Contents

Preface vii
Summary of Notation ix

1 Fundamentals of Measure and Integration Theory 1
  1.1 Introduction 1
  1.2 Fields, σ-Fields, and Measures 3
  1.3 Extension of Measures 12
  1.4 Lebesgue–Stieltjes Measures and Distribution Functions 22
  1.5 Measurable Functions and Integration 35
  1.6 Basic Integration Theorems 45
  1.7 Comparison of Lebesgue and Riemann Integrals 55

2 Further Results in Measure and Integration Theory 60
  2.1 Introduction 60
  2.2 Radon–Nikodym Theorem and Related Results 64
  2.3 Applications to Real Analysis 72
  2.4 $L^p$ Spaces 83
  2.5 Convergence of Sequences of Measurable Functions 96
  2.6 Product Measures and Fubini’s Theorem 101
  2.7 Measures on Infinite Product Spaces 113
  2.8 Weak Convergence of Measures 121
  2.9 References 125

3 Introduction to Functional Analysis 127
  3.1 Introduction 127
  3.2 Basic Properties of Hilbert Spaces 130
  3.3 Linear Operators on Normed Linear Spaces 141
  3.4 Basic Theorems of Functional Analysis 152
  3.5 References 165
4 Basic Concepts of Probability

4.1 Introduction 166
4.2 Discrete Probability Spaces 167
4.3 Independence 167
4.4 Bernoulli Trials 170
4.5 Conditional Probability 171
4.6 Random Variables 173
4.7 Random Vectors 176
4.8 Independent Random Variables 178
4.9 Some Examples from Basic Probability 181
4.10 Expectation 188
4.11 Infinite Sequences of Random Variables 196
4.12 References 200

5 Conditional Probability and Expectation

5.1 Introduction 201
5.2 Applications 202
5.3 The General Concept of Conditional Probability and Expectation 204
5.4 Conditional Expectation Given a $\sigma$-Field 215
5.5 Properties of Conditional Expectation 220
5.6 Regular Conditional Probabilities 228

6 Strong Laws of Large Numbers and Martingale Theory

6.1 Introduction 235
6.2 Convergence Theorems 239
6.3 Martingales 248
6.4 Martingale Convergence Theorems 257
6.5 Uniform Integrability 262
6.6 Uniform Integrability and Martingale Theory 266
6.7 Optional Sampling Theorems 270
6.8 Applications of Martingale Theory 277
6.9 Applications to Markov Chains 285
6.10 References 288

7 The Central Limit Theorem

7.1 Introduction 290
7.2 The Fundamental Weak Compactness Theorem 300
7.3 Convergence to a Normal Distribution 307
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4 Stable Distributions</td>
<td>317</td>
</tr>
<tr>
<td>7.5 Infinitely Divisible Distributions</td>
<td>320</td>
</tr>
<tr>
<td>7.6 Uniform Convergence in the Central Limit Theorem</td>
<td>329</td>
</tr>
<tr>
<td>7.7 The Skorokhod Construction and Other Convergence Theorems</td>
<td>332</td>
</tr>
<tr>
<td>7.8 The k-Dimensional Central Limit Theorem</td>
<td>336</td>
</tr>
<tr>
<td>7.9 References</td>
<td>344</td>
</tr>
<tr>
<td>8 Ergodic Theory</td>
<td>345</td>
</tr>
<tr>
<td>8.1 Introduction</td>
<td>345</td>
</tr>
<tr>
<td>8.2 Ergodicity and Mixing</td>
<td>350</td>
</tr>
<tr>
<td>8.3 The Pointwise Ergodic Theorem</td>
<td>356</td>
</tr>
<tr>
<td>8.4 Applications to Markov Chains</td>
<td>368</td>
</tr>
<tr>
<td>8.5 The Shannon–McMillan Theorem</td>
<td>374</td>
</tr>
<tr>
<td>8.6 Entropy of a Transformation</td>
<td>386</td>
</tr>
<tr>
<td>8.7 Bernoulli Shifts</td>
<td>394</td>
</tr>
<tr>
<td>8.8 References</td>
<td>397</td>
</tr>
<tr>
<td>9 Brownian Motion and Stochastic Integrals</td>
<td>399</td>
</tr>
<tr>
<td>9.1 Stochastic Processes</td>
<td>399</td>
</tr>
<tr>
<td>9.2 Brownian Motion</td>
<td>401</td>
</tr>
<tr>
<td>9.3 Nowhere Differentiability and Quadratic Variation of Paths</td>
<td>408</td>
</tr>
<tr>
<td>9.4 Law of the Iterated Logarithm</td>
<td>410</td>
</tr>
<tr>
<td>9.5 The Markov Property</td>
<td>414</td>
</tr>
<tr>
<td>9.6 Martingales</td>
<td>420</td>
</tr>
<tr>
<td>9.7 Itô Integrals</td>
<td>426</td>
</tr>
<tr>
<td>9.8 Itô's Differentiation Formula</td>
<td>432</td>
</tr>
<tr>
<td>9.9 References</td>
<td>437</td>
</tr>
<tr>
<td>Appendices</td>
<td>438</td>
</tr>
<tr>
<td>1. The Symmetric Random Walk in ( \mathbb{R}^d )</td>
<td>438</td>
</tr>
<tr>
<td>2. Semicontinuous Functions</td>
<td>441</td>
</tr>
<tr>
<td>3. Completion of the Proof of Theorem 7.3.2</td>
<td>443</td>
</tr>
<tr>
<td>4. Proof of the Convergence of Types Theorem 7.3.4</td>
<td>447</td>
</tr>
<tr>
<td>5. The Multivariate Normal Distribution</td>
<td>449</td>
</tr>
<tr>
<td>Bibliography</td>
<td>454</td>
</tr>
<tr>
<td>Solutions to Problems</td>
<td>456</td>
</tr>
<tr>
<td>Index</td>
<td>512</td>
</tr>
</tbody>
</table>
Preface

