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 $\{x \in \mathbb{R}: a < x \le b\}$, and (a, ∞) as $\{x \in \mathbb{R}: x > a\}$; other types of intervals are defined similarly. If $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ are points in $\mathbb{R}^n$, $a \le b$ will mean $a_i \le b_i$ for all i. The interval $(a, b]$ is defined as $\{x \in R^n: a_i < x_i \le b_i, i = 1, \dots, n\}$, and other types of intervals are defined similarly.

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

$$a + \infty = \infty + a = \infty, \qquad a - \infty = -\infty + a = -\infty, \qquad a \in \mathbb{R},$$

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

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

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

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

The rules are convenient when developing the properties of the abstract Lebesgue integral, but it should be emphasized that $\overline{\mathbb{R}}$ 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 Ω to Ω' (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 $f^{-1}(\bigcup_i B_i) = \bigcup_i f^{-1}(B_i)$, $f^{-1}(\bigcap_i B_i) = \bigcap_i f^{-1}(B_i)$, $f^{-1}(A - B) = f^{-1}(A) - f^{-1}(B)$; hence $f^{-1}(A^c) = [f^{-1}(A)]^c$. If $\mathscr{C}$ is a class of sets, $f^{-1}(\mathscr{C})$ means the collection of sets $f^{-1}(B), B \in \mathscr{C}$.

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

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

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 ω], we write $f_n \uparrow f$. (Similarly, $f_n \downarrow f$ is used for a decreasing sequence.)

Sometimes, a set such as $\{\omega \in \Omega\colon f(\omega) \le g(\omega)\}$ is abbreviated as $\{f \le g\}$; similarly, the preimage $\{\omega \in \Omega\colon f(\omega) \in B\}$ 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 $y \to f(x, y)$ 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: $\Omega \to \overline{\mathbb{R}}$, the *positive* and *negative parts* of f are defined by $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$, that is,

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

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

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 $d(x, y) \ge 0$, $d(x, y) = 0$ iff $x = y$, $d(x, y) = d(y, x)$, and $d(x, z) \le d(x, y) + d(y, z)$. If $d(x, y)$ can be 0 for $x \ne y$, 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 $B(x, r) = \{y \in \Omega\colon d(x, y) < r\}$ 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 $\overline{B}(x, r) = \{y \in \Omega\colon d(x, y) \le r\}$.

Sequences in Ω 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 $C \subset S$, 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 $m \neq a$.

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 Ω. 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, $\left(\bigcup_n A_n\right)^c = \bigcap_n A_n^c$, $\left(\bigcap_n A_n\right)^c = \bigcup_n A_n^c$, 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 $A_1, A_2, \ldots$ be subsets of Ω. For each n we have

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

Furthermore,

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

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

(4) $\bigcup_{i=1}^{n} A_i = A_1 \cup (A_2 - A_1) \cup \cdots \cup (A_n - A_{n-1})$
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 Ω, 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) $\omega \in \limsup_n A_n$ iff $\omega \in A_n$ 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) $\omega \in \liminf_n A_n$ iff $\omega \in A_n$ 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) $(\limsup_n A_n)^c = \liminf_n A_n^c$

(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 $\liminf_n A_n = \limsup_n A_n = A$, 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 $\limsup_n A_n$ and $\liminf_n A_n$.
3. Let $\Omega = \mathbb{R}^2$, 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 $\limsup_{n\to\infty} x_n$ 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, 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, as 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. 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 $\mathscr{F}$ be a collection of subsets of a set Ω. Then $\mathscr{F}$ is called a *field* (the term *algebra* is also used) iff $\Omega \in \mathscr{F}$ and $\mathscr{F}$ is closed under complementation and finite union, that is,

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

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

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

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

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

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

If $\mathscr{F}$ is a field, a countable union of sets in $\mathscr{F}$ can be expressed as the limit of an increasing sequence of sets in $\mathscr{F}$, 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 $\mathscr{F} = \{\emptyset, \Omega, A, A^c\}$. Then $\mathscr{F}$ is the smallest σ-field containing A. For if $\mathscr{G}$ is a σ-field and $A \in \mathscr{G}$, then by definition of a σ-field, Ω, $\emptyset$, and A^c belong to $\mathscr{G}$, hence $\mathscr{F} \subset \mathscr{G}$. But $\mathscr{F}$ is a σ-field, for if we form complements or unions of sets in $\mathscr{F}$, we invariably obtain sets in $\mathscr{F}$. Thus $\mathscr{F}$ is a σ-field that is included in any σ-field containing A, and the result follows.

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

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

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

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

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

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

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

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

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

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

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

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

1.2.3 Definitions and Comments. A *measure* on a σ-field $\mathscr{F}$ is a nonnegative, extended real-valued function μ on $\mathscr{F}$ such that whenever $A_1, A_2, \ldots$ form a finite or countably infinite collection of disjoint sets in $\mathscr{F}$, 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 $(\Omega, \mathscr{F}, \mu)$ where Ω is a set, $\mathscr{F}$ is a σ-field of subsets of Ω, and μ is a measure on $\mathscr{F}$. If μ is a probability measure, $(\Omega, \mathscr{F}, \mu)$ 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 $\mathscr{F}$ be a *field*, μ a set function on $\mathscr{F}$ (a map from $\mathscr{F}$ to $\overline{\mathbb{R}}$). We say that μ is *countably additive* on $\mathscr{F}$ iff whenever $A_1, A_2, \ldots$ form a finite or countably infinite collection of disjoint sets in $\mathscr{F}$ whose union also belongs to $\mathscr{F}$ (this will always be the case if $\mathscr{F}$ 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 $\mathscr{F}$, μ is said to be *finitely additive* on $\mathscr{F}$. 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 $A \in \mathscr{F}$, μ is called a *measure* on $\mathscr{F}$, a *probability measure* if $\mu(\Omega) = 1$.

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

Although the set function given by $\mu(A) = +\infty$ for all $A \in \mathscr{F}$ satisfies the definition of a measure, and similarly $\mu(A) = -\infty$ for all $A \in \mathscr{F}$ 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 $A \in \mathscr{F}$ and $\mu(A^c) = 0$, we can frequently ignore A^c; we say that μ is *concentrated* on A.

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

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 $\mathscr{F}$ as all subsets of Ω, and define

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

Thus if $A = \{x_{i_1}, x_{i_2}, \ldots\}$, then $\mu(A) = p_{i_1} + p_{i_2} + \cdots$. The set function μ is a measure on $\mathscr{F}$ 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 μ 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 μ be a measure, determine μ on a large class of sets.

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

Note that $\mathscr{B}(\mathbb{R})$ 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

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

Thus $\mathscr{B}(\mathbb{R})$ 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 $[a, b)$, $a, b \in \mathbb{R}$,
all intervals (a, ∞), $a \in \mathbb{R}$,
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 σ-field that contains all intervals of a given type contains all intervals of any other type, $\mathscr{B}(\mathbb{R})$ may be described as the smallest σ-field that contains the class of all intervals of $\mathbb{R}$. Similarly, $\mathscr{B}(\mathbb{R})$ is the smallest σ-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, $\mathscr{B}(\mathbb{R})$ is the smallest σ-field containing all closed sets of $\mathbb{R}$. Finally, if $\mathscr{F}_0$ is the field of finite disjoint unions of right-semiclosed intervals (see 1.2.2), then $\mathscr{B}(\mathbb{R})$ is the smallest σ-field containing the sets of $\mathscr{F}_0$.

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

If $E \in \mathscr{B}(\mathbb{R})$, $\mathscr{B}(E)$ will denote $\{B \in \mathscr{B}(\mathbb{R})\colon\ B \subset E\}$; this coincides with $\{A \cap E\colon\ A \in \mathscr{B}(\mathbb{R})\}$ (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 $\mathscr{F}$.

(a) $\mu(\emptyset) = 0$.
(b) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ for all $A, B \in \mathscr{F}$.
(c) If $A, B \in \mathscr{F}$ 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 $\mu(B) \le \mu(A)$ if $\mu(A - B) \ge 0$).

(d) If μ is nonnegative,

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mu(A_i) \qquad \text{for all} \qquad A_1, \ldots, A_n \in \mathscr{F}.$$

If μ is a measure,

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

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

PROOF. (a) Pick $A \in \mathscr{F}$ 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

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

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

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

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

$$\mu(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) \le \mu(A_n) \qquad \text{by (c).}$$

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

1.2.6 Definitions. A set function μ defined on $\mathscr{F}$ is said to be *finite* iff $\mu(A)$ is finite, that is, not $\pm\infty$, for each $A \in \mathscr{F}$. 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 $\mathscr{F}$ is said to be *σ-finite* on $\mathscr{F}$ iff Ω can be written as $\bigcup_{n=1}^{\infty} A_n$ where the A_n belong to $\mathscr{F}$ 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 $\mathscr{F}$ is finite iff it is bounded; that is, $\sup\{|\mu(A)|: A \in \mathscr{F}\} < \infty$. 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 $\mathscr{F}$.

