from the above equation to obtain

$$h(u) = \sum_{k=0}^n \frac{h^{(k)}(0)}{k!} u^k + R_n(u),$$

where

$$R_n(u) = u^n \int_0^1 [h^{(n)}(ut) - h^{(n)}(0)] \frac{(1-t)^{n-1}}{(n-1)!} dt.$$

Since $R_n(u)/u^n \rightarrow 0$ as $u \rightarrow 0$ by the dominated convergence theorem, the result follows.

- (b) By 8.1.5(e), h has n continuous derivatives, so as in part (a),

$$h(u) - \sum_{k=0}^{n-1} \frac{h^{(k)}(0)}{k!} u^k = u^n \int_0^1 h^{(n)}(ut) \frac{(1-t)^{n-1}}{(n-1)!} dt.$$

Now $h^{(k)}(u) = \int_R (ix)^k e^{iux} dF(x)$ by 8.1.5(e); hence $|h^{(n)}(ut)| \leq E(|X|^n)$; the result follows.

$$\begin{aligned} 7. (a) \quad & \left(\frac{E_r - E_{-r}}{2r} \right)^2 h(0) = \frac{1}{4r^2} (h(2r) - 2h(0) + h(-2r)) \\ & = \int_{-\infty}^{\infty} \left(\frac{e^{irx} - e^{-irx}}{2r} \right)^2 dF(x) \\ & = - \int_{-\infty}^{\infty} \left(\frac{\sin rx}{rx} \right)^2 x^2 dF(x). \end{aligned}$$

(b) By l'Hospital's rule,

$$\begin{aligned} & \lim_{r \rightarrow 0} \frac{1}{4r^2} [h(2r) - 2h(0) + h(-2r)] \\ &= \lim_{r \rightarrow 0} \frac{2h'(2r) - 2h'(-2r)}{8r} \\ &= \lim_{r \rightarrow 0} \frac{h'(2r) - h'(0) + h'(0) - h'(-2r)}{2(2r)} = h''(0). \end{aligned}$$

Thus

$$\begin{aligned} -h''(0) &= \lim_{r \rightarrow 0} \int_{-\infty}^{\infty} \left(\frac{\sin rx}{rx} \right)^2 x^2 dF(x) \\ &\geq \int_{-\infty}^{\infty} \lim_{r \rightarrow 0} \left(\frac{\sin rx}{rx} \right)^2 x^2 dF(x) \quad \text{by Fatou's lemma} \\ &= \int_{-\infty}^{\infty} x^2 dF(x). \end{aligned}$$

(c) Assume the result holds up to the integer n , and assume $h^{(2n+2)}(0)$ exists and is finite. Then $\int_{-\infty}^{\infty} x^{2n} dF(x) < \infty$, so by 8.1.5(e),

$$h^{(2n)}(u) = \int_{-\infty}^{\infty} (ix)^{2n} e^{iux} dF(x)$$

or

$$(-1)^n h^{(2n)}(u) = \int_{-\infty}^{\infty} e^{iux} dG(x),$$

where

$$G(x) = \int_{-\infty}^x t^{2n} dF(t).$$

Since G is a bounded distribution function with characteristic function $(-1)^n h^{(2n)}$, part (b) shows that $\int_{-\infty}^{\infty} x^2 dG(x) < \infty$, that is, $\int_{-\infty}^{\infty} x^{2n+2} dF(x) < \infty$. The result follows by induction.

Section 8.2

4. (a) Choose $a, b \in I$ such that $L = \int_a^b g(u) du \neq 0$. (If this is not possible, the integral of g is 0 on all subintervals of I ; hence on all Borel subsets of I , and therefore $g = 0$ a.e., contradicting $|\exp(iua_n)| = 1$.) We may assume that $\exp(iua_n)$ converges when $u = a$ and $u = b$ (since $\int_a^b g(u) du$ is continuous in a and b). Now

$$ia_n = ia_n \frac{\int_a^b \exp(iua_n) du}{\int_a^b \exp(iua_n) du} = \frac{\exp(ib a_n) - \exp(ia a_n)}{\int_a^b \exp(iua_n) du} \rightarrow \frac{g(b) - g(a)}{L},$$

- (b) This is immediate from (a).

5. (a) Let $F_n(x) = 1$, $x \geq n$; $F_n(x) = 0$, $x < n$ (corresponding to a random variable $X_n \equiv n$). Let $F_0(x) = 0$ for all x . Then $F_n(x) \rightarrow F_0(x)$ for all $x \in R \cup \{-\infty\}$, but $F_n(\infty) \rightarrow F_0(\infty)$, so F_n does not converge weakly to F_0 .
- (b) If n is even, let $F_n(x) = 1$, $x \geq n$; $F_n(x) = 0$, $x < n$ (corresponding to $X_n \equiv n$). If n is odd, let $F_n(x) = 1$, $x \geq -n$; $F_n(x) = 0$, $x < -n$ ($X_n \equiv -n$). Let $F_0(x) \equiv 0$. Then $F_n(a, b] \rightarrow F_0(a, b]$ for all $a, b \in R$, but $F_n(-\infty, \infty] \rightarrow F(-\infty, \infty]$. Furthermore, $\lim_{n \rightarrow \infty} F_n(x)$ does not exist for any $x \in R$.
6. If F_n converges weakly to F_0 , then $\{F_n\}$ is relatively compact, and hence tight by 8.2.4. Thus assume $\{F_n\}$ tight. Given $\varepsilon > 0$, let a and b be finite continuity points of F_0 such that $F_n(R - (a, b)) < \varepsilon$ for all n . Then

$$\begin{aligned}\limsup_{n \rightarrow \infty} F_n(R) &\leq \varepsilon + \limsup_{n \rightarrow \infty} F_n(a, b] \\ &= \varepsilon + F_0(a, b] \\ &\leq \varepsilon + F_0(R).\end{aligned}$$