It is a pleasure to accept the invitation of Harcourt/Academic Press to publish a second edition. The first edition has been used mainly in graduate courses in measure and probability, offered by departments of mathematics and statistics and frequently taken by engineers. We have prepared the present text with this audience in mind, and the title has been changed from *Real Analysis and Probability* to *Probability and Measure Theory* to reflect the revisions we have made.

Chapters 1 and 2 develop the fundamentals of measure and integration theory. Included are several results that are crucial in constructing the foundations of probability: the Radon–Nikodym theorem, the product measure theorem, the Kolmogorov extension theorem and the theory of weak convergence of measures. We remain convinced that it is best to assemble a complete set of measure-theoretic tools before going into probability, rather than try to develop both areas simultaneously. The gain in efficiency far outweighs any temporary loss in motivation. Those who wish to reach probability as quickly as possible may omit Chapter 3, which gives a brief introduction to functional analysis, and Section 2.3, which gives some applications to real analysis. In addition, instructors may wish to summarize or sketch some of the intricate constructions in Sections 1.3, 1.4, and 2.7.

The study of probability begins with Chapter 4, which offers a summary of an undergraduate probability course from a measure-theoretic point of view. Chapter 5 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 6 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 7 considers the central limit problem, emphasizing the fundamental role of Prokhorov’s weak compactness theorem. The last two sections of this chapter cover some material (not in the first edition) of special interest to statisticians: Slutsky’s theorem, the Skorokhod construction, convergence of transformed sequences and a $k$-dimensional central limit theorem.
Chapters 8 and 9 have been added in the second edition, and should be of interest to the entire prospective audience: mathematicians, statisticians, and engineers. Chapter 8 covers ergodic theory, which is developed far enough so that connections with information theory are clearly visible. The Shannon–McMillan theorem is proved and the isomorphism problem for Bernoulli shifts is discussed. Chapter 9 treats the one-dimensional Brownian motion in detail, and then introduces stochastic integrals and the Itô differentiation formula.

To make room for the new material, the appendix on general topology and the old Chapter 4 on the interplay between measure theory and topology have been removed, along with the section on topological vector spaces in Chapter 3. We assume that the reader has had a course in basic analysis and is familiar with metric spaces, but not with general topology. All the necessary background appears in Real Variables With Basic Metric Space Topology by Robert B. Ash, IEEE Press, 1993. (The few exercises that require additional background are marked with an asterisk.)

It is theoretically possible to read the text without any prior exposure to probability, picking up the necessary equipment in Chapter 4. But we expect that in practice, almost all readers will have taken a standard undergraduate probability course. We believe that discrete time, discrete state Markov chains, and random walks are best covered in a second undergraduate probability course, without measure theory. But instructors and students usually find this area appealing, and we discuss the symmetric random walk on \( \mathbb{R}^1 \) in Appendix 1.

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 in Chapters 1–8 not worked out in the text by writing to the publisher.

Catherine Doleans–Dade wrote Chapter 9, and offered valuable advice and criticism for the other chapters. Mel Gardner kindly allowed some material from Topics in Stochastic Processes by Ash and Gardner to be used in Chapter 8. We appreciate the encouragement and support provided by the staff at Harcourt/Academic Press.

Robert B. Ash
Catherine Doleans–Dade
Urbana, Illinois, 1999
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 appendices, the letter A is used; thus A2.3 means Part 3 of Appendix 2.

The symbol □ is used to mark the end of a proof.

1 Sets

If A and B are subsets of a set Ω, A ∪ B will denote the union of A and B, and A ∩ 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 ⊆ A; the inclusion need not be proper, that is, we have A ⊆ A for any set A. We also write B ⊆ A as A ⊇ 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 ⊆ A.

The symmetric difference between A and B is by definition the union of A − B and B − A; it is denoted by A Δ B.

If A_1 ⊆ A_2 ⊆ ⋯ and \( \bigcup_{n=1}^{∞} A_n = A \), we say that the A_n form an increasing sequence of sets (increasing to A) and write A_n ↑ A. Similarly, if A_1 ⊇ A_2 ⊇ ⋯ and \( \bigcap_{n=1}^{∞} A_n = A \), we say that the A_n form a decreasing sequence of sets (decreasing to A) and write A_n ↓ 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, “\( \mathcal{D} \) includes \( \mathcal{C} \)” means that \( \mathcal{D} \subseteq \mathcal{C} \). 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.

The empty set \( \emptyset \) is the set with no members. The sets A_i, i ∈ I, are disjoint if A_i ∩ A_j = \( \emptyset \) for all i ≠ j.
2 Real Numbers

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

The interval \((a, b]\) is defined as \( \{x \in \mathbb{R}^n : a_1 < x_1 \leq b_1, i = 1, \ldots, n\} \), and other types of intervals are defined similarly.