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

(b) If $A_1, A_2, \ldots \in \mathscr{F}, A_n \downarrow A$, 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 $\mathscr{F}$ is only assumed to be a field, if we add the hypothesis that the limit sets A belong to $\mathscr{F}$. [If $A \notin \mathscr{F}$ 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),

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

(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 μ be a finitely additive set function on the field $\mathscr{F}$.

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

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

PROOF. (a) Let $A_1, A_2, \ldots$ be disjoint sets in $\mathscr{F}$ whose union A belongs to $\mathscr{F}$. 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 $\mathscr{F}$ whose union A belongs to $\mathscr{F}$, 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 μ_1 and μ_2 are measures on the σ-field $\mathscr{F}$, then $\mu = \mu_1 - \mu_2$ is countably additive on $\mathscr{F}$, 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

1. Let Ω be a countably infinite set, and let $\mathscr{F}$ consist of all subsets of Ω. Define $\mu(A) = 0$ if A is finite, $\mu(A) = \infty$ if A is infinite.
   (a) Show that μ is finitely additive but not countably additive.
   (b) Show that Ω is the limit of an increasing sequence of sets A_n with $\mu(A_n) = 0$ for all n, but $\mu(\Omega) = \infty$.
2. 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 $\lim_{n\to\infty} \mu(A_n) \neq 0$.
3. Let Ω be a countably infinite set, and let $\mathscr{F}$ 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 $\mu(A) = 1$.
   (a) Show that μ is finitely additive but not countably additive on $\mathscr{F}$.

(b) Show that Ω is the limit of an increasing sequence of sets $A_n \in \mathscr{F}$ with $\mu(A_n) = 0$ for all n, but $\mu(\Omega) = 1$.

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

$$\mu(-\infty, a] = a, \qquad a \in \mathbb{R},$$
$$\mu(a, b] = b - a, \qquad a, b \in \mathbb{R}, \qquad a < b,$$
$$\mu(b, \infty) = -b, \qquad b \in \mathbb{R},$$
$$\mu(\mathbb{R}) = 0,$$
$$\mu\left(\bigcup_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} \mu(I_i)$$

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

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

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

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

6. Let $f\colon \Omega \to \Omega'$, and let $\mathscr{C}$ be a class of subsets of Ω'. Show that

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

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

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

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

9. (a) Let $\mathscr{C}$ be an arbitrary class of subsets of Ω, and let $\mathscr{G}$ 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 $\bigcap_{j=1}^{r} B_{ij}$, with B_{ij} or its complement a set in $\mathscr{C}$. Show that $\mathscr{G}$ is the minimal field (not σ-field) over $\mathscr{C}$.

(b) Show that the minimal field can also be described as the collection $\mathscr{D}$ of all finite *disjoint* unions $\bigcup_{i=1}^{n} A_i$, where the A_i are as above.

(c) If $\mathscr{F}_1, \ldots, \mathscr{F}_n$ are fields of subsets of Ω, show that the smallest field including $\mathscr{F}_1, \ldots, \mathscr{F}_n$ consists of all finite (disjoint) unions of sets $A_1 \cap \cdots \cap A_n$ with $A_i \in \mathscr{F}_i$, $i = 1, \ldots, n$.

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

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

$$\mathscr{C}_\alpha = \left(\bigcup\{\mathscr{C}_\beta\colon \beta < \alpha\}\right)',$$

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

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

(a) Show that $\mathscr{S}$ is a σ-field (hence $\mathscr{S} = \mathscr{F}$ by minimality of $\mathscr{F}$).

(b) If the cardinality of $\mathscr{C}$ is at most c, the cardinality of the reals, show that card $\mathscr{F} \le c$ also.

12. Show that if μ 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 $\mathscr{F}_0$ is the field of finite disjoint unions of right-semiclosed intervals, there is no problem extending μ to $\mathscr{F}_0$: 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 $\mathscr{F}_0$ is finitely additive, but countable additivity is not clear at this point. Even if we can prove countable additivity on $\mathscr{F}_0$, we still have the problem of extending μ to the minimal σ-field over $\mathscr{F}_0$, 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 $\mathscr{F}_0$ to the minimal σ-field over $\mathscr{F}_0$.

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

$$\lim_{m\to\infty} P(A_m) \le \lim_{n\to\infty} P(A_n').$$

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 $P(A_m \cap A_n') \le P(A_n')$ by 1.2.5(c), hence

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

Let $m \to \infty$ 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 $\mathscr{F}_0$. Let $\mathscr{G}$ be the collection of all limits of increasing sequences of sets in $\mathscr{F}_0$, that is, $A \in \mathscr{G}$ iff there are sets $A_n \in \mathscr{F}_0$, $n = 1, 2, \ldots$, such that $A_n \uparrow A$. (Note that $\mathscr{G}$ can also be described as the collection of all countable unions of sets in $\mathscr{F}_0$; see 1.2.1.)

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

(a) $\emptyset \in \mathscr{G}$ and $\mu(\emptyset) = 0$; $\Omega \in \mathscr{G}$ and $\mu(\Omega) = 1$; $0 \le \mu(A) \le 1$ for all $A \in \mathscr{G}$.

(b) If $G_1, G_2 \in \mathscr{G}$, then $G_1 \cup G_2, G_1 \cap G_2 \in \mathscr{G}$ and

$$\mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) = \mu(G_1) + \mu(G_2).$$

(c) If $G_1, G_2 \in \mathscr{G}$ and $G_1 \subset G_2$, then $\mu(G_1) \le \mu(G_2)$.

(d) If $G_n \in \mathscr{G}$, $n = 1, 2, \ldots$, and $G_n \uparrow G$,
then $G \in \mathscr{G}$ and $\mu(G_n) \to \mu(G)$.

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

(b) Let $A_{n1} \in \mathscr{F}_0$, $A_{n1} \uparrow G_1$; $A_{n2} \in \mathscr{F}_0$, $A_{n2} \uparrow G_2$. 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 $n \to \infty$ to complete the argument.

(c) This follows from 1.3.1.

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

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

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

$$A_{nm} \subset D_m \subset G_m \qquad \text{for} \qquad n \le m \tag{1}$$

and, therefore,

$$P(A_{nm}) \le P(D_m) \le \mu(G_m) \qquad \text{for} \qquad n \le m. \tag{2}$$

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 μ. Now let $m \to \infty$ in (2) to obtain $\mu(G_n) \le \lim_{m\to\infty} P(D_m) \le \lim_{m\to\infty} \mu(G_m)$; then let $n \to \infty$ to conclude that $\lim_{n\to\infty} \mu(G_n) = \lim_{m\to\infty} P(D_m) = \mu(G)$. $\square$

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

$$\mu^*(A) = \inf\{\mu(G)\colon\ G \in \mathscr{G}, \quad G \supset A\}.$$

Then:

(a) $\mu^* = \mu$ on $\mathscr{G}$, $0 \le \mu^*(A) \le 1$ for all $A \subset \Omega$.

(b) $\mu^*(A \cup B) + \mu^*(A \cap B) \le \mu^*(A) + \mu^*(B)$; in particular, $\mu^*(A) + \mu^*(A^c) \ge \mu^*(\Omega) + \mu^*(\emptyset) = \mu(\Omega) + \mu(\emptyset) = 1$ by 1.3.2(a).

(c) If $A \subset B$, then $\mu^*(A) \le \mu^*(B)$.

(d) If $A_n \uparrow A$, then $\mu^*(A_n) \to \mu^*(A)$.

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

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

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

Since ε is arbitrary, the result follows.

(c) This follows from the definition of μ^*.

(d) By (c), $\mu^*(A) \ge \lim_{n\to\infty} \mu^*(A_n)$. If $\varepsilon > 0$, for each n we may choose $G_n \in \mathscr{G}$, $G_n \supset A_n$, such that

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

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

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

The proof will be accomplished if we prove that

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

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

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

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

$$\mu\left(\bigcup_{i=1}^{n+1} G_i\right) \le \mu^*(A_n) + \varepsilon \sum_{i=1}^{n} 2^{-i} + \mu^*(A_{n+1}) + \varepsilon 2^{-(n+1)} - \mu^*(A_n)$$

$$\le \mu^*(A_{n+1}) + \varepsilon \sum_{i=1}^{n+1} 2^{-i}. \quad \square$$

Our aim in this section is to prove that a σ-finite measure on a field $\mathscr{F}_0$ has a unique extension to the minimal σ-field over $\mathscr{F}_0$. In fact an arbitrary measure μ on $\mathscr{F}_0$ can be extended to $\sigma(\mathscr{F}_0)$, 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 $\lambda(A) \le \lambda(B)$ (monotonicity), and
(c) $\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \lambda(A_n)$ (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:

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu^*\left(\bigcup_{i=1}^{n} A_i\right) \qquad \text{by 1.3.3(d).}$$

$$\le \lim_{n\to\infty} \sum_{i=1}^{n} \mu^*(A_i) \qquad \text{by 1.3.3(b),}$$

as desired.