But $F_n(R) \geq F_n(a, b]$; hence

$$\liminf_{n \rightarrow \infty} F_n(R) \geq F_0(a, b].$$

Since ε is arbitrary and b may be taken arbitrarily large and a arbitrarily small, we have $F_n(R) \rightarrow F_0(R)$. A similar argument shows that $F_n(-\infty, b] \rightarrow F_0(-\infty, b]$ and $F_n(a, \infty] \rightarrow F_0(a, \infty]$ if a and b are finite continuity points of F_0 . Therefore F_n converges weakly to F_0 .

9. (a) $E(X_{n+1} | \mathcal{F}_n) = E(X_n h_{n+1}^{-1}(u) \exp(iu Y_{n+1}) | \mathcal{F}_n)$
- $$\begin{aligned}&= X_n h_{n+1}^{-1}(u) E[\exp(iu Y_{n+1})] \\ &\quad \text{since } X_n \text{ is } \mathcal{F}_n\text{-measurable and the } Y_k \text{ are independent} \\ &= X_n.\end{aligned}$$
- (b) By hypothesis, $\prod_{k=1}^n h_k \rightarrow h_X$ uniformly on bounded intervals. Thus if I is a bounded open interval containing 0 on which $|h_X| \geq \delta > 0$, then $\prod_{k=1}^n h_k$ is bounded away from 0 on I . Thus for any fixed $u \in I$, $\{X_n, \mathcal{F}_n\}$ is a bounded martingale, and hence converges a.e. But

$$X_n(\omega) = \left[\prod_{k=1}^n h_k(u) \right]^{-1} \exp \left[iu \sum_{k=1}^n Y_k(\omega) \right],$$

and the result follows.

- (c) Let C be the set of pairs (u, ω) , $u \in I$, $\omega \in \Omega$, such that

$$\exp[iu \sum_{k=1}^n Y_k(\omega)]$$

fails to converge. By (b), $\{\omega : (u, \omega) \in C\}$ has probability 0 for each $u \in I$, so by Problem 4, Section 2.6, $\{u : (u, \omega) \in C\}$ has Lebesgue measure 0 for almost every ω .

- (d) Convergence a.e. implies convergence in probability since a probability measure is finite, and convergence in probability implies convergence in distribution by Problem 8. By parts (b) and (c), convergence in distribution implies convergence a.e.
10. If h_n is the characteristic function of the measure μ_n , let μ_{nj} be the marginal measure on the j th coordinate space, defined by $\mu_{nj}(B_j) = \mu_n\{x \in R^k : x_j \in B_j\}$. If $h_n \rightarrow h$, with h continuous at 0, we show that $\{\mu_n\}$ is tight. If K is a compact subset of R^k , there are compact subsets K_1, \dots, K_k of R such that $K \subset \prod_{j=1}^k K_j$. Since $\mu_n(R^k - \prod_{j=1}^k K_j) \leq \sum_{j=1}^k \mu_{nj}(R - K_j)$, it suffices to show that $\{\mu_{nj}\}$ is tight for each $j = 1, \dots, k$. But the characteristic function of μ_{nj} is $h_{nj}(u_j) = h_n(0, \dots, 0, u_j, 0, \dots, 0)$. (To see this, note that if $X_n = (X_{n1}, \dots, X_{nk})$, then $E[\exp(iu_j X_{nj})] = E[\exp(i\langle u, X_n \rangle)]$ if $u_r = 0$, $r \neq j$.) Since h_{nj} converges to a limit that is continuous at 0, the proof of 8.2.8 shows that $\{\mu_{nj}\}$ is tight. The remainder of the proof is carried out exactly as in 8.2.8 and 8.2.9.

Section 8.3

3. Let the X_{nk} be uan. Then

$$\begin{aligned} \max_{1 \leq k \leq n} |h_{nk}(u) - 1| &= \max_{1 \leq k \leq n} \left| \int_{-\infty}^{\infty} (e^{iux} - 1) dF_{nk}(x) \right| \\ &\leq \max_{1 \leq k \leq n} \int_{|x| < \varepsilon} |e^{iux} - 1| dF_{nk}(x) \\ &\quad + \max_{1 \leq k \leq n} \int_{|x| \geq \varepsilon} |e^{iux} - 1| dF_{nk}(x). \end{aligned}$$

Now $|e^{iux} - 1| \leq |ux|$; hence

$$\begin{aligned} \max_{1 \leq k \leq n} |h_{nk}(u) - 1| &\leq \max_{1 \leq k \leq n} \int_{|x| < \varepsilon} |ux| dF_{nk}(x) \\ &\quad + \max_{1 \leq k \leq n} \int_{|x| \geq \varepsilon} 2 dF_{nk}(x) \\ &\leq |u| \varepsilon + 2 \max_{1 \leq k \leq n} P\{|X_{nk}| \geq \varepsilon\}. \end{aligned}$$

The second term approaches 0 as $n \rightarrow \infty$ for any $\varepsilon > 0$ by the van hypothesis, and thus

$$\max_{1 \leq k \leq n} |h_{nk}(u) - 1| \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

uniformly for u in a bounded interval.