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

\[
\begin{align*}
\infty + \infty &= \infty + a = \infty, & a - \infty &= -\infty + a = -\infty, & a \in \mathbb{R}, \\
\infty + \infty &= \infty, & -\infty - \infty &= -\infty & (\infty - \infty \text{ is not defined}), \\
b \cdot \infty &= \infty \cdot b = \begin{cases} 
\infty & \text{if } b \in \mathbb{R}, \quad b > 0, \\
-\infty & \text{if } b \in \mathbb{R}, \quad b < 0,
\end{cases} \\
\frac{a}{\infty} &= -\infty = 0, & a \in \mathbb{R}, & \left(\frac{\infty}{\infty}\right) \text{ is not defined}, \\
0 \cdot \infty &= \infty \cdot 0 = 0.
\end{align*}
\]

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

Unless otherwise specified, positive means (strictly) greater than zero, and nonnegative means greater than or equal to zero.

The set of complex numbers is denoted by \( \mathbb{C} \), and the set of \( n \)-tuples of complex numbers by \( \mathbb{C}^n \).

3 Functions

If \( f \) is a function from \( \Omega \) to \( \Omega' \) (written as \( f : \Omega \to \Omega' \)) 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 \( f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) \)
\( = f^{-1}(B_1) \cap f^{-1}(B_2) \) (the intersection of \( f^{-1}(B_1) \) and \( f^{-1}(B_2) \))
\( = f^{-1}(B_1, B_2) \) (the union of \( f^{-1}(B_1) \) and \( f^{-1}(B_2) \)); hence \( f^{-1}(A^c) = (f^{-1}(A))^c \).

If \( \mathcal{F} \) is a class of sets, \( f^{-1}(\mathcal{F}) \) means the collection of sets \( f^{-1}(B), B \in \mathcal{F} \).

If \( f : \mathbb{R} \to \mathbb{R}, f \) 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 \( f_n : \Omega \to \mathbb{R}, n = 1, 2, \ldots, \) the \( f_n \) are said to form an increasing sequence iff \( f_n(\omega) \leq f_{n+1}(\omega) \) for all \( n \) and \( \omega \); a decreasing sequence is defined similarly.
If \( f \) and \( g \) are functions from \( \Omega \) to \( \mathbb{R} \), 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 \( f_i : \Omega \to \mathbb{R} \) for each \( i \in I \), sup, \( f_i \) is the function whose value at \( \omega \) is \( \sup \{ f_i(\omega) : i \in I \} \).

If \( f_1, f_2, \ldots \) form an increasing sequence of functions with limit \( f \) (that is, \( \lim_{n \to \infty} f_n(\omega) = f(\omega) \) for all \( \omega \)), 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 \( A \subseteq \Omega \), 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 \( y \mapsto f(x, y) \) with \( x \) fixed.

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

If \( f : \Omega \to \mathbb{R} \), 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 \( \Omega \) 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 \( \Omega \), and \( r \), the radius, is a positive real number. A closed ball is a set of the form \( B(x, r) = \{ y \in \Omega : d(x, y) \leq r \} \).

Sequences in \( \Omega \) are denoted by \( \{ x_n, n = 1, 2, \ldots \} \). The term “lower semicontinuous” is abbreviated LSC, and “upper semicontinuous” is abbreviated USC.

No knowledge of general topology (beyond metric spaces) is assumed, and the few comments that refer to general topological spaces can safely be ignored.
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 term “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 \( x_1, \ldots, x_n \in B, n = 1, 2, \ldots, \) and \( c_1, \ldots, c_n \) 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 metric on \( L \) and \( M \) is a closed subset of \( L \), 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 metric 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 \( C \subseteq S \), \( C \) 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 \( m \neq a \).

Zorn’s lemma is actually an axiom of set theory, equivalent to the axiom of choice.
CHAPTER 1

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 $A_1, A_2, \ldots$ be subsets of a set $\Omega$. 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 with limit $A$, or that the $A_n$ increase to $A$; we write $A_n \uparrow A$. 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 with limit $A$, or that the $A_n$ decrease to $A$; we write $A_n \downarrow A$.

The De Morgan laws, namely, $(\bigcup_{n=1}^{\infty} A_n)^c = \bigcap_{n=1}^{\infty} A_n^c$, $(\bigcap_{n=1}^{\infty} A_n)^c = \bigcup_{n=1}^{\infty} A_n^c$, 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 $A_1, A_2, \ldots$ be subsets of $\Omega$. For each $n$ we have

\begin{align*}
(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) \\
&\quad \cdots \cup (A_1^c \cap \cdots A_{n-1}^c \cap A_n).
\end{align*}

Furthermore,

\begin{align*}
(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).
\end{align*}

In (2) and (3), the sets on the right are disjoint. If the $A_n$ form an increasing sequence, the formulas become
2.1  FUNDAMENTALS OF MEASURE AND INTEGRATION THEORY