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

1.3.5 Theorem. Under the hypothesis of 1.3.2, with μ^* defined as in 1.3.3, let $\mathscr{H} = \{H \subset \Omega\colon\ \mu^*(H) + \mu^*(H^c) = 1\}$
[$\mathscr{H} = \{H \subset \Omega\colon\ \mu^*(H) + \mu^*(H^c) \le 1$ by 1.3.3(b).]
Then $\mathscr{H}$ is a σ-field and μ^* is a probability measure on $\mathscr{H}$.

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

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

$$\mu^*(H_1 \cup H_2) + \mu^*(H_1 \cap H_2) \le \mu^*(H_1) + \mu^*(H_2) \tag{1}$$

and since

$$(H_1 \cup H_2)^c = H_1^c \cap H_2^c, \qquad (H_1 \cap H_2)^c = H_1^c \cup H_2^c,$$

we have

$$\mu^*(H_1 \cup H_2)^c + \mu^*(H_1 \cap H_2)^c \le \mu^*(H_1^c) + \mu^*(H_2^c). \tag{2}$$

If H_1, $H_2 \in \mathscr{H}$, 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 $a \le 1$ or $b \le 1$. If $a \le 1$, then $a = 1$, so $b = 1$ also. Consequently $H_1 \cup H_2 \in \mathscr{H}$ and $H_1 \cap H_2 \in \mathscr{H}$. We have therefore shown that $\mathscr{H}$ 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 $\mathscr{H}$.

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

Since $\mu^*(H_n) \to \mu^*(H)$, 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 $\mathscr{F}_0$ can be extended to a measure on $\sigma(\mathscr{F}_0)$.

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

In fact there is very little difference between $\sigma(\mathscr{F}_0)$ and $\mathscr{H}$; if $B \in \mathscr{H}$, then B can be expressed as $A \cup N$, where $A \in \sigma(\mathscr{F}_0)$ and N is a subset of a set $M \in \sigma(\mathscr{F}_0)$ 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 $\mathscr{F}$ is said to be *complete* iff whenever $A \in \mathscr{F}$ and $\mu(A) = 0$ we have $B \in \mathscr{F}$ for all $B \subset A$.

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

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

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

We extend μ to $\mathscr{F}_\mu$ by setting $\mu(A \cup N) = \mu(A)$. This is a valid definition, for if $A_1 \cup N_1 = A_2 \cup N_2 \in \mathscr{F}_\mu$, 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 $(\Omega, \mathscr{F}_\mu, \mu)$ is called the *completion* of $(\Omega, \mathscr{F}, \mu)$, and $\mathscr{F}_\mu$ the completion of $\mathscr{F}$ relative to μ.

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

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

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

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

To prove the uniqueness of the extension from $\mathscr{F}_0$ to $\mathscr{F}$, we need the following basic result.

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

PROOF. The technique of the proof might be called "boot strapping." Let $\mathscr{F} = \sigma(\mathscr{F}_0)$ and let $\mathscr{M}$ 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 $\mathcal{M}_A = \{B \in \mathcal{M}\colon A \cap B, A \cap B^c \text{ and } A^c \cap B \in \mathcal{M}\}$; 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 $A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n$, 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 $\mathcal{F}$. As $\mu = \sum_n \mu_n$, 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 $\mathcal{C} = \{A \in \mathcal{F}\colon \lambda_n(A) = \mu_n^*(A)\}$ 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 $\mathscr{G}$ be the class of all countable unions of sets of $\mathscr{F}_0$. The conclusion of 1.3.11 holds for any $A \in \mathscr{G}$, by 1.2.7(a). By 1.3.3, if μ is finite and $A \in \mathscr{F}$, A can be approximated arbitrarily closely (in the sense of 1.3.11) by a set in $\mathscr{G}$, and therefore 1.3.11 is proved for finite μ. In general, let Ω be the disjoint union of sets $A_n \in \mathscr{F}_0$ with $\mu(A_n) < \infty$, and let $\mu_n(C) = \mu(C \cap A_n)$, $C \in \mathscr{F}$.

Then μ_n is a finite measure on $\mathscr{F}$, hence if $A \in \mathscr{F}$, there is a set $B_n \in \mathscr{F}_0$ 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 $B_n \cap A_n \in \mathscr{F}_0$, we may assume that $B_n \subset A_n$. (The observation that $B_n \cap A_n \in \mathscr{F}_0$ is the point where we use the hypothesis that μ is σ-finite on $\mathscr{F}_0$, not merely on $\mathscr{F}$.) 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

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

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

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

(a) $\mathscr{F}$ 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 $\mathscr{F}$ but not on $\mathscr{F}_0$.

(c) There are sets $A \in \mathscr{F}$ of finite measure that cannot be approximated by sets in $\mathscr{F}_0$, that is, there is no sequence $A_n \in \mathscr{F}_0$ with $\mu(A \,\Delta\, A_n) \to 0$.

(d) If $\lambda = 2\mu$, then $\lambda = \mu$ on $\mathscr{F}_0$ but not on $\mathscr{F}$.

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 $\mathscr{F}$. But then all sets are in $\mathscr{F}$ since Ω is countable.

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

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

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

Problems

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

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

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

2. Show that the monotone class theorem (1.3.9) fails if $\mathscr{F}_0$ 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

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

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

$$\lambda(A \cap E) = \sum_n \lambda(A \cap E_n). \tag{1}$$

   In particular, λ is a measure on $\mathscr{M}$. [Use the definition of λ-measurability to show that $\mathscr{M}$ is a field and that (1) holds for finite sequences. If $E_1, E_2, \ldots$ are disjoint sets in $\mathscr{M}$ 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 $n \to \infty$.]

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

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

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

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

(d) In (b), if μ is σ-finite on $\mathscr{F}_0$, show that $(\Omega, \mathscr{M}, \mu^*)$ is the completion of $[\Omega, \sigma(\mathscr{F}_0), \mu^*]$.

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 $\mathbb{R}$. If F is an increasing, right-continuous function from $\mathbb{R}$ to $\mathbb{R}$, we set $\mu(a, b] = F(b) - F(a)$; we then extend μ to a finitely additive set function on the field $\mathscr{F}_0(\mathbb{R})$ of finite disjoint unions of right-semiclosed intervals. If we can show that μ is countably additive on $\mathscr{F}_0(\mathbb{R})$, the Carathéodory extension theorem extends μ to $\mathscr{B}(\mathbb{R})$.

1.4.1 Definitions. A *Lebesgue–Stieltjes measure* on $\mathbb{R}$ is a measure μ on $\mathscr{B}(\mathbb{R})$ such that $\mu(I) < \infty$ for each bounded interval I. A *distribution function* on $\mathbb{R}$ is a map $F\colon \mathbb{R} \to \mathbb{R}$ that is *increasing* [$a < b$ implies $F(a) \le F(b)$] and *right-continuous* [$\lim_{x \to x_0^+} F(x) = F(x_0)$]. 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 $\mathbb{R}$. Let $F\colon \mathbb{R} \to \mathbb{R}$ 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 $F(b) - F(a) = \mu(a, b] \ge 0$. If $\{x_n\}$ is a sequence of points such that $x_1 > x_2 > \cdots \to x$, then $F(x_n) - F(x) = \mu(x, x_n] \to 0$ by 1.2.7(b). □

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

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

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

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

Thus we can find sets $B_n \in \mathscr{F}_0(\overline{\mathbb{R}})$ whose closures $\overline{B}_n$ (in $\overline{\mathbb{R}}$) 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 $\bigcap_{n=1}^\infty \overline{B}_n = \emptyset$, and it follows that $\bigcap_{k=1}^n \overline{B}_k = \emptyset$ for sufficiently large n. (Perhaps the easiest way to see this is to note that the sets $\overline{\mathbb{R}} - \overline{B}_n$ form an open covering of the compact set $\overline{\mathbb{R}}$, hence there is a finite subcovering, so that $\bigcup_{k=1}^n (\overline{R} - \overline{B}_k) = \overline{\mathbb{R}}$ for some n. Therefore $\bigcap_{k=1}^n \overline{B}_k = \emptyset$.) Now

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

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

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 $\mu_n \to \mu$ on $\mathscr{F}_0(\overline{\mathbb{R}})$. Let $A_1, A_2, \ldots$ be disjoint sets in $\mathscr{F}_0(\overline{\mathbb{R}})$ such that $A = \bigcup_{n=1}^\infty A_n \in \mathscr{F}_0(\overline{\mathbb{R}})$. 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

$$\mu(A) = \lim_{n\to\infty} \mu_n(A) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \mu_n(A_k)$$

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

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

We now complete the construction of Lebesgue–Stieltjes measures.