Conversely, assume

$$\max_{1 \leq k \leq n} |h_{nk}(u) - 1| \rightarrow 0.$$

By the truncation inequality 8.2.7,

$$\begin{aligned} \max_{1 \leq k \leq n} P\{|X_{nk}| \geq \varepsilon\} &= \max_{1 \leq k \leq n} \int_{|x| \geq \varepsilon} dF_{nk}(x) \\ &\leq \max_{1 \leq k \leq n} 7\varepsilon \int_0^{1/\varepsilon} [1 - \operatorname{Re} h_{nk}(v)] dv \\ &\leq 7\varepsilon \int_0^{1/\varepsilon} \max_{1 \leq k \leq n} |1 - h_{nk}(v)| dv \end{aligned}$$

since $|1 - \operatorname{Re} z| = |\operatorname{Re}(1 - z)| \leq |1 - z|$.

The integral approaches 0 as $n \rightarrow \infty$ by the hypothesis and the dominated convergence theorem, and the result follows.

4. (a) $\operatorname{Var} X_k = 1$ for all n , so $c_n = \sqrt{n}$. Now, for a given $\varepsilon > 0$, if n is large enough so that $\varepsilon\sqrt{n} > 1$, then

$$\begin{aligned} P\left\{\left|\frac{X_k}{c_n}\right| \geq \varepsilon\right\} &= P\{|X_k| \geq \varepsilon\sqrt{n}\} \\ &= \begin{cases} 0 & \text{if } k < \varepsilon\sqrt{n}, \\ \frac{1}{k^2} \left(1 - \frac{1}{c}\right) & \text{if } k \geq \varepsilon\sqrt{n}. \end{cases} \end{aligned}$$

Thus

$$\max_{1 \leq k \leq n} P\left\{\left|\frac{X_k}{c_n}\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2 n} \left(1 - \frac{1}{c}\right) \rightarrow 0.$$

- (b) $\frac{1}{c_n^2} \sum_{k=1}^n \int_{|x| \geq \varepsilon c_n} x^2 dF_k(x) = \frac{1}{n} \sum_{k=1}^n E[X_k^2 I_{\{|X_k| \geq \varepsilon\sqrt{n}\}}]$.

Again let n be large enough so that $\varepsilon\sqrt{n} > 1$. We obtain

$$\frac{1}{n} \sum_{k=\varepsilon\sqrt{n}}^n k^2 P\{|X_k| = k\} \sim \frac{1}{n} (n - \varepsilon\sqrt{n}) \left(1 - \frac{1}{c}\right) \rightarrow 1 - \frac{1}{c} > 0.$$

Thus the Lindeberg condition fails for the X_k . Now

$$\begin{aligned}\text{Var } X'_{nk} &= E[X_k^2 I_{\{|X_k| \leq \sqrt{n}\}}] \\ &= E(X_k^2) = 1 \quad \text{if } k \leq \sqrt{n} \\ &= P\{|X_k| = 1\} = \frac{1}{c} \quad \text{if } k > \sqrt{n}.\end{aligned}$$

Thus

$$(c_n')^2 = \sum_{k=1}^n \text{Var } X'_{nk} = [\sqrt{n}] + \frac{1}{c}(n - [\sqrt{n}]) \sim \frac{n}{c}.$$

The Lindeberg sum for the X'_{nk} is

$$\begin{aligned}\frac{1}{(c_n')^2} \sum_{k=1}^n E[(X'_{nk})^2 I_{\{|X'_{nk}| \leq \varepsilon c_n'\}}] &\sim \frac{c}{n} \sum_{k=\varepsilon c_n'}^{\sqrt{n}} k^2 P\{|X_k| = k\} \\ &\sim \frac{c}{n} \left(\sqrt{n} - \varepsilon \sqrt{\frac{n}{c}} \right) \left(1 - \frac{1}{c} \right) \rightarrow 0.\end{aligned}$$

By 8.3.1, $S_n'/c_n' \xrightarrow{d} \text{normal}(0, 1)$.

$$\begin{aligned}(c) \quad P\{S_n \neq S_n'\} &\leq \sum_{k=1}^n P\{X_k \neq X'_{nk}\} \leq \sum_{k=1}^n P\{|X_k| > \sqrt{n}\} \\ &\leq \sum_{k=\sqrt{n}}^n P\{|X_k| = k\} \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty\end{aligned}$$

since

$$P\{|X_k| = k\} = \frac{1}{k^2} \left(1 - \frac{1}{c} \right) \quad \text{and} \quad \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

