Measure Theory

John K. Hunter

Department of Mathematics, University of California at Davis

Abstract. These are some brief notes on measure theory, concentrating on Lebesgue measure on $\mathbb{R}^n$. Some missing topics I would have liked to have included had time permitted are: the change of variable formula for the Lebesgue integral on $\mathbb{R}^n$; absolutely continuous functions and functions of bounded variation of a single variable and their connection with Lebesgue-Stieltjes measures on $\mathbb{R}$; Radon measures on $\mathbb{R}^n$, and other locally compact Hausdorff topological spaces, and the Riesz representation theorem for bounded linear functionals on spaces of continuous functions; and other examples of measures, including k-dimensional Hausdorff measure in $\mathbb{R}^n$, Wiener measure and Brownian motion, and Haar measure on topological groups. All these topics can be found in the references.

Contents

CHAPTER 1

Measures

Measures are a generalization of volume; the fundamental example is Lebesgue measure on $\mathbb{R}^n$, which we discuss in detail in the next Chapter. Moreover, as formalized by Kolmogorov (1933), measure theory provides the foundation of probability. Measures are important not only because of their intrinsic geometrical and probabilistic significance, but because they allow us to define integrals.

This connection, in fact, goes in both directions: we can define an integral in terms of a measure; or, in the Daniell-Stone approach, we can start with an integral (a linear functional acting on functions) and use it to define a measure. In probability theory, this corresponds to taking the expectation of random variables as the fundamental concept from which the probability of events is derived.

In these notes, we develop the theory of measures first, and then define integrals. This is (arguably) the more concrete and natural approach; it is also (unarguably) the original approach of Lebesgue. We begin, in this Chapter, with some preliminary definitions and terminology related to measures on arbitrary sets. See Folland [**4**] for further discussion.

1.1. Sets

We use standard definitions and notations from set theory and will assume the axiom of choice when needed. The words 'collection' and 'family' are synonymous with 'set' — we use them when talking about sets of sets. We denote the collection of subsets, or power set, of a set X by $\mathcal{P}(X)$. The notation 2^X is also used.

If $E \subset X$ and the set X is understood, we denote the complement of E in X by $E^c = X \setminus E$. De Morgan's laws state that

$$\left(\bigcup_{\alpha\in I} E_\alpha\right)^c = \bigcap_{\alpha\in I} E_\alpha^c, \qquad \left(\bigcap_{\alpha\in I} E_\alpha\right)^c = \bigcup_{\alpha\in I}^{\infty} E_\alpha^c.$$

We say that a collection

$$\mathcal{C} = \{E_\alpha \subset X : \alpha \in I\}$$

of subsets of a set X, indexed by a set I, covers $E \subset X$ if

$$\bigcup_{\alpha\in I} E_\alpha \supset E.$$

The collection $\mathcal{C}$ is disjoint if $E_\alpha \cap E_\beta = \varnothing$ for $\alpha \neq \beta$.

The Cartesian product, or product, of sets X, Y is the collection of all ordered pairs

$$X \times Y = \{(x, y) : x \in X,\, y \in Y\}.$$

1.2. Topological spaces

A topological space is a set equipped with a collection of open subsets that satisfies appropriate conditions.

Definition 1.1. A topological space $(X, \mathcal{T})$ is a set X and a collection $\mathcal{T} \subset \mathcal{P}(X)$ of subsets of X, called open sets, such that

(a) $\varnothing, X \in \mathcal{T}$;
(b) if $\{U_\alpha \in \mathcal{T} : \alpha \in I\}$ is an arbitrary collection of open sets, then their union

$$\bigcup_{\alpha \in I} U_\alpha \in \mathcal{T}$$

is open;
(c) if $\{U_i \in \mathcal{T} : i = 1, 2, \dots, N\}$ is a finite collection of open sets, then their intersection

$$\bigcap_{i=1}^{N} U_i \in \mathcal{T}$$

is open.

The complement of an open set in X is called a closed set, and $\mathcal{T}$ is called a topology on X.

1.3. Extended real numbers

It is convenient to use the extended real numbers

$$\overline{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}.$$

This allows us, for example, to talk about sets with infinite measure or non-negative functions with infinite integral. The extended real numbers are totally ordered in the obvious way: ∞ is the largest element, $-\infty$ is the smallest element, and real numbers are ordered as in $\mathbb{R}$. Algebraic operations on $\overline{\mathbb{R}}$ are defined when they are unambiguous *e.g.* $\infty + x = \infty$ for every $x \in \overline{\mathbb{R}}$ except $x = -\infty$, but $\infty - \infty$ is undefined.

We define a topology on $\overline{\mathbb{R}}$ in a natural way, making $\overline{\mathbb{R}}$ homeomorphic to a compact interval. For example, the function $\phi : \overline{\mathbb{R}} \to [-1, 1]$ defined by

$$\phi(x) = \begin{cases} 1 & \text{if } x = \infty \\ x/\sqrt{1+x^2} & \text{if } -\infty < x < \infty \\ -1 & \text{if } x = -\infty \end{cases}$$

is a homeomorphism.