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

PROOF. Extend μ to a countably additive set function on $\mathscr{F}_0(\overline{\mathbb{R}})$ as above. Let $\mathscr{F}_0(\mathbb{R})$ be the field of all finite disjoint unions of right-semiclosed intervals of $\mathbb{R}$ [counting (a, ∞) as right-semiclosed; see 1.2.2], and extend μ to $\mathscr{F}_0(\mathbb{R})$ as in the discussion that follows 1.4.2. [Take $\mu(a, \infty) = F(\infty) - F(a)$; $\mu(-\infty, b] = F(b) - F(-\infty)$, $a, b \in \mathbb{R}$; $\mu(\mathbb{R}) = F(\infty) - F(-\infty)$; note that there is no other possible choice for μ on these sets, by 1.2.7(a).] Now the map

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

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

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 $\lim_{y\to x^-} F(y)$, then

(1) $\mu(a, b] = F(b) - F(a)$, (3) $\mu[a, b] = F(b) - F(a^-)$,

(2) $\mu(a, b) = F(b^-) - F(a)$, (4) $\mu[a, b) = F(b^-) - F(a^-)$.

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

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

Statement (3) follows because

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

(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 $+\infty$.

(6) $\mu(-\infty, x] = F(x) - F(-\infty)$, (9) $\mu[x, \infty) = F(\infty) - F(x^-)$,

(7) $\mu(-\infty, x) = F(x^-) - F(-\infty)$, (10) $\mu(\mathbb{R}) = F(\infty) - F(-\infty)$.

(8) $\mu(x, \infty) = F(\infty) - F(x)$,

(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 $\mathscr{B}(\mathbb{R})$. For example, if $f\colon \mathbb{R} \to \mathbb{R}$, $f \geq 0$, and f is integrable (Riemann for now) on any finite interval, then if we fix $F(0)$ arbitrarily and define

$$F(x) - F(0) = \int_0^x f(t)\,dt, \qquad x > 0;$$

$$F(0) - F(x) = \int_x^0 f(t)\,dt, \qquad x < 0,$$

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 $\mathscr{B}(\mathbb{R})$. The completion of $\mathscr{B}(\mathbb{R})$ relative to Lebesgue measure is called the class of *Lebesgue measurable sets*, written $\overline{\mathscr{B}}(\mathbb{R})$. 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 $\overline{\mathscr{B}}(\mathbb{R})$ is also called "Lebesgue measure."

Now let μ be a Lebesgue–Stieltjes measure that is concentrated on a countable set $S = \{x_1, x_2, \ldots\}$, that is, $\mu(\mathbb{R} - S) = 0$. [In general if $(\Omega, \mathscr{F}, \mu)$ is a measure space and $B \in \mathscr{F}$, we say that μ is concentrated on B iff $\mu(\Omega - B) = 0$.] In the present case, such a measure is easily constructed: If $a_1, a_2, \ldots$ are nonnegative numbers and $A \subset \mathbb{R}$, set $\mu(A) = \sum\{a_i\colon x_i \in A\}$; μ is a measure on all subsets of $\mathbb{R}$, 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 $\mathscr{B}(\mathbb{R})$; if $\sum_i a_i < \infty$, μ will be a finite measure. The distribution function F corresponding to μ is continuous on $\mathbb{R} - S$; 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 $x \le b < y$, 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 $\mathbb{R}$ to $\mathbb{R}$ 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 $\mu(a', b'] = F(b') - F(a')$, $a \le a' < b' \le b$. The easiest way to establish the correspondence is to extend F by defining $F(x) = F(b), x \ge b$; $F(x) = F(a), x \le a$; then take μ as the Lebesgue–Stieltjes measure corresponding to F, restricted to $\mathscr{B}[a, b]$.

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

1.4.6 Definitions and Comments. If $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in \mathbb{R}^n$, the interval $(a, b]$ is defined as $\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n\colon a_i < x_i \le b_i$ for all $i = 1, \ldots, n\}$; (a, ∞) is defined as $\{x \in \mathbb{R}^n\colon x_i > a_i$ for all $i = 1, \ldots, n\}$, $(-\infty, b]$ as $\{x \in \mathbb{R}^n\colon x_i \le b_i$ for all $i = 1, \ldots, n\}$; other types of intervals are defined similarly. The smallest σ-field containing all intervals $(a, b]$, $a, b \in \mathbb{R}^n$, is called the class of *Borel sets* of $\mathbb{R}^n$, written $\mathscr{B}(\mathbb{R}^n)$. 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 $\overline{\mathbb{R}}^n$, written $\mathscr{B}(\overline{\mathbb{R}}^n)$, is defined similarly.

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

The notion of a distribution function on $\mathbb{R}^n$, $n \geq 2$, 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 $\mathscr{B}(\mathbb{R}^3)$. Define

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

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 $G\colon\ \mathbb{R}^n \to \mathbb{R}$, $\triangle_{b_i a_i} G(x_1, \ldots, x_n)$ is defined as

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

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) $$\begin{aligned}\triangle_{b_1 a_1} \triangle_{b_2 a_2} \triangle_{b_3 a_3} & F(x_1, x_2, x_3) \\ &= F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ &\quad + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) - F(a_1, a_2, a_3)\end{aligned}$$

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

PROOF.

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

Similarly,

$$\triangle_{b_2a_2}\triangle_{b_3a_3}F(x_1,x_2,x_3)=\mu\{\omega\colon\ \omega_1\le x_1,\quad a_2<\omega_2\le b_2,\quad a_3<\omega_3\le b_3\}$$

and

$$\triangle_{b_1a_1}\triangle_{b_2a_2}\triangle_{b_3a_3}F(x_1,x_2,x_3)=\mu\{\omega\colon\ a_1<\omega_1\le b_1,\quad a_2<\omega_2\le b_2,\quad a_3<\omega_3\le b_3\}.$$

(b) $$\triangle_{b_3a_3}F(x_1,x_2,x_3)=F(x_1,x_2,b_3)-F(x_1,x_2,a_3),$$

$$\triangle_{b_2a_2}\triangle_{b_3a_3}F(x_1,x_2,x_3)=F(x_1,b_2,b_3)-F(x_1,a_2,b_3)-F(x_1,b_2,a_3)+F(x_1,a_2,a_3).$$

Thus $\triangle_{b_1a_1}\triangle_{b_2a_2}\triangle_{b_3a_3}F(x_1,x_2,x_3)$ 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 $\mathscr{B}(\mathbb{R}^n)$ and define

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

If $a\le b$, then

(a) $\mu(a,b]=\triangle_{b_1a_1}\cdots\triangle_{b_na_n}F(x_1,\dots,x_n)$, where

(b) $\triangle_{b_1a_1}\cdots\triangle_{b_na_n}F(x_1,\dots,x_n) = F_0-F_1+F_2-\cdots+(-1)^nF_n$;

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 of $\mathbb{R}$ determines a corresponding Lebesgue–Stieltjes measure. This is true in n dimensions if we change the definition of increasing.

Let $F\colon\ \mathbb{R}^n\to\mathbb{R}$, and, for $a\le b$, let $F(a,b]$ denote

$$\triangle_{b_1a_1}\cdots\triangle_{b_na_n}F(x_1,\dots,x_n).$$

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

An increasing right-continuous $F: \mathbb{R}^n \to \mathbb{R}$ is said to be a *distribution function* on $\mathbb{R}^n$. (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 $\mathbb{R}^n$, 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 $\mathbb{R}^n$. 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 $a \le a' \le b' \le b, I = (a, b]$ 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:

$$a_1 < x \le a_1', \qquad a_1' < x \le b_1', \qquad b_1' < x \le b_1,$$

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

$$a_2 < y \le a_2', \qquad a_2' < y \le b_2', \qquad b_2' < y \le b_2.$$

For example, a typical set in the union is

$$\{(x, y)\colon\ b_1' < x \le b_1, \qquad a_2 < y \le a_2'\};$$

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

Result (1) may be verified by looking at Fig. 1.4.1.

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

This is verified by brute force, using 1.4.8.