- (d) If $Y_n \xrightarrow{d} Y$ and $a_n \rightarrow 1$, then $a_n Y_n \xrightarrow{d} Y$; for if h_n is the characteristic function of Y_n and h is the characteristic function of Y , we have $h_n \rightarrow h$ uniformly on bounded intervals; hence $h_n(a_n u) \rightarrow h(u)$. Now $\sqrt{c} S_n'/\sqrt{n} = a_n(S_n'/c_n')$, where $a_n \rightarrow 1$, so that $(\sqrt{c}/\sqrt{n}) S_n' \xrightarrow{d} \text{normal}(0, 1)$ by (b). Also, if $Y_n \xrightarrow{d} Y$ and $P\{Y_n \neq Y_n'\} \rightarrow 0$, then $Y_n' \xrightarrow{d} Y$ because

$$\begin{aligned}P\{Y_n \leq y\} &= P\{Y_n \leq y, Y_n' \leq y\} + P\{Y_n \leq y, Y_n' > y\} \\ &\leq P\{Y_n' \leq y\} + P\{Y_n \neq Y_n'\}.\end{aligned}$$

Thus by (c), $\sqrt{c} S_n/\sqrt{n} \xrightarrow{d} \text{normal}(0, 1)$. But by (a) and (b), $S_n/c_n = S_n/\sqrt{n} \xrightarrow{d} \text{normal}(0, 1)$.

Section 8.4

1. (a) $E[\exp(iuk(\operatorname{sgn} X_1)|X_1|^{-r})]$

$$\begin{aligned}&= \int_{-n}^n \frac{1}{2n} \exp(iuk(\operatorname{sgn} x)|x|^{-r}) dx \\&= \frac{1}{n} \int_0^n \cos(ukx^{-r}) dx \\&= 1 - \frac{1}{n} \int_0^n [1 - \cos(ukx^{-r})] dx \\&= 1 - \frac{1}{n} \left[\int_0^\infty [1 - \cos(ukx^{-r})] dx - g(n) \right],\end{aligned}$$

where $g(n) = \int_n^\infty [1 - \cos(ukx^{-r})] dx \rightarrow 0$ as $n \rightarrow \infty$ since

$$1 - \cos(ukx^{-r}) = 2 \sin^2 \frac{1}{2} ukx^{-r} \sim cx^{-2r}, \quad 2r > 1.$$

The result follows from Theorem 8.1.2.

- (b) Let $I = \int_0^\infty [1 - \cos(kux^{-r})] dx$; then [see Eq. (3) of the proof of 8.3.1]

$$n \ln(1 - \frac{1}{n} [I - g(n)]) = g(n) - I + \frac{\theta}{n} |I - g(n)|^2 \rightarrow -I;$$

hence $h_n(u) \rightarrow e^{-I}$.

- (c) We have

$$dy = -|u|^{1/r} k^{1/r} x^{-2} dx = -|u|^{-1/r} k^{-1/r} y^2 dx,$$

hence

$$h(u) = \exp(-|u|^{1/r} k^{1/r} \int_0^\infty (1 - \cos y^r) y^{-2} dy).$$

This is of the form $\exp[-d|u|^\alpha]$, $d > 0$, $0 < \alpha < 2$.

Section 8.5

1. A logarithm of $h(u)$ is given by

$$iu + \operatorname{Log}(1 - q) - \operatorname{Log}(1 - qe^{iu}) = iu + \sum_{k=1}^{\infty} \frac{q^k}{k} (e^{iuk} - 1).$$

Thus

$$h(u) = \lim_{n \rightarrow \infty} \exp \left(\sum_{k=1}^n \left[\frac{iu}{n} + \frac{q^k}{k} (e^{iuk} - 1) \right] \right).$$

But

$$\exp\left[\frac{iu}{n} + \frac{q^k}{k}(e^{iuk} - 1)\right]$$

is of the Poisson type [see 8.5.3(b)] with $\lambda = q^k/k$, $a = k$, $b = 1/n$, and the result follows from 8.5.7.

$$\begin{aligned} 4. \quad F_3(z) &= P_{XY}\{(x, y) \in R^2 : x + y \leq z\} \\ &= \iint_{x+y \leq z} dP_{XY}(x, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{\{x+y \leq z\}} dP_X(x) dP_Y(y) \\ &= \int_{-\infty}^{\infty} P_X\{x : x \leq z - y\} dP_Y(y) \\ &= \int_{-\infty}^{\infty} F_1(z - y) dF_2(y) \\ &= \int_{-\infty}^{\infty} F_2(z - x) dF_1(x) \quad \text{by a symmetrical argument.} \end{aligned}$$

In the case where X has a density,

$$F_3(z) = \int_{-\infty}^{\infty} F_1(z - y) dF_2(y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-y} f_1(x) dx \right) dF_2(y).$$

Let $x = u - y$ to obtain

$$F_3(z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^z f_1(u - y) du \right) dF_2(y) = \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f_1(u - y) dF_2(y) \right) du,$$

and the result follows.