(4) \[ \bigcup_{i=1}^{n} A_i = A_1 \cup (A_2 - A_1) \cup \cdots \cup (A_n - A_{n-1}) \]
and
(5) \[ \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_n - A_{n-1}) \]
(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 \( A_1, A_2, \ldots \) are subsets of \( \Omega \), we define
(6) \[ \limsup_{n} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \]
Thus \( \omega \in \limsup_{n} A_n \) iff for every \( n \), \( \omega \in A_k \) for some \( k \geq n \), in other words,
(7) \[ \omega \in \limsup_{n} A_n \text{ iff } \omega \in A_n \text{ for infinitely many } n. \]
Also define
(8) \[ \liminf_{n} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \]
Thus \( \omega \in \liminf_{n} A_n \) iff for some \( n \), \( \omega \in A_k \) for all \( k \geq n \), in other words,
(9) \[ \omega \in \liminf_{n} A_n \text{ iff } \omega \in A_n \text{ eventually, that is, for all but finitely many } n. \]

We shall 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) \( \left( \limsup_{n} A_n \right)^c = \liminf_{n} A_n^c \)
(11) \( \left( \liminf_{n} A_n \right)^c = \limsup_{n} A_n^c \)
(12) \( \lim inf_{n} A_n \subseteq \lim sup_{n} A_n \)
(13) If \( A_n \uparrow A \) or \( A_n \downarrow A \), then \( \lim inf_{n} A_n = \lim sup_{n} A_n = A. \)

In general, if \( \lim inf_{n} A_n = \lim sup_{n} A_n = A \), then \( A \) is said to be the limit of the sequence \( A_1, A_2, \ldots \); 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 \( \lim sup_{n} A_n \) and \( \lim inf_{n} A_n. \)
3. Let \( \Omega = \mathbb{R}^2, A_n \) the interior of the circle with center at \((-1)^n/12, 0)\) and radius 1. Find \( \lim sup_{n} A_n \) and \( \lim inf_{n} A_n. \)
4. Let \( \{x_n\} \) be a sequence of real numbers, and let \( A_n = (-\infty, x_n) \). What is the connection between \( \limsup_{n \to \infty} x_n \) and \( \limsup_n A_n \) (similarly for \( \liminf \))? 

5. Establish formulas (10)–(13).

6. Let \( A = (a, b) \) and \( B = (c, d) \) be disjoint open intervals of \( \mathbb{R} \), and let \( C_n = A \) if \( n \) is odd, \( C_n = B \) if \( n \) is even. Find \( \limsup_n C_n \) and \( \liminf_n C_n \).

### 1.2 Fields, \( \sigma \)-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 \( \mu \) is defined, and probability considerations provide a good motivation for the type of structure required. If \( \Omega \) is a set whose points correspond to the possible outcomes of a random experiment, certain subsets of \( \Omega \) will be called “events” and assigned a probability. Intuitively, \( A \) is an event if the question “Does \( \omega \) 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 \( i = 1, \ldots, n \), we can decide whether or not \( \omega \) belongs to \( A_i \), then we can determine whether or not \( \omega \) 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, as the answer to the question “Is \( \omega \in \Omega \)” is always “yes,” the entire space \( \Omega \) 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. Specifically, we are able to assert that the limit of a sequence of events is an event; see 1.2.1.

#### 1.2.1 Definitions

Let \( \mathcal{F} \) be a collection of subsets of a set \( \Omega \). 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 \( A_1, A_2, \ldots, A_n \in \mathcal{F} \), then \( \bigcup_{i=1}^n A_i \in \mathcal{F} \).

It follows that \( \mathcal{F} \) is closed under finite intersection. For if \( A_1, \ldots, A_n \in \mathcal{F} \), 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 \( A_1, A_2, \ldots \in \mathcal{F} \), then \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{F} \).

\( \mathcal{F} \) is called a \( \sigma \)-field (the term \( \sigma \)-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 \( \bigcup_{n=1}^{\infty} A_n \uparrow A \); conversely, if \( A_n \uparrow A \), then \( A = \bigcup_{n=1}^{\infty} A_n \).

This shows that a \( \sigma \)-field is a field that is closed under limits of increasing sequences.

1.2.2 Examples. The largest \( \sigma \)-field of subsets of a fixed set \( \Omega \) is the collection of all subsets of \( \Omega \). The smallest \( \sigma \)-field consists of the two sets \( \emptyset \) and \( \Omega \).

Let \( A \) be a nonempty proper subset of \( \Omega \), and let \( \mathcal{F} = \{ \emptyset, \Omega, A, A' \} \). Then \( \mathcal{F} \) is the smallest \( \sigma \)-field containing \( A \). For if \( \mathcal{F} \) is a \( \sigma \)-field and \( A \in \mathcal{F} \), then by definition of a \( \sigma \)-field, \( \Omega \), \( \emptyset \), and \( A' \) belong to \( \mathcal{F} \), hence \( \mathcal{F} \subseteq \mathcal{F} \). But \( \mathcal{F} \) is a \( \sigma \)-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 \( \sigma \)-field that is included in any \( \sigma \)-field containing \( A \), and the result follows.

If \( A_1, \ldots, A_n \) are arbitrary subsets of \( \Omega \), the smallest \( \sigma \)-field containing \( A_1, \ldots, A_n \) may be described explicitly; see Problem 8.

If \( \mathcal{F} \) is a class of sets, the smallest \( \sigma \)-field containing the sets of \( \mathcal{F} \) will be written as \( \sigma(\mathcal{F}) \), and sometimes called the minimal \( \sigma \)-field over \( \mathcal{F} \). We also call \( \sigma(\mathcal{F}) \) the \( \sigma \)-field generated by \( \mathcal{F} \), and currently this is probably the most common terminology.