Now a *right-semiclosed interval* $(a, b]$ in $\overline{\mathbb{R}}^n$ is, by convention, a set of the form $\{(x_1, \dots, x_n)\colon\ a_i < x_i \le b_i, i = 1, \dots, n\}, a, b \in \overline{\mathbb{R}}^n$, with the proviso that $a_i < x_i \le b_i$ can be replaced by $a_i \le x_i \le b_i$ if $a_i = -\infty$. With this assumption, the set $\mathscr{F}_0(\overline{\mathbb{R}}^n)$ of finite disjoint unions of right-semiclosed intervals is a field. (The corresponding convention in $\mathbb{R}^n$ is that $a_i < x_i \le b_i$ can 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 $\overline{\mathbb{R}}^n$ but not to $\mathbb{R}^n$, we define $F(a, b]$ as the limit of $F(a', b']$ where $a', b' \in \mathbb{R}^n, a'$ decreases to a, and b' increases to b. [The definition is sensible because of the monotonicity property in (2).] Similarly if $a \in \mathbb{R}^n, b \in \overline{\mathbb{R}}^n - \mathbb{R}^n$, we take $F(a, b] = \lim_{b' \uparrow b} F(a, b']$; if $a \in \overline{\mathbb{R}}^n - \mathbb{R}^n$, $b \in \mathbb{R}^n$, $F(a, b] = \lim_{a' \downarrow a} F(a', b]$.

Thus we define μ on right-semiclosed intervals of $\overline{\mathbb{R}}^n$; μ extends to a finitely additive set function on $\mathscr{F}_0(\overline{\mathbb{R}}^n)$, as in the discussion after 1.4.2. [There is a

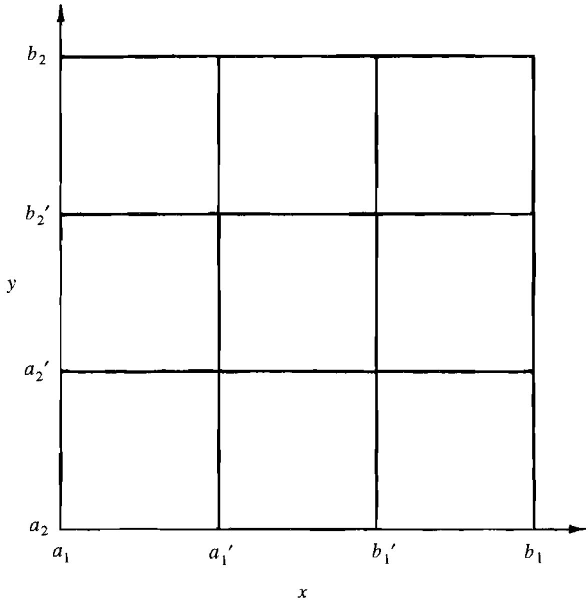

Figure 1.4.1.

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 $\mathscr{F}_0(\overline{\mathbb{R}}^n)$.

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

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

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

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

$$\mu\{(x, y, z)\colon a_1 < x \le b_1, \quad a_2 < y < \infty, \quad a_3 < z < \infty\} = \lim_{b_2, b_3 \to \infty} F(a, b].$$

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

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

also, if the interval $\{(x_1, \dots, x_n)\colon a_i < x_i \le b_i\colon i = 1, \dots, n\}$ has some $b_i = \infty$, the corresponding interval in $\mathbb{R}^n$ 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 $\mathbb{R}$, 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 $\mathbb{R}^n$ 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 $\mathscr{B}(\mathbb{R})$. In this case we have $F(x_1, \dots, x_n) = x_1x_2\cdots x_n$ 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 $\mathscr{B}(\mathbb{R}^n)$. Just as in one dimension, the completion of $\mathscr{B}(\mathbb{R}^n)$ relative to Lebesgue measure is called the class of *Lebesgue measurable sets* in R^n, written $\overline{\mathscr{B}}(\mathbb{R}^n)$.

(b) Let f be a nonnegative function from $\mathbb{R}^n$ to $\mathbb{R}$ 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

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

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

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.

1.4.11 Theorem. If μ is a σ-finite measure on $\mathscr{B}(\mathbb{R}^n)$, then for each $B \in \mathscr{B}(\mathbb{R}^n)$,

(a) $\mu(B) = \sup\{\mu(K)\colon K \subset B, K \text{ compact}\}$.

If μ is in fact a Lebesgue–Stieltjes measure, then

(b) $\mu(B) = \inf\{\mu(V)\colon V \supset B, B \text{ open}\}$.

(c) There is an example of a σ-finite measure on $\mathscr{B}(\mathbb{R}^n)$ that is not a Lebesgue–Stieltjes measure and for which (b) fails.

PROOF.

(a) First assume that μ is finite. Let $\mathscr{C}$ be the class of subsets of $\mathbb{R}^n$ having the desired property; we show that $\mathscr{C}$ is a monotone class. Indeed, let $B_n \in \mathscr{C}, B_n \uparrow B$. Let K_n be a compact subset of B_n with $\mu(B_n) \le \mu(K_n) + \varepsilon, \varepsilon > 0$ preassigned. By replacing K_n by $\bigcup_{i=1}^n K_i$, we may assume the K_n form an increasing sequence. Then $\mu(B) = \lim_{n\to\infty} \mu(B_n) \le \lim_{n\to\infty} \mu(K_n) + \varepsilon$, so that

$$\mu(B) = \sup\{\mu(K)\colon K \text{ a compact subset of } B\},$$

and $B \in \mathscr{C}$. If $B_n \in \mathscr{C}$, $B_n \downarrow B$, let K_n be a compact subset of B_n such that $\mu(B_n) \le \mu(K_n) + \varepsilon 2^{-n}$, and set $K = \bigcap_{n=1}^{\infty} K_n$. Then

$$\mu(B) - \mu(K) = \mu(B - K) \le \mu\left(\bigcup_{n=1}^{\infty}(B_n - K_n)\right) \le \sum_{n=1}^{\infty} \mu(B_n - K_n) \le \varepsilon;$$

thus $B \in \mathscr{C}$. Therefore $\mathscr{C}$ 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 $\mathscr{C}$ contains all Borel sets.

If μ is σ-finite, each $B \in \mathscr{B}(\mathbb{R}^n)$ 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 $\mu_i(A) = \mu(A \cap B_i)$, $A \in \mathscr{B}(\mathbb{R}^n)$], and the preceding argument that $\mathscr{C}$ is closed under limits of increasing sequences shows that $B \in \mathscr{C}$.

(b) We have $\mu(B) \le \inf\{\mu(V)\colon\ V \supset B,\quad V \text{ open}\}$

$$\le \inf\{\mu(W)\colon\ W \supset B,\quad W = K^c,\quad K \text{ compact}\}.$$

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 $\mathbb{R}^n = \bigcup_{k=1}^{\infty} B_k$, where the B_k are disjoint bounded sets; then $B_k \subset C_k$ for some bounded open set C_k. The measure $\mu_k(A) = \mu(A \cap C_k)$, $A \in \mathscr{B}(\mathbb{R}^n)$, 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 $\mu_k(W_k) \le \mu_k(B) + \varepsilon 2^{-k}$. 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 $\mu(V_k) \le \mu(B) + \varepsilon 2^{-k}$. For any $A \in \mathscr{B}(\mathbb{R}^n)$, let V_k be an open set with $V_k \supset A \cap B_k$ and $\mu(V_k) \le \mu(A \cap B_k) + \varepsilon 2^{-k}$. Then $V = \bigcup_{k=1}^{\infty} V_k$ is open, $V \supset A$, and $\mu(V) \le \sum_{k=1}^{\infty} \mu(V_k) \le \mu(A) + \varepsilon$.

(c) Construct a measure μ on $\mathscr{B}(\mathbb{R})$ as follows. Let μ be concentrated on $S = \{1/n\colon\ n = 1, 2, \ldots\}$ and take $\mu\{1/n\} = 1/n$ for all n. Since $\mathbb{R} = \bigcup_{n=1}^{\infty}\{1/n\} \cup S^c$ 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

$$\begin{aligned} \mu(V) &\ge \mu(-\varepsilon, \varepsilon) && \text{for some } \varepsilon \\ &\ge \sum_{k=r}^{\infty} \frac{1}{k} && \text{for some } r \\ &= \infty. \end{aligned}$$

Thus (b) fails. [Another example: Let $\mu(A)$ be the number of rational points in A.]

Problems

1. Let F be the distribution function on $\mathbb{R}$ given by $F(x) = 0$, $x < -1$; $F(x) = 1 + x$, $-1 \le x < 0$; $F(x) = 2 + x^2$, $0 \le x < 2$; $F(x) = 9$, $x \ge 2$. If μ is the Lebesgue–Stieltjes measure corresponding to F, compute the measure of each of the following sets:

   (a) $\{2\}$,
   (b) $\left[-\frac{1}{2}, 3\right)$,
   (c) $(-1, 0] \cup (1, 2)$,
   (d) $\left[0, \frac{1}{2}\right) \cup (1, 2]$,
   (e) $\{x\colon\ |x| + 2x^2 > 1\}$.