Section 8.6

- By the strong law of large numbers, $F_n(x, \omega)$ converges a.e. to

$$E[I_{\{X_1 \leq x\}}] = F(x), \quad \text{and} \quad F_n(x^-, \omega) = \frac{1}{n} \sum_{k=1}^n I_{\{X_k < x\}}(\omega) \rightarrow F(x^-) \quad \text{a.e.}$$

(This holds for each fixed $x \in R$, and the exceptional set of measure 0 depends on x .)

Assume that $F_n(x, \omega) \rightarrow F(x)$ and $F_n(x^-, \omega) \rightarrow F(x^-)$ for $\omega \notin A_x$, where $P(A_x) = 0$. Let S be a countable dense subset of R containing all discontinuity points of F . If $A = \bigcup \{A_x : x \in S\}$, then $P(A) = 0$ since S is countable. If $\omega \notin A$, then $F_n(x, \omega) \rightarrow F(x)$ and $F_n(x^-, \omega) \rightarrow F(x^-)$ for all $x \in S$. Furthermore, $F_n(\infty, \omega) \equiv 1$, $F(\infty) = 1$, $F_n(-\infty, \omega) \equiv 0$, $F(-\infty) = 0$. By 8.6.1, $F_n(x, \omega) \rightarrow F(x)$ uniformly for $x \in R$.

Subject Index

A

Absolute continuity
 of functions, 70, 210
 of measures, 59, 63, 210
Absolute G_δ , 401
Absolute homogeneity, 114
Absorption of one set by another,
 167
Accumulation point
 of filterbase, 374
 of net, 372
Additivity theorem for integrals, 45
Adjoint of linear operator, 149
Algebra of sets, 4
Almost everywhere, 46
Annihilator of subset of normed linear
 space, 149
Approximation
 of Baire or Borel sets by closed, compact,
 or open sets, 34, 179–183
 by continuous functions, 88, 185–188
 by simple functions, 38, 88, 90
Arzela–Ascoli theorem, 396
Atom, 259

B

Baire category theorem, 398
Baire sets, 178
Baire space, 399
Banach space, 114
Banach–Alaoglu theorem, 162
Bernoulli trials, 205, 234
 generalized, 206
Bessel's inequality, 119
Beta function, 237
Bilinear form, 150
Bochner's theorem, 326
Borel–Cantelli lemma, 66
 extension, 316
 second, 272
Borel equivalence, 195, 265
Borel measurable functions, 35, 36
 complex-valued, 80
 properties of, 38–40
Borel sets, 7, 27, 178
Bornivore, 167
Bornological space, 167
Bounded linear operators, 128
 weakly, 150

Bounded set, in topological vector space, 167
 Bounded variation, 71
 Branching process, 285, 314

C

c, space of convergent sequences of complex numbers, 115, 136
 Cantor function, 77, 78
 Cantor sets, 33, 34
 Carathéodory extension theorem, 19
 Cardinality arguments, 13, 34, 42
 Cauchy density, 211
 Cauchy in measure, 93
 Cauchy-Schwarz inequality, 82, 116 for sums, 87
 Central limit theorem, 321ff., 336, 356–358
 Chain rule, 69
 Change of variable formula for multiple integrals, 78
 Characteristic function(s), 322 inversion formula for, 324, 359 moment-generating property, 328 properties of, 325
 Chebyshev's inequality, 84, 227
 Circled set, 151
 Closed graph theorem, 148, 166
 Closed linear operator, 147 unbounded, 150
 Cluster point, 371
 Compact topological space, 381 countably, 384 locally, 385 relatively, 385 sequentially, 385 σ , 386
 Compactification, one-point, 388
 Compactness of families of measures, 330
 Complete metric space, 398
 Complete orthonormal set, 122
 Completeness of L^p spaces, 85, 90
 Completion of measure space, 18
 Complex measure, 69
 Composition of measurable functions, 39
 Conditional density, 242
 Conditional expectation, 244ff. general definition, 245, 252 given σ -field, 249ff., 252 properties of, 254ff.
 Conditional probability, 207, 236ff. general definition, 240, 253 given σ -field, 253 regular, 264
 Conjugate isometry, 131
 Conjugate linear map, 131
 Conjugate space, 142 second, 142
 Consistent probability measures, 190–191
 Continuity of countably additive set functions, 10, 11
 Continuity point of distribution function, 198
 Continuous functions dense in L^p , 88, 185, 188
 Continuous linear functionals, 130, 135 extension of, 140, 156 representations of, 130–133, 136, 137, 184–186, 188 space of, 131, 141
 Convergence in distribution, 322 of filterbases, 373 of net, 371 in normed linear spaces strong, 144 weak, 144 of sequences of linear operators, 134 strong, 134, 144, 149 uniform, 134 of sequences of measurable functions, 92ff. almost everywhere, 47 almost uniform, 93 in L^p , 88 in L^∞ , 89 in measure, 92 in probability, 92 of types theorem, 342, 364
 Convergence theorems for integrals, 44, 47, 49
 Convex functions, 286
 Convex sets, 119 in topological vector spaces, 154ff.
 Convolution, 355–356
 Correlation coefficient, 229
 Countably additive set function, 6, 43, 62 expressed as difference of measures, 11, 44, 61
 Counting measure, 7