Let \( \Omega \) be the set \( \mathbb{R} \) 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 \( (a, \infty) \) 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 \( (a, \infty) \) 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 \( \sigma \)-field; for example, \( A_n = (0, 1 - (1/n)] \) \( \in \mathcal{F} \), \( n = 1, 2, \ldots \), and \( \bigcup_{n=1}^{\infty} A_n = (0, 1) \notin \mathcal{F} \).

If \( \Omega \) is the set \( \mathbb{R} = [-\infty, \infty] \) of extended real numbers, then just as above, the collection of finite disjoint unions of right-semiclosed intervals forms a field but not a \( \sigma \)-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 \( \sigma \)-fields that it deserves to be displayed explicitly, as in the following typical illustration.
1.2 Fields, $\sigma$-Fields, and Measures

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

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

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

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

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

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

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

1.2.3 Definitions and Comments. A measure on a $\sigma$-field $\mathcal{F}$ is a nonnegative, extended real-valued function $\mu$ on $\mathcal{F}$ such that whenever $A_1, A_2, \ldots$ 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 $\mu(\Omega) = 1$, $\mu$ is called a probability measure.

A measure space is a triple $(\Omega, \mathcal{F}, \mu)$ where $\Omega$ is a set, $\mathcal{F}$ is a $\sigma$-field of subsets of $\Omega$, and $\mu$ is a measure on $\mathcal{F}$. If $\mu$ 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 $\sigma$-field. Let $\mathcal{F}$ be a field, $\mu$ a set function on $\mathcal{F}$ (a map from $\mathcal{F}$ to $\mathbb{R}$). We say that $\mu$ is countably additive on $\mathcal{F}$ iff whenever $A_1, A_2, \ldots$ 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 $\sigma$-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 $\mathcal{F}$, $\mu$ is said to be finitely additive on $\mathcal{F}$. To avoid the appearance of terms of the form
In the summation, we always assume that $+\infty$ and $-\infty$ cannot both belong to the range of $\mu$.

If $\mu$ is countably additive and $\mu(A) \geq 0$ for all $A \in \mathcal{F}$, $\mu$ 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 $A, 0, 0, \ldots$, we find that $\mu(0) = 0$, and finite additivity is now established by considering the sequence $A_1, \ldots, A_n, 0, 0, \ldots$, where $A_1, \ldots, A_n$ 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 $\mu(\emptyset) = 0$.

If $A \in \mathcal{F}$ and $\mu(A^c) = 0$, we can frequently ignore $A^c$; we say that $\mu$ is concentrated on $A$.

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

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

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

Thus if $A = \{x_1, x_2, \ldots\}$, then $\mu(A) = p_1 + p_2 + \cdots$. The set function $\mu$ is a measure on $\mathcal{F}$ and $\mu(x_i) = p_i$, $i = 1, 2, \ldots$. A probability measure will be obtained iff $\sum_i p_i = 1$; if all $p_i = 1$, then $\mu$ is counting measure.

Now if $A$ is a subset of $\mathbb{R}$, 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 $a, b \in \mathbb{R}, a < b$, and that $\mu$ be a measure, determine $\mu$ on a large class of sets.

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

Note that $\mathcal{B}(\mathbb{R})$ is guaranteed to exist; it may be described (admittedly in a rather ethereal way) as the intersection of all $\sigma$-fields containing the intervals
1.2 Fields, $\sigma$-fields, and Measures

(a, b]. Also, if a $\sigma$-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 $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-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 $[a, b], a, b \in \mathbb{R},$
- all intervals $(a, \infty), a \in \mathbb{R},$
- all intervals $[a, \infty), a \in \mathbb{R},$
- all intervals $(-\infty, b), b \in \mathbb{R},$
- all intervals $(-\infty, b], b \in \mathbb{R}.$

Since a $\sigma$-field that contains all intervals of a given type contains all intervals of any other type, $\mathcal{B}(\mathbb{R})$ may be described as the smallest $\sigma$-field that contains the class of all intervals of $\mathbb{R}$. Similarly, $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-field containing all open sets of $\mathbb{R}$. (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, $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-field containing all closed sets of $\mathbb{R}$. Finally, if $\mathcal{F}_0$ is the field of finite disjoint unions of right-semiclosed intervals (see 1.2.2), then $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-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 $\mathbb{R}$, denoted by $\mathcal{B}(\mathbb{R})$, is defined as the smallest $\sigma$-field of subsets of $\mathbb{R}$ containing all intervals $(a, b], a, b \in \mathbb{R}$. The above discussion concerning the replacement of the right-semiclosed intervals by other classes of sets applies equally well to $\mathbb{R}$.

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

We now begin to develop some properties of set functions.