2. Let μ be a Lebesgue–Stieltjes measure on $\mathbb{R}$ corresponding to a continuous distribution function.

   (a) If A is a countable subset of $\mathbb{R}$, show that $\mu(A) = 0$.
   (b) If $\mu(A) > 0$, must A include an open interval?
   (c) If $\mu(A) > 0$ and $\mu(\mathbb{R} - A) = 0$, must A be dense in $\mathbb{R}$?
   (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 $\mathbb{R}^n$ and $a \in \mathbb{R}^n$, show that $a + B = \{a + x\colon\ x \in B\}$ is a Borel set, and $-B = \{-x\colon\ x \in B\}$ is a Borel set. (Use the good sets principle.)

4. Show that if $B \in \overline{\mathscr{B}}(\mathbb{R}^n)$, $a \in \mathbb{R}^n$, then $a + B \in \overline{\mathscr{B}}(\mathbb{R}^n)$ 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 $\mathscr{B}(\mathbb{R}^n)$ 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 $B_x = \{y\colon\ y - x \text{ rational}\}$ to form a set A (this requires the axiom of choice). 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 $\mathbb{R} = \bigcup\{r + A\colon\ r \text{ rational}\}$.
   (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 $\mathbb{R}$ 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, $E_1 = \left(\frac{1}{3}, \frac{2}{3}\right)$; 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, $E_2 = \left(\frac{1}{9}, \frac{2}{9}\right) \cup \left(\frac{7}{9}, \frac{8}{9}\right)$; 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 $E_1, \ldots, E_{n-1}$ 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 $m(E_n) = \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{n-1}$, $n = 1, 2, \ldots$, where m is Lebesgue measure. If $0 < \alpha < 1$, the procedure may be altered slightly so that $m(E_n) = \alpha\left(\frac{1}{2}\right)^n$. 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: $\mathbb{R}^2 \to \mathbb{R}$ such that F is right-continuous and is increasing in each coordinate separately, but F is not a distribution function on $\mathbb{R}^2$.

9. A distribution function on $\mathbb{R}$ is monotone and thus has only countably many points of discontinuity. Is this also true for a distribution function on $\mathbb{R}^n$, $n > 1$?

10. (a) Let F and G be distribution functions on $\mathbb{R}^n$. If $F(a, b] = G(a, b]$ for all $a, b \in \mathbb{R}^n$, $a \le b$, does it follow that F and G differ by a constant?
    (b) Must a distribution function on $\mathbb{R}^n$ be increasing in each coordinate separately?

*11. If c is the cardinality of the reals, show that there are only c Borel subsets of $\mathbb{R}^n$, but 2^c Lebesgue measurable sets.

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 $(\Omega, \mathscr{F}, P)$ is a probability space, and that h is a function from Ω to $\overline{\mathbb{R}}$. 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 $a \le h(\omega) \le b$, in other words, we wish to compute $P\{\omega\colon h(\omega) \in B\}$ where $B = [a, b]$. For this to be possible, the set $\{\omega\colon h(\omega) \in B\} = h^{-1}(B)$ must belong to the σ-field $\mathscr{F}$. If $h^{-1}(B) \in \mathscr{F}$ 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 $h\colon \Omega_1 \to \Omega_2$, h is *measurable* relative to the σ-fields $\mathscr{F}_j$ of subsets of Ω_j, $j = 1, 2$, iff $h^{-1}(A) \in \mathscr{F}_1$ for each $A \in \mathscr{F}_2$.

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

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

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

Notice that measurability of h does not imply that $h(A) \in \mathscr{F}_2$ for each $A \in \mathscr{F}_1$. For example, if $\mathscr{F}_2 = \{\emptyset, \Omega_2\}$, then any $h\colon \Omega_1 \to \Omega_2$ is measurable, regardless of $\mathscr{F}_1$, but if $A \in \mathscr{F}_1$ and $h(A)$ is a nonempty proper subset of Ω_2, then $h(A) \notin \mathscr{F}_2$. 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 $h^{-1}\left(\bigcup_i B_i\right) = \bigcup_i h^{-1}(B_i)$, $h^{-1}\left(\bigcap_i B_i\right) = \bigcap_i h^{-1}(B_i)$, and $h^{-1}(B^c) = [h^{-1}(B)]^c$. We also have $h\left(\bigcup_i B_i\right) = \bigcup_i h(A_i)$, but $h\left(\bigcap_i A_i\right) \subset \bigcap_i h(A_i)$, 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 $(\Omega, \mathscr{F})$ is a measurable space and $h\colon \Omega \to \mathbb{R}^n$ (or $\overline{\mathbb{R}}^n$), h is said to be *Borel measurable* [on $(\Omega, \mathscr{F})$] iff h is measurable relative to the σ-fields $\mathscr{F}$ and $\mathscr{B}$, the class of Borel sets. If Ω is a Borel subset of $\mathbb{R}^k$ (or $\overline{\mathbb{R}}^k$) and we use the term "Borel measurable," we always assume that $\mathscr{F} = \mathscr{B}$.

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

If A is a subset of $\mathbb{R}$ 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 $\{\omega\colon I_A(\omega) = 1\} = A \notin \mathscr{B}(\mathbb{R})$.

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

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

1.5.2 Definition. Let $(\Omega, \mathscr{F})$ be a measurable space, fixed throughout the discussion. If $h\colon \Omega \to \overline{\mathbb{R}}$, 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 $\mathscr{F}$ and I_{A_i} is the indicator of A_i; the x_i need not be distinct.

We assume the standard arithmetic of $\overline{\mathbb{R}}$; if $a \in \mathbb{R}$, $a + \infty = \infty$, $a - \infty = -\infty$, $a/\infty = a/-\infty = 0$, $a \cdot \infty = \infty$ 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 $\mathscr{F}$, again fixed throughout the discussion. If $h\colon \Omega \to \overline{\mathbb{R}}$ 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 $\mathscr{F}$. 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

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

If h is nonnegative Borel measurable, define

$$\int_\Omega h\,d\mu = \sup\left\{\int_\Omega s\,d\mu\colon\ s \quad \text{simple,} \quad 0 \le s \le h\right\}.$$

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,

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

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, $\{\omega\colon h^+(\omega) \in B\} = \{\omega\colon h(\omega) \ge 0, h(\omega) \in B\} \cup \{\omega\colon h(\omega) < 0, 0 \in B\}$. The first set is $h^{-1}[0, \infty] \cap h^{-1}(B) \in \mathscr{F}$; the second is $h^{-1}[-\infty, 0)$ if $0 \in B$, and $\emptyset$ if $0 \notin B$. Thus $(h^+)^{-1}(B) \in \mathscr{F}$ for each $B \in \mathscr{B}(\overline{\mathbb{R}})$, 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

$$\{\omega\colon\ \max(h_1(\omega), h_2(\omega)) \le c\} = \{\omega\colon h_1(\omega) \le c\} \cap \{\omega\colon\ h_2(\omega) \le c\}$$

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

We define

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

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 $A \in \mathscr{F}$, we define

$$\int_A h\,d\mu = \int_\Omega hI_A\,d\mu.$$

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

If h is a step function from $\mathbb{R}$ to $\mathbb{R}$ and μ is Lebesgue measure, $\int_\mathbb{R} h\,d\mu$ 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 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 $h_1, h_2, \ldots$ are Borel measurable functions from Ω to $\overline{\mathbb{R}}$ and $h_n(\omega) \to h(\omega)$ for all $\omega \in \Omega$, then h is Borel measurable.

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

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

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

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

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 $\overline{\mathbb{R}}$, 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 $s_{1n} \to h_1$, $s_{2n} \to h_2$. Then $s_{1n} + s_{2n} \to h_1 + h_2$,

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

and

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

Since

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

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 $\overline{\mathbb{R}}^n$; to do this, we need the following useful result.

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

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

Since some books contain the statement "A composition of measurable functions need not be measurable," some explanation is called for. If h: $\mathbb{R} \to \mathbb{R}$, 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. Let $\mathscr{B}$ be the Borel sets, and $\overline{\mathscr{B}}$ the Lebesgue measurable sets. If $B \in \mathscr{B}$ then $h^{-1}(B) \in \overline{\mathscr{B}}$; but $g^{-1}(h^{-1}(B))$ is known to belong to $\overline{\mathscr{B}}$ only when $h^{-1}(B) \in \mathscr{B}$, so we cannot conclude that $(h \circ g)^{-1}(B) \in \overline{\mathscr{B}}$. For an explicit example, see Royden (1968, p. 70). If $g^{-1}(A) \in \overline{\mathscr{B}}$ for all $A \in \overline{\mathscr{B}}$, not just for all $A \in \mathscr{B}$, 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 $\mathbb{R}$ to $\mathbb{R}$ (or $\mathbb{R}^k$ to $\mathbb{R}^n$) by the slightly more general concept of Lebesgue measurability; in this book, the only instance is in 1.7. The integration theory that we are developing works for extended real-valued functions on an arbitrary measure space $(\Omega, \mathscr{F}, \mu)$. Thus there is no problem in integrating Lebesgue measurable functions; $\Omega = \mathbb{R}, \mathscr{F} = \overline{\mathscr{B}}$.