Covariance, 229
 Cylinder, 108, 189

D

Daniell integral, 170ff., 175
 Daniell representation theorem, 175
 Decreasing sequence of sets, 1
 De Morgan laws, 1
 Density, 66
 function, of random variable or random vector, 210
 examples, 211
 Derivative
 of function of bounded variation, 76
 Radon–Nikodym, 66
 of signed measure, 74
 Difference operator, 27
 Differentiation
 of measures, 74ff.
 under integral sign, 52
 Dini's theorem, 181
 Directed set, 371
 Discontinuous linear functional, 135
 Discrete distribution function, 76
 Discrete probability space, 203
 Distribution function(s), 5, 23, 29
 bounded sequences of, 329
 decomposition of, 76
 empirical, 358
 proper, 222
 of random variable or random vector, 209, 212
 Distribution of random object, 220
 Dominated convergence theorem, 49
 for conditional expectations, 257
 extension of, 96, 295
 Dynkin system, 168, 169

E

Egoroff's theorem, 94
 Equicontinuity, 396
 Event(s), 201
 mutually exclusive and exhaustive, 207
 prior to stopping time, 303
 Expectation, 223ff.
 conditional, *see* Conditional expectation

Exponential density, 211
 Extension of finitely additive set functions, 149
 Extension theorems for measures, 13ff., 18, 19, 22, 183, 184

F

F_σ set, 42, 178
 Fatou's lemma, 48
 for conditional expectations, 257
 extension of, 295
 Feller's theorem, 342, 361
 Field of sets, 4
 Filter, 373
 Filterbase, 373
 subordinate, 374
 Finitely additive set function, 6
 not countably additive, 11, 12
 σ -finite, 9
 First countable space, 372, 380
 Fubini's differentiation theorem, 80
 Fubini's theorem, 101, 104
 classical, 103, 106
 Functional analysis, 113ff.
 basic theorems of, 138ff.

G

G_δ set, 178
 Gambler's ruin problem, 235
 Gambling model, 315
 Gamma distribution, 351
 Gamma function, 238
 Gauge space, 394, 405
 Glivenko–Cantelli theorem, 358
 Good sets principle, 5
 Gram–Schmidt process, 125
 Gramian, 125

H

Hahn–Banach theorem, 139, 140, 149
 Hausdorff space, 379
 Heine–Borel theorem, 381
 Helly's theorem, 329
 Hermite polynomials, 125

Hewitt–Savage zero–one law, 279, 280

Hilbert spaces, 114, 116ff.

classification of, 123, 124

separable, 124, 133

Hitting time, 303

Hölder inequality, 82

for sums, 87

I

Identically distributed random variables, 275

Identification topology, 377

Increasing sequence of set, 1

Independence

of classes, 216

of events, 204

of random objects, 213, 214

Independent, identically distributed (iid) random variables, 275

Indicator, 35

Infinitely divisible distributions, 348–355

Initial distribution, 231

Inner product, 114

Inner product space, 114

Integrable function, 37, 81

Integral, 36ff.

as countably additive set function, 43

indefinite, 59, 73

Integration of series, 46, 52, 256

Internal point, 154

Inversion formula for characteristic

functions, 324, 359

Isometric isomorphism, 123, 133, 137, 142, 163, 186, 189

J

Jensen's inequality, 287

Joint distribution function, 212

Jordan–Hahn decomposition theorem, 60

K

Kolmogorov extension theorem, 191, 194

Kolmogorov strong law of large numbers, 274

Kolmogorov three series theorem, 313

Kolmogorov zero–one law, 278

Kolmogorov's inequality, 271, 308

Kronecker lemma, 270

L

L^p spaces, 80ff.

completeness of, 85

continuous linear functionals on,

131–133, 137, 165

l^p , $l^p(\Omega)$, 87

L^∞ , 89

l^∞ , $l^\infty(\Omega)$, 90

Lattice distribution, 327

Lattice operations, 170

Law of large numbers

strong, 234, 269ff., 274, 275, 310

weak, 233

Lebesgue decomposition theorem, 68, 76

Lebesgue integrable function, 51

Lebesgue integral

abstract, 36ff.

comparison with Riemann integral, 53

Lebesgue measurable function, 39

Lebesgue measurable sets, 26, 31, 33, 54

Lebesgue measure, 26, 31, 100, 106

Lebesgue set, 78

Lebesgue–Stieltjes measures, 23, 27

Legendre polynomials, 125

Lévy–Khintchine representation, 355

Lévy's theorem, 332–333

Liapounov's condition, 337

Lim inf (lower limit), 2

Lim sup (upper limit), 2

Limit

under integral sign, 52

of sequence of sets, 3, 12, 52

Lindeberg's theorem, 336

Lindelöf space, 370

Line of support theorem, 286

Linear functionals, 130

continuous, 130, *see also* Continuous linear functionals

positive, 170

Linear manifold, 119

generated by set, 121

Linear operator(s), 127ff.