1.2.5 Theorem. Let $\mu$ 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 \( \mu \) is nonnegative,
\[
\mu \left( \bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} \mu(A_i) \quad \text{for all} \quad A_1, \ldots, A_n \in \mathcal{F}.
\]

If \( \mu \) is a measure,
\[
\mu \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu(A_n)
\]
for all \( A_1, A_2, \ldots \in \mathcal{F} \) 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,
\[
\mu(A) = \mu(A \cap B) + \mu(A - B),
\]
\[
\mu(B) = \mu(A \cap B) + \mu(B - A).
\]

Add the above equations to obtain
\[
\mu(A) + \mu(B) = \mu(A \cap B) + [\mu(A - B) + \mu(B - A) + \mu(A \cap B)]
\]
\[
= \mu(A \cap B) + \mu(A \cup B).
\]

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

(d) We have
\[
\bigcup_{i=1}^{n} A_i = A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \cup \cdots \cup (A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n)
\]
[see Section 1.1, formula (2)]. The sets on the right are disjoint and
\[
\mu(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) \leq \mu(A_n) \quad \text{by (c)}.
\]

The case in which \( \mu \) is a measure is handled using identity (3) of Section 1.1. \( \square \)

**1.2.6 Definitions.** A set function \( \mu \) 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 \( \mu \) is finitely additive, it is
1.2 FIELDS, $\sigma$-FIELDS, AND MEASURES

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 $\mu$ on the field $\mathcal{F}$ is said to be $\sigma$-finite on $\mathcal{F}$ iff $\Omega$ 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 $\sigma$-finite measures.

It follows from 1.2.5(c) that a nonnegative, finitely additive set function $\mu$ 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).

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

1.2.7 Theorem. Let $\mu$ be a countably additive set function on the $\sigma$-field $\mathcal{F}$.

(a) If $A_1, A_2, \ldots \in \mathcal{F}$ and $A_n \uparrow A$, then $\mu(A_n) \to \mu(A)$ as $n \to \infty$.

(b) If $A_1, A_2, \ldots \in \mathcal{F}$, $A_n \downarrow A$, 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 $\mu(A_n) \to \mu(A)$ as $n \to \infty$.

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 identity (5) of Section 1.1:

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

Therefore, by 1.2.5(c),

$$\mu(A) = \mu(A_1) + \mu(A_2) - \mu(A_1) + \cdots + \mu(A_n) - \mu(A_{n-1}) + \cdots$$

$$= \lim_{n \to \infty} \mu(A_n).$$

(b) If $A_n \downarrow A$, then $A_1 - A_n \uparrow A_1 - A$, hence $\mu(A_1 - A_n) \to \mu(A_1 - A)$ 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 $\mu$ be a finitely additive set function on the field $\mathcal{F}$.

(a) Assume that $\mu$ is continuous from below at each $A \in \mathcal{F}$, that is, if $A_1, A_2, \ldots \in \mathcal{F}$, $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, and $A_n \uparrow A$, then $\mu(A_n) \to \mu(A)$. It follows that $\mu$ is countably additive on $\mathcal{F}$.

(b) Assume that $\mu$ is continuous from above at the empty set, that is, if $A_1, A_2, \ldots \in \mathcal{F}$ and $A_n \downarrow \emptyset$, then $\mu(A_n) \to 0$. It follows that $\mu$ is countably additive on $\mathcal{F}$.

Proof. (a) Let $A_1, A_2, \ldots$ 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 $\mu(B_n) \to \mu(A)$ by hypothesis. But $\mu(B_n) = \sum_{i=1}^{n} \mu(A_i)$ by finite additivity, hence $\mu(A) = \lim_{n \to \infty} \sum_{i=1}^{n} \mu(A_i)$, the desired result.

(b) Let $A_1, A_2, \ldots$ 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, $\mu(A - B_n) \to 0$. Thus $\mu(B_n) \to \mu(A)$, and the result follows as in (a). □

If $\mu_1$ and $\mu_2$ are measures on the $\sigma$-field $\mathcal{F}$, then $\mu = \mu_1 - \mu_2$ is countably additive on $\mathcal{F}$, assuming either $\mu_1$ or $\mu_2$ is finite-valued. We shall see later (in 2.1.3) that any countably additive set function on a $\sigma$-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

1. Let $\Omega$ be a countably infinite set, and let $\mathcal{F}$ consist of all subsets of $\Omega$. Define $\mu(A) = 0$ if $A$ is finite, $\mu(A) = \infty$ if $A$ is infinite.

(a) Show that $\mu$ is finitely additive but not countably additive.

(b) Show that $\Omega$ is the limit of an increasing sequence of sets $A_n$ with $\mu(A_n) = 0$ for all $n$, but $\mu(\Omega) = \infty$.

2. Let $\mu$ be counting measure on $\Omega$, where $\Omega$ is an infinite set. Show that there is a sequence of sets $A_n \downarrow \emptyset$ with $\lim_{n \to \infty} \mu(A_n) \neq 0$.

3. Let $\Omega$ be a countably infinite set, and let $\mathcal{F}$ be the field consisting of all finite subsets of $\Omega$ and their complements. If $A$ is finite, set $\mu(A) = 0$, and if $A'$ is finite, set $\mu(A) = 1$.

(a) Show that $\mu$ is finitely additive but not countably additive on $\mathcal{F}$.
1.2 FIELDS, $\sigma$-FIELDS, AND MEASURES

(b) Show that $\Omega$ 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 $\mathbb{R}$, and define the set function $\mu$ on $\mathcal{F}$ as follows.

$$
\begin{align*}
\mu(-\infty, a] &= a, & a \in \mathbb{R}, \\
\mu(a, b] &= b - a, & a, b \in \mathbb{R}, & a < b, \\
\mu(b, \infty) &= -b, & b \in \mathbb{R}, \\
\mu(\mathbb{R}) &= 0,
\end{align*}
$$

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

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

(a) Show that $\mu$ is finitely additive but not countably additive on $\mathcal{F}$.