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

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

PROOF. Assume h Borel measurable. Since

$$p_i^{-1}\{x_i\colon a_i \le x_i \le b_i\} = \{x \in \overline{\mathbb{R}}^n\colon a_i \le x_i \le b_i, \quad -\infty \le x_j \le \infty, \quad j \ne i\},$$

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

$$h\colon (\Omega, \mathscr{F}) \to (\overline{\mathbb{R}}^n, \mathscr{B}(\overline{\mathbb{R}}^n)), \qquad p_i\colon (\overline{\mathbb{R}}^n, \mathscr{B}(\overline{\mathbb{R}}^n)) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}})),$$

and therefore by 1.5.7, h_i: $(\Omega, \mathscr{F}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$.

Conversely, assume each h_i to be Borel measurable. Then

$$h^{-1}\{x \in \overline{\mathbb{R}}^n\colon\ a_i \le x_i \le b_i,\ i = 1, \ldots, n\}$$
$$= \bigcap_{i=1}^{n} \{\omega \in \Omega\colon\ a_i \le h_i(\omega) \le b_i\} \in \mathscr{F},$$

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 $\overline{\mathbb{R}}$.

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

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

(c) If $\int_\Omega h\,d\mu$ exists, then $|\int_\Omega h\,d\mu| \le \int_\Omega |h|\,d\mu$.

(d) If $h \ge 0$ and $B \in \mathscr{F}$, then $\int_B h\,d\mu = \sup\{\int_B s\,d\mu\colon\ 0 \le s \le h, s \text{ simple}\}$.

(e) If $\int_\Omega h\,d\mu$ exists, so does $\int_A h\,d\mu$ for each $A \in \mathscr{F}$; if $\int_\Omega h\,d\mu$ is finite, then $\int_A h\,d\mu$ is also finite for each $A \in \mathscr{F}$.

Proof. (a) It is immediate that this holds when h is simple. If h is nonnegative and $c > 0$, then

$$\int_\Omega ch\,d\mu = \sup\left\{\int_\Omega s\,d\mu;\qquad 0 \le s \le ch,\quad s\quad \text{simple}\right\}$$
$$= c\sup\left\{\int_\Omega \frac{s}{c}\,d\mu;\qquad 0 \le \frac{s}{c} \le h,\qquad \frac{s}{c}\ \text{simple}\right\} = c\int_\Omega h\,d\mu.$$

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

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

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 $0 \le s \le h$, s simple, then $0 \le s \le g$; hence $\int_\Omega h\,d\mu \le \int_\Omega g\,d\mu$. In general, $h \le g$ implies $h^+ \le g^+$, $h^- \ge g^-$. If

$\int_\Omega h\,d\mu > -\infty$, we have $\int_\Omega g^-\,d\mu \le \int_\Omega h^-\,d\mu < \infty$; hence $\int_\Omega g\,d\mu$ exists and equals

$$\int_\Omega g^+\,d\mu - \int_\Omega g^-\,d\mu \ge \int_\Omega h^+\,d\mu - \int_\Omega h^-\,d\mu = \int_\Omega h\,d\mu.$$

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

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

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

$$\int_B h\,d\mu \ge \sup\left\{\int_B s\,d\mu\colon\ 0 \le s \le h\right\}.$$

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

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

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 $(\Omega, \mathscr{F})$, and define

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

where A is a set in $\mathscr{F}$. Show that h is Borel measurable.
3. If $f_1, f_2, \ldots$ are extended real-valued Borel measurable functions on $(\Omega, \mathscr{F})$, $n = 1, 2, \ldots,$ show that $\sup_n f_n$ and $\inf_n f_n$ are Borel measurable (hence $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are Borel measurable).
4. Let $(\Omega, \mathscr{F}, \mu)$ be a complete measure space. If $f\colon (\Omega, \mathscr{F}) \to (\Omega', \mathscr{F}')$ and $g\colon \Omega \to \Omega'$, $g = f$ except on a subset of a set $A \in \mathscr{F}$ with $\mu(A) = 0$, show that g is measurable (relative to $\mathscr{F}$ and $\mathscr{F}'$).

*5. (a) Let f be a function from $\mathbb{R}^k$ to $\mathbb{R}^m$, not necessarily Borel measurable. Show that $\{x\colon\ f \text{ is discontinuous at } x\}$ is an F_σ (a countable union of closed subsets of $\mathbb{R}^k$), and hence is a Borel set. Does this result hold in spaces more general than the Euclidean space $\mathbb{R}^n$?

(b) Show that there is no function from $\mathbb{R}$ to $\mathbb{R}$ 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 $\mathbb{R}^n$ to $\mathbb{R}^k$?

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) $\sum_{k=1}^{\infty} |a_{nk}| \le c < \infty$ for all n, and
(iii) $a_{nk} \to 0$ as $n \to \infty$ 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 $\sum_{k \le k_r} |a_{n_{r+1}}k| < \frac{1}{8}$; 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 $x_k = 0$, $k_{2s-1} < k \le k_{2s}$, $x_k = 1$, $k_{2s} < k \le k_{2s+1}$, $s = 1, 2, \dots$. 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 $k \le k_r$, h_2 corresponds to $k_r < k \le k_{r+1}$, and h_3 to $k > k_{r+1}$. If r is odd, then $x_k = 0$, $k_r < k \le k_{r+1}$; hence $|t_{n_{r+1}}| < \frac{1}{4}$. If r is even, then $h_2 = \sum_{k_r<k\le k_{r+1}} a_{n_{r+1}k}$; hence by (i),

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

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 $(\Omega, \mathscr{F})$ be a measurable space, and let P_n, $n = 1, 2, \dots$, be probability measures on $\mathscr{F}$. If $P_n(A) \to P(A)$ for all $A \in \mathscr{F}$, then P is a probability measure on $\mathscr{F}$; furthermore, if $\{B_k\}$ is a sequence of sets in $\mathscr{F}$ decreasing to $\emptyset$, then $\sup_n P_n(B_k) \downarrow 0$ as $k \to \infty$. [Let A be the disjoint union of sets $A_k \in \mathscr{F}$; 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, $\alpha = \sum_k P(A_k) \le P(\Omega) = 1$. 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 $P_n(\Omega) \le c < \infty$ 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, $(\Omega, \mathscr{F}, \mu)$ is a fixed measure space, and all functions to be considered map Ω to $\overline{\mathbb{R}}$.

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

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

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

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

Now $B_n \subset B$, hence $I_{B_n} \le I_B$, so by 1.5.9(b), $\lambda(B_n) \le \lambda(B)$. 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 1.59(b), (d) and the fact that the maximum of a finite number of simple functions is simple that we can find a simple function s, $0 \le s \le h$, such that

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

Now

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

by what we have proved for nonnegative simple functions, hence

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

Since $\lambda(B) \geq \lambda\left(\bigcup_{i=1}^{n} B_i\right)$ 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 $\lambda^+(B) = \int_B h^+ \, d\mu$, $\lambda^- = \int_B h^- \, d\mu$; at least one of the measures λ^+ and λ^- must be finite.

1.6.2 Monotone Convergence Theorem. Let $h_1, h_2, \ldots$ form an increasing sequence of nonnegative Borel measurable functions, and let $h(\omega) = \lim_{n\to\infty} h_n(\omega)$, $\omega \in \Omega$. Then $\int_\Omega h_n \, d\mu \to \int_\Omega h \, d\mu$. [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 $k = \lim_{n\to\infty} \int_\Omega h_n \, d\mu \leq \int_\Omega h \, d\mu$. Let $0 < b < 1$, and let s be a nonnegative, finite-valued, simple function with $s \leq h$. Let $B_n = \{\omega\colon\ h_n(\omega) \geq bs(\omega)\}$. 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, $\int_{B_n} s \, d\mu \to \int_\Omega s \, d\mu$, hence (let $b \to 1$) $k \geq \int_\Omega s \, d\mu$. 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 $0 \le s_n = t_n + u_n \uparrow f + g$. 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 $f \ge 0$, $g \le 0$, $h = f + g \ge 0$ (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 $h \ge 0$,

$$\int_\Omega f\,d\mu \ge -\int_\Omega g\,d\mu = \infty,$$

contradicting the hypothesis that $\int_\Omega f\,d\mu + \int_\Omega g\,d\mu$ is well-defined. Similarly, if $f \ge 0$, $g \le 0$, $h \le 0$, 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, $-g \ge 0$, $-f \le 0$, $-h = -f - g \ge 0$, and the above argument applies.)