bounded, 128

- Linear operator(s) (*Cont.*)
 closed, 147
 continuous, 128
 with discontinuous inverse, 147
 idempotent, 130
 range and null space of, 149
 spaces of, 133
- Lipschitz condition, 78
- Locally compact space, 385
- Locally convex topological vector space,
 153
 characterization of, 156
- Lusin's theorem, 187
- M**
- Markov chain, 231, 232, 316
 irreducible, 317
- Martingale(s), 281
 applications of, 309ff.
 convergence theorems for, 290ff.
 inequalities, 308
 L^1 bounded, 293
 last element of, 299
 reverse, 284
 uniformly integrable, 298–302
- Martingale differences, 311
- Measurable cylinder, 108, 189
- Measurable function, 35
 Borel, 35
 jointly, 101, 107
 Lebesgue, 39
- Measurable rectangle, 97, 108, 189
- Measurable set, 35
- Measurable space, 35
- Measure(s), 6
 absolutely continuous, 59, 63
 complete, 18
 complex, 69
 concentrated on set, 26
 extension of, 13ff., 183–184
 on field, 6
 finite, 9
 on infinite product spaces, 108ff., 189ff.
 Lebesgue, 26, 31, 100, 106
 Lebesgue–Stieltjes, 23, 27
 outer, 16, 22
 probability, 6
 product, 97, 100, 104, 106, 109, 111
 regular, 183, 189
- σ -finite, 9
 signed, 62
 singular, 59, 66
 spaces of, 186, 189
 on topological spaces, 178ff.
 uniformly σ -finite, 97
- Measure-preserving transformation, 50
- Measure space, 6
- Minkowski functional, 154
- Minkowski inequality, 83
 for sums, 87
- Moments, 226
- Monotone class theorem, 19
- Monotone convergence theorem, 44
 for conditional expectations, 256
 extended, 47
- Monotone set function, 16
- N**
- Negative part
 of countably additive set function, 62
 of function, 37
- Neighborhood, 369
- Net, 371
- Norm(s), 84, 114
 on finite-dimensional space, 134
 inducing some topology, 134–136
 of linear operator, 128, 133
- Normal distribution, 211, 217, 227, 336
- Normal topological space, 178, 379
- Normed linear space, 114
 linear operators on, 127ff.
- Normed sums, 345
- O**
- Open mapping theorem, 147, 159
- Optional sampling theorems, 304–307
- Optional skipping theorem, 290
- Optional stopping theorem, 311
- Orthogonal complement, 121
- Orthogonal direct sum, 121
- Orthogonal elements, 118
- Orthogonal set, 118
- Orthonormal basis, 122
- Orthonormal set, 118
 complete, 122
- Outer measure, 16, 22
- Overneighborhood, 369

P

- Parallelogram law, 118
 Parseval relation, 122
 Poisson distribution, 348
 Poisson type, 351
 Polarization identity, 124
 Polish space, 180
 Polya urn scheme, 294
 Positive homogeneity, 138
 Positive linear functional, 170
 Positive part
 of countably additive set function, 62
 of function, 37
 Pre-Hilbert space, 114
 Probability function, 210
 Probability measure, 6
 induced by random object, 209, 212, 220,
 225, 240
 Probability space, 6, 202
 universal, 222
 Product measure theorem, 97, 104
 classical, 100, 106, 111
 infinite-dimensional, 109, 111
 Product σ -field, 97, 108, 189
 Product topology, 376
 Projection, 119, 120, 130, 148
 Projection theorem, 121
 Prokhorov's theorem, 330
 Pseudometric, 84, 394
 Pseudonorm, 84
 Pythagorean relation, 118

Q

- Queueing, 318
 Quotient space, 135
 Quotient topology, 210

R

- Rademacher functions, 281
 Radial, 154
 Radial kernel, 154
 Radon-Nikodym derivative, 66
 Radon-Nikodym theorem, 63
 Random objects, 214
 infinite sequences of, 231
 Random sequence, 233

Random signs problem, 273

- Random variable(s), 208
 absolutely continuous, 210
 continuous, 210
 degenerate, 342
 discrete, 209
 distribution function of, 209
 extended, 209
 infinite sequences of, 230
 simple, 209
 Random vector, 212
 absolutely continuous, 213
 discrete, 212
 distribution function of, 212
 Random walk, 235
 Rectangle, 96, 97, 108, 189
 Reflexivity, 142, 163, 164
 Regular conditional distribution function,
 263
 Regular conditional probability, 264
 Regular measure, 183, 189
 Regular topological space, 379
 completely, 380
 Relatively compact family of measures, 330
 Riemann integral, 53–57
 Riemann-Stieltjes integral, 56
 Riesz lemma, 136
 Riesz representation theorem, 130, 133,
 181, 182, 184–186, 188

S

- Sample space, 201
 Second countable space, 380
 Section of set, 98
 Semicontinuous functions, 388
 Semimetric, 84
 Seminorm(s), 84, 113
 family of, generating locally convex
 topology, 153
 Separable Hilbert spaces, 124, 133
 Separable topological spaces, 380
 Separation properties for topological
 spaces, 379
 Separation theorems, 159–161
 strong, 160
 Series of independent random variables,
 310–314, 335