(b) Show that $\mu$ is finite but unbounded on $\mathcal{F}$.

5. Let $\mu$ be a nonnegative, finitely additive set function on the field $\mathcal{F}$. If $A_1, A_2, \ldots$ 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 $f: \Omega \to \Omega'$, and let $\mathcal{E}$ be a class of subsets of $\Omega'$. Show that

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

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

7. If $A$ is a Borel subset of $\mathbb{R}$, show that the smallest $\sigma$-field of subsets of $A$ containing the sets open in $A$ (in the relative topology inherited from $\mathbb{R}$) is $\{B \in \mathcal{B}(\mathbb{R}) : B \subset A\}$.

8. Let $A_1, \ldots, A_n$ be arbitrary subsets of a set $\Omega$. Describe (explicitly) the smallest $\sigma$-field $\mathcal{F}$ containing $A_1, \ldots, A_n$. How many sets are there in $\mathcal{F}$? (Give an upper bound that is attainable under certain conditions.) List all the sets in $\mathcal{F}$ when $n = 2$.

9. (a) Let $\mathcal{E}$ be an arbitrary class of subsets of $\Omega$, and let $\mathcal{F}$ be the collection of all finite unions $\bigcup_{i=1}^{n} A_i$, $n = 1, 2, \ldots$, where each $A_i$ is a finite intersection $\bigcap_{j=1}^{m} B_{ij}$, with $B_{ij}$ or its complement a set in $\mathcal{E}$. Show that $\mathcal{F}$ is the minimal field (not $\sigma$-field) over $\mathcal{E}$.

(b) Show that the minimal field can also be described as the collection $\mathcal{F}$ of all finite disjoint unions $\bigcup_{i=1}^{n} A_i$, where the $A_i$ are as above.
(c) If $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are fields of subsets of $\Omega$, show that the smallest field including $\mathcal{F}_1, \ldots, \mathcal{F}_n$ consists of all finite (disjoint) unions of sets $A_1 \cap \cdots \cap A_n$ with $A_i \in \mathcal{F}_i$, $i = 1, \ldots, n$. 

10. Let $\mu$ be a finite measure on the $\sigma$-field $\mathcal{F}$. If $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ and $A = \lim_{n \to \infty} A_n$ (see Section 1.1), show that $\mu(A) = \lim_{n \to \infty} \mu(A_n)$. 

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

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

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

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

(a) Show that $\mathcal{F}$ is a $\sigma$-field (hence $\mathcal{F} = \mathcal{F}$ by minimality of $\mathcal{F}$). 

(b) If the cardinality of $\mathcal{G}$ is at most $c$, the cardinality of the reals, show that $\text{card } \mathcal{F} \leq c$ also. 

12. Show that if $\mu$ is a finite measure, there cannot be uncountably many disjoint sets $A$ such that $\mu(A) > 0$. 

### 1.3 Extension of Measures

In 1.2.4, we discussed the concept of length of a subset of $\mathbb{R}$. 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 $\mu$ to $\mathcal{F}_0$: if $A_1, \ldots, A_n$ are disjoint right-semiclosed intervals, we set $\mu \left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} \mu(A_i)$. 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 $\mu$ to the minimal $\sigma$-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 $\mathbb{R}$ to $\mathbb{R}$. The extension technique to be developed is not restricted to set functions defined on subsets of $\mathbb{R}$; we shall prove a general result concerning the extension of a measure from a field $\mathcal{F}_0$ to the minimal $\sigma$-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 $\Omega$, and let $P$ be a probability measure on $\mathcal{F}_0$. Suppose that the sets $A_1, A_2, \ldots$ belong to $\mathcal{F}_0$ and
increase to a limit \( A \), and that the sets \( A_1', A_2', \ldots \) belong to \( \mathcal{F}_0 \) and increase to \( A' \). (\( A \) and \( A' \) need not belong to \( \mathcal{F}_0 \).) If \( A \subset A' \), then
\[
\lim_{m \to \infty} P(A_m) \leq \lim_{n \to \infty} P(A_n').
\]
Thus if \( A_n \) and \( A_n' \) both increase to the same limit \( A \), then
\[
\lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} P(A_n').
\]

**Proof.** If \( m \) is fixed, \( A_m \cap A_n' \uparrow A_m \cap A' = A_m \) as \( n \to \infty \), hence
\[
P(A_m \cap A_n') \to P(A_m)
\]
by 1.2.7(a). But \( P(A_m \cap A_n') \leq P(A_n') \) by 1.2.5(c), hence
\[
P(A_m) = \lim_{n \to \infty} P(A_m \cap A_n') \leq \lim_{n \to \infty} P(A_n').
\]
Let \( m \to \infty \) to finish the proof. \( \Box \)

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 \( A_n \in \mathcal{F}_0, \ n = 1, 2, \ldots, \) 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 \( \mu \) on \( \mathcal{G} \) as follows. If \( A_n \in \mathcal{F}_0, \ n = 1, 2, \ldots, A_n \uparrow A \ (\in \mathcal{G}) \), set
\[
\mu(A) = \lim_{n \to \infty} P(A_n);
\]
\( \mu \) 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 \( G_n \in \mathcal{G}, \ n = 1, 2, \ldots, \) and \( G_n \uparrow G \),
then \( G \in \mathcal{G} \) and \( \mu(G_n) \to \mu(G) \).