Let

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

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 $\int_\Omega f^+\,d\mu \ge \int_{E_i} f^+\,d\mu$.) Similarly $\int_\Omega f\,d\mu$ or $\int_\Omega g\,d\mu = -\infty$, and this is a contradiction. $\square$

1.6.4 Corollaries. (a) If $h_1, h_2, \ldots$ 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 $|g| \le h$, 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 $B \in \mathscr{F}$ 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 $x_i \ne 0$ implies $\mu(A_i) = 0$ by hypothesis, hence $\int_\Omega f \, d\mu = 0$. If $f \ge 0$ and $0 \le s \le f$, 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 $A = \{\omega\colon\ g(\omega) = h(\omega)\}$, $B = A^c$. 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 $\int_\Omega h_n \, d\mu \to \int_\Omega h \, d\mu$.

Another example: If g and h are Borel measurable and $g \ge h$ a.e., then $\int_\Omega g \, d\mu \ge \int_\Omega h \, d\mu$ [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 $A = \{\omega\colon |h(\omega)| = \infty\}$. If $\mu(A) > 0$, then $\int_\Omega |h|\,d\mu \geq \int_A |h|\,d\mu = \infty\mu(A) = \infty$, a contradiction.

(b) Let $B = \{\omega\colon h(\omega) > 0\}$, $B_n = \{\omega\colon h(\omega) \geq 1/n\} \uparrow B$. 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 $g_1, g_2, \ldots, g, h$ 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

$$\int_\Omega g_n\,d\mu \downarrow \int_\Omega g\,d\mu.$$

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 $f_1, f_2, \ldots$ are functions from Ω to $\overline{R}$, $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ are defined pointwise, that is,

$$\left(\liminf_{n\to\infty} f_n\right)(\omega) = \sup_n \inf_{k\geq n} f_k(\omega),$$

$$\left(\limsup_{n\to\infty} f_n\right)(\omega) = \inf_n \sup_{k\geq n} f_k(\omega).$$

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

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

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

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

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

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, $\int_\Omega g_n\, d\mu \uparrow \int_\Omega (\liminf_{n\to\infty} f_n)\, d\mu$. But $g_n \leq f_n$, so

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

(b) We may write

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

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 $f_1, f_2, \ldots, f, g$ are Borel measurable, $|f_n| \leq g$ for all n, where g is μ-integrable, and $f_n \to f$ a.e. $[\mu]$, then f is μ-integrable and $\int_\Omega f_n\, d\mu \to \int_\Omega f\, d\mu$.

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

$$\int_\Omega \left(\liminf_{n\to\infty} f_n\right) d\mu \le \liminf_{n\to\infty} \int_\Omega f_n \, d\mu \le \limsup_{n\to\infty} \int_\Omega f_n \, d\mu$$
$$\le \int_\Omega \left(\limsup_{n\to\infty} f_n\right) d\mu.$$

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

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

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

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

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

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

$$A_n = \left\{\omega\colon\ g(\omega) \ge h(\omega) + \frac{1}{n}, \qquad |h(\omega)| \le n\right\}.$$

Then

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

But

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

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

$\{\omega\colon h(\omega)=\infty\}$, and by taking $C_n=\{\omega\colon h(\omega)=-\infty, g(\omega)\geq -n\}$ 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

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

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

If g and h are integrable, the proof is simpler. Let $B=\{\omega\colon g(\omega)>h(\omega)\}$. 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 $T\colon (\Omega,\mathscr{F})\to(\Omega_0,\mathscr{F}_0)$ be a measurable mapping, and let μ be a measure on $\mathscr{F}$. Define a measure $\mu_0=\mu T^{-1}$ on $\mathscr{F}_0$ by

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

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

If $f\colon (\Omega_0,\mathscr{F}_0)\to(\overline{R},\mathscr{B}(\overline{R}))$ and $A\in\mathscr{F}_0$, 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

$$\int_{T^{-1}A} f(T(\omega))\,d\mu(\omega) = \sum_{i=1}^n x_i \int_{T^{-1}A} I_{B_i}(T(\omega))\,d\mu(\omega) \qquad \text{by 1.6.3}$$

$$= \sum_{i=1}^n x_i \int_A I_{B_i}(\omega)\,d\mu_0(\omega)$$

by what we have proved for indicators

$$= \int_A f(\omega)\,d\mu_0(\omega) \qquad \text{by 1.6.3.}$$

If f is a non-negative 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 $\int_a^b f(x)\,dx$ (or $\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \cdots, x_n)dx_1 \cdots dx_n$ if we are integrating functions on $\mathbb{R}^n$). 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: $(a, b) \to \mathbb{R}$ such that $|f(x, y)| \le g(y)$ for all x, y, and $\int_a^b g(y)dy < \infty$. If $x_0 \in (c, d)$ and $\lim_{x\to x_0} f(x, y)$ exists for all $y \in (a, b)$, show that

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

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

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

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

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: $(a, b) \to \mathbb{R}$ such that $|f_1(x, y)| \le h(y)$ 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 $(\Omega, \mathscr{F})$ and $A_1, A_2, \dots$ is a sequence of sets in $\mathscr{F}$, use Fatou's lemma to show that

$$\mu\left(\liminf_n A_n\right) \le \liminf_{n\to\infty} \mu(A_n).$$

   If μ is finite, show that

$$\mu\left(\limsup_n A_n\right) \ge \limsup_{n\to\infty} \mu(A_n).$$

   Thus if μ is finite and $A = \lim_n A_n$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$. (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

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

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

6. (a) Show that $\int_1^\infty e^{-t} \ln t\, dt = \lim_{n\to\infty} \int_1^n [1 - (t/n)]^n \ln t\, dt$.

   (b) Show that $\int_0^1 e^{-t} \ln t\, dt = \lim_{n\to\infty} \int_0^1 [1 - (t/n)]^n \ln t\, dt$.

7. If $(\Omega, \mathscr{F}, \mu)$ is the completion of $(\Omega, \mathscr{F}_0, \mu)$ and f is a Borel measurable function on $(\Omega, \mathscr{F})$, show that there is a Borel measurable function g on $(\Omega, \mathscr{F}_0)$ such that $f = g$, except on a subset of a set in $\mathscr{F}_0$ of measure 0. (Start with indicators.)

8. If f is a Borel measurable function from $\mathbb{R}$ to $\mathbb{R}$ and $a \in \mathbb{R}$, 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 $[a, b]$, assumed fixed throughout the discussion. If P: $a = x_0 < x_1 < \cdots < x_n = b$ is a partition of $[a, b]$, we may construct the upper and lower sums of f relative to P as follows.

Let

$$M_i = \sup\{f(y)\colon\ x_{i-1} < y \le x_i\}, \qquad i = 1, \ldots, n,$$
$$m_i = \inf\{f(y)\colon\ x_{i-1} < y \le x_i\}, \qquad i = 1, \ldots, n,$$

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

$$\alpha(x) = M_i \quad \text{if} \quad x_{i-1} < x \le x_i, \qquad i = 1, \ldots, n,$$
$$\beta(x) = m_i \quad \text{if} \quad x_{i-1} < x \le x_i, \qquad i = 1, \ldots, n$$

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

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

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

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

$$\mathrm{U}(P) = \int_a^b \alpha\, d\mu, \qquad \mathrm{L}(P) = \int_a^b \beta\, d\mu.$$

Now let $P_1, P_2, \ldots$ 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 $k \to \infty$. 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,

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

and

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

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

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

This follows by a standard ε–δ argument.

If $\lim_{k\to\infty} \mathrm{U}(P_k) = \lim_{k\to\infty} \mathrm{L}(P_k) =$ 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,

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

As $\beta \le f \le \alpha$, 1.6.6(b) applied to $\alpha - \beta$ yields $\alpha = f = \beta$ a.e.; hence f is continuous a.e. Conversely, assume f is 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

$$\int_a^b \alpha \, d\mu = \int_a^b \beta \, d\mu = \int_a^b f \, d\mu, \tag{1}$$

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 $\mathbb{R}^n$; 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

1. 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.

2. 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 1.7, and take

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

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

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

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

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

   where $\alpha = \lim_{k\to\infty} \alpha_k$, $\beta = \lim_{k\to\infty} \beta_k$. If $\mathrm{U}(P_k)$ and $\mathrm{L}(P_k)$ 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 $[a, b]$.

(a) Show that f is Riemann–Stieltjes integrable iff f is continuous a.e. $[\mu]$ on $[a, b]$.

(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: $\mathbb{R} \to \mathbb{R}$, the improper Riemann integral of f may be defined as

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

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 $\mathbb{R}$, 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, $f_n \to f$ everywhere, but f is not Riemann integrable.

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