- Set function, 3
 countably additive, 6
 finite, 9
 finitely additive, 6
 Shift operator, 129
 one-sided (unilateral), 129
 two-sided (bilateral), 129
 σ -field (σ -algebra), 4
 countably generated, 111, 148
 induced by random object, 250
 minimal, 5, 12
 Simple functions, 36
 dense in L^p , 88, 90
 Singular distribution function, 77, 78
 Singular measures, 66
 Solvability theorem, 150
 Space spanned by set, 121
 Stable distributions, 344–348
 Standard deviation, 226
 State space, 231
 Steinhaus' lemma, 42
 Stochastic matrix, 231
 Stone–Weierstrass theorem, 393
 Stone's theorem, 200
 Stopping time, 302
 event prior to, 303
 Strong convergence
 in normed linear space, 144
 of operators, 134, 144, 149
 Strong law of large numbers, 234, 269ff.,
 274, 275, 310
 Strong topology, 144, 161
 Subadditivity, 114, 138
 countable, 16
 Sublinear functional, 138
 Submartingales, 282, *see also* Martingale(s)
 Subnet, 372
 Subspace, 119
 closed, 119
 Supermartingales, 282, *see also*
 Martingale(s)
 Symmetric events, 279
 Symmetric functions, 280
 Symmetrization, 313
- U
- Ultrafilter, 374
 Uniform asymptotic negligibility (uan),
 341, 344, 350
 Uniform boundedness principle, 143
 Uniform convergence of operators, 134
 Uniform density, 211, 216
 Uniform integrability, 295
 Martingale theory and, 298
 Uniform space, 405
 Uniform structure, 402
 Uniformity, 402
 Upcrossing theorem, 291, 308
 Urysohn metrization theorem, 387
 Urysohn's lemma, 379

T

- T_i spaces, 379, 380
 Tail events, 278

V

- Vague (= weak) convergence of measures,
 198

- Variance, 226
 of sum, 229
- Variation
 bounded, 71
 of function, 71
 lower, 62
 total, 62, 69
 upper, 62
- Vitali–Hahn–Saks theorem, 43
- Weak compactness, 162–164
 of families of measures, 329
- Weak* compactness, 162–164
- Weak convergence, 144, 161
 of distribution functions, 199
 of measures, 196–199
- Weak* convergence and weak* topology, 161
- Weak law of large numbers, 233
- Weak topology, 144, 161

W

Wald's theorem, 308

Z

Zero–one laws, 278–280

Probability and Mathematical Statistics

A Series of Monographs and Textbooks

Editors **Z. W. Birnbaum** **E. Lukacs**
University of Washington Bowling Green State University
Seattle, Washington Bowling Green, Ohio

1. Thomas Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. 1967
2. Howard Tucker. *A Graduate Course in Probability*. 1967
3. K. R. Parthasarathy. *Probability Measures on Metric Spaces*. 1967
4. P. Révész. *The Laws of Large Numbers*. 1968
5. H. P. McKean, Jr. *Stochastic Integrals*. 1969
6. B. V. Gnedenko, Yu. K. Belyayev, and A. D. Solov'yev. *Mathematical Methods of Reliability Theory*. 1969
7. Demetrios A. Kappos. *Probability Algebras and Stochastic Spaces*. 1969
8. Ivan N. Pesin. *Classical and Modern Integration Theories*. 1970
9. S. Vajda. *Probabilistic Programming*. 1972
10. Sheldon M. Ross. *Introduction to Probability Models*. 1972
11. Robert B. Ash. *Real Analysis and Probability*. 1972
12. V. V. Fedorov. *Theory of Optimal Experiments*. 1972
13. K. V. Mardia. *Statistics of Directional Data*. 1972
14. H. Dym and H. P. McKean. *Fourier Series and Integrals*. 1972
15. Tatsuo Kawata. *Fourier Analysis in Probability Theory*. 1972
16. Fritz Oberhettinger. *Fourier Transforms of Distributions and Their Inverses: A Collection of Tables*. 1973
17. Paul Erdős and Joel Spencer. *Probabilistic Methods in Combinatorics*. 1973
18. K. Sarkadi and I. Vincze. *Mathematical Methods of Statistical Quality Control*. 1973
19. Michael R. Anderberg. *Cluster Analysis for Applications*. 1973
20. W. Hengartner and R. Theodorescu. *Concentration Functions*. 1973
21. Kai Lai Chung. *A Course in Probability Theory, Second Edition*. 1974
22. L. H. Koopmans. *The Spectral Analysis of Time Series*. 1974
23. L. E. Maistrov. *Probability Theory: A Historical Sketch*. 1974
24. William F. Stout. *Almost Sure Convergence*. 1974

25. E. J. McShane. Stochastic Calculus and Stochastic Models. 1974
26. Z. Govindarajulu. Sequential Statistical Procedures. 1975
27. Robert B. Ash and Melvin F. Gardner. Topics in Stochastic Processes. 1975
28. Avner Friedman, Stochastic Differential Equations and Applications, Volume 1, 1975; Volume 2, 1976
29. Roger Cuppens. Decomposition of Multivariate Probabilities. 1975
30. Eugene Lukacs. Stochastic Convergence, Second Edition. 1975
31. H. Dym and H. P. McKean. Gaussian Processes, Function Theory, and the Inverse Spectral Problem. 1976

In Preparation

N. C. Giri. Multivariate Statistical Inference

D 9
E 0
F 1
G 2
H 3
I 4
J 5