**Proof.**

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

(b) Let \( A_n \in \mathcal{F}_0, \ A_n \uparrow G_1; \ A_n \in \mathcal{F}_0, \ A_n \uparrow G_2 \). We have \( P(A_1 \cup A_n) + P(A_1 \cap A_n) = P(A_1) + P(A_n) \) by 1.2.5(b); let \( n \to \infty \) to complete the argument.

(c) This follows from 1.3.1.
Since $G$ is a countable union of sets in $\mathcal{F}_0$, $G \in \mathcal{F}$. Now for each $n$ we can find sets $A_{nm} \in \mathcal{F}_0$, $m = 1, 2, \ldots$, with $A_{nm} \uparrow G_n$ as $m \to \infty$. The situation may be represented schematically as follows:

\[
\begin{array}{cccc}
A_{11} & A_{12} & \cdots & A_{1m} & \cdots \\
A_{21} & A_{22} & \cdots & A_{2m} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \ddots \\
A_{n1} & A_{n2} & \cdots & A_{nm} & \cdots \\
& & & \uparrow G_n & \\
\end{array}
\]

Let $D_m = A_{1m} \cup A_{2m} \cup \cdots \cup A_{nm}$ (the $D_m$ form an increasing sequence).

The key step in the proof is the observation that

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

and, therefore,

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

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

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

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

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

Then:

(a) $\mu^* = \mu$ on $\mathcal{F}$, $0 \leq \mu^*(A) \leq 1$ 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).
1.3 Extension of Measures

(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 $\mu^*$ and from 1.3.2(c).

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

$$\mu^*(A) + \mu^*(B) + \varepsilon \geq \mu(G_1) + \mu(G_2) = \mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) \geq \mu^*(A \cup B) + \mu^*(A \cap B).$$

Since $\varepsilon$ is arbitrary, the result follows.

(c) This follows from the definition of $\mu^*$.

(d) By (c), $\mu^*(A) \geq \lim_{n \to \infty} \mu^*(A_n)$. If $\varepsilon > 0$, for each $n$ we may choose $G_n \in \mathcal{F}$, $G_n \supset A_n$, such that $\mu(G_n) \leq \mu^*(A_n) + \varepsilon 2^{-n}$.

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

$$\mu^*(A) \leq \mu^* \left( \bigcup_{n=1}^{\infty} G_n \right) \quad \text{by (c)}$$

$$= \mu \left( \bigcup_{n=1}^{\infty} G_n \right) \quad \text{by (a)}$$

$$= \lim_{n \to \infty} \mu \left( \bigcup_{k=1}^{n} G_k \right) \quad \text{by 1.3.2(d)}.$$  

The proof will be accomplished if we prove that

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

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

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

\[
\mu\left(\bigcup_{i=1}^{n+1} G_i\right) \leq \mu^\ast(A_n) + \varepsilon \sum_{i=1}^{n} 2^{-i} + \mu^\ast(A_{n+1}) + \varepsilon 2^{-(n+1)} - \mu^\ast(A_n)
\]

\[
\leq \mu^\ast(A_{n+1}) + \varepsilon \sum_{i=1}^{n+1} 2^{-i}. \quad \square
\]

Our aim in this section is to prove that a \(\sigma\)-finite measure on a field \(\mathcal{G}_0\) has a unique extension to the minimal \(\sigma\)-field over \(\mathcal{G}_0\). In fact an arbitrary measure \(\mu\) on \(\mathcal{G}_0\) can be extended to \(\sigma(\mathcal{G}_0)\), 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 \(\Omega\) is a nonnegative, extended real-valued set function \(\lambda\) on the class of all subsets of \(\Omega\), satisfying

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

The set function \(\mu^\ast\) of 1.3.3 is an outer measure on \(\Omega\). 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:

\[
\mu^\ast\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} \mu^\ast\left(\bigcup_{i=1}^{n} A_i\right) \quad \text{by 1.3.3(d).}
\]

\[
\leq \lim_{n \to \infty} \sum_{i=1}^{n} \mu^\ast(A_i) \quad \text{by 1.3.3(b),}
\]

as desired.

We now identify a \(\sigma\)-field on which \(\mu^\ast\) is countably additive:

**1.3.5 Theorem.** Under the hypothesis of 1.3.2, with \(\mu^\ast\) defined as in 1.3.3, let \(\mathcal{H} = \{H \subset \Omega: \mu^\ast(H) + \mu^\ast(H^c) = 1\}\)

\(\mathcal{H} = \{H \subset \Omega: \mu^\ast(H) + \mu^\ast(H^c) \leq 1\} \) by 1.3.3(b),]

Then \(\mathcal{H}\) is a \(\sigma\)-field and \(\mu^\ast\) is a probability measure on \(\mathcal{H}\).

**Proof.** First note that \(\mathcal{F} \subset \mathcal{H}\). For if \(A_n \in \mathcal{F}_0\) and \(A_n \uparrow G \in \mathcal{F}\), then \(G^c \subset A_n^c\), so \(P(A_n) + \mu^\ast(G^c) \leq P(A_n) + P(A_n^c) = 1\). By 1.3.3(d), \(\mu^\ast(G) + \mu^\ast(G^c) \leq 1\).