A primary reason to use the extended real numbers is that upper and lower bounds always exist. Every subset of $\overline{\mathbb{R}}$ has a supremum (equal to ∞ if the subset contains ∞ or is not bounded from above in $\mathbb{R}$) and infimum (equal to $-\infty$ if the subset contains $-\infty$ or is not bounded from below in $\mathbb{R}$). Every increasing sequence of extended real numbers converges to its supremum, and every decreasing sequence converges to its infimum. Similarly, if $\{a_n\}$ is a sequence of extended real-numbers then

$$\limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}} \left(\sup_{i\geq n} a_i \right), \qquad \liminf_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} \left(\inf_{i\geq n} a_i \right)$$

both exist as extended real numbers.

Every sum $\sum_{i=1}^{\infty} x_i$ with non-negative terms $x_i \geq 0$ converges in $\overline{\mathbb{R}}$ (to ∞ if $x_i = \infty$ for some $i \in \mathbb{N}$ or the series diverges in $\mathbb{R}$), where the sum is defined by

$$\sum_{i=1}^{\infty} x_i = \sup\left\{\sum_{i\in F} x_i : F \subset \mathbb{N} \text{ is finite}\right\}.$$

As for non-negative sums of real numbers, non-negative sums of extended real numbers are unconditionally convergent (the order of the terms does not matter); we can rearrange sums of non-negative extended real numbers

$$\sum_{i=1}^{\infty}(x_i + y_i) = \sum_{i=1}^{\infty} x_i + \sum_{i=1}^{\infty} y_i;$$

and double sums may be evaluated as iterated single sums

$$\begin{aligned}\sum_{i,j=1}^{\infty} x_{ij} &= \sup\left\{\sum_{(i,j)\in F} x_{ij} : F \subset \mathbb{N}\times\mathbb{N} \text{ is finite}\right\}\\ &= \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} x_{ij}\right)\\ &= \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} x_{ij}\right).\end{aligned}$$

Our use of extended real numbers is closely tied to the order and monotonicity properties of $\mathbb{R}$. In dealing with complex numbers or elements of a vector space, we will always require that they are strictly finite.

1.4. Outer measures

As stated in the following definition, an outer measure is a monotone, countably subadditive, non-negative, extended real-valued function defined on all subsets of a set.

Definition 1.2. An outer measure μ^* on a set X is a function

$$\mu^* : \mathcal{P}(X) \to [0, \infty]$$

such that:

(a) $\mu^*(\varnothing) = 0$;
(b) if $E \subset F \subset X$, then $\mu^*(E) \leq \mu^*(F)$;
(c) if $\{E_i \subset X : i \in \mathbb{N}\}$ is a countable collection of subsets of X, then

$$\mu^*\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mu^*(E_i).$$

We obtain a statement about finite unions from a statement about infinite unions by taking all but finitely many sets in the union equal to the empty set. Note that μ^* is not assumed to be additive even if the collection $\{E_i\}$ is disjoint.

1.5. σ-algebras

A σ-algebra on a set X is a collection of subsets of a set X that contains $\varnothing$ and X, and is closed under complements, finite unions, countable unions, and countable intersections.

Definition 1.3. A σ-algebra on a set X is a collection $\mathcal{A}$ of subsets of X such that:

(a) $\varnothing, X \in \mathcal{A}$;
(b) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$;
(c) if $A_i \in \mathcal{A}$ for $i \in \mathbb{N}$ then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}, \qquad \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$

From de Morgan's laws, a collection of subsets is σ-algebra if it contains $\varnothing$ and is closed under the operations of taking complements and countable unions (or, equivalently, countable intersections).

Example 1.4. If X is a set, then $\{\varnothing, X\}$ and $\mathcal{P}(X)$ are σ-algebras on X; they are the smallest and largest σ-algebras on X, respectively.

Measurable spaces provide the domain of measures, defined below.

Definition 1.5. A measurable space $(X, \mathcal{A})$ is a non-empty set X equipped with a σ-algebra $\mathcal{A}$ on X.

It is useful to compare the definition of a σ-algebra with that of a topology in Definition 1.1. There are two significant differences. First, the complement of a measurable set is measurable, but the complement of an open set is not, in general, open, excluding special cases such as the discrete topology $\mathcal{T} = \mathcal{P}(X)$. Second, countable intersections and unions of measurable sets are measurable, but only finite intersections of open sets are open while arbitrary (even uncountable) unions of open sets are open. Despite the formal similarities, the properties of measurable and open sets are very different, and they do not combine in a straightforward way.

If $\mathcal{F}$ is any collection of subsets of a set X, then there is a smallest σ-algebra on X that contains $\mathcal{F}$, denoted by $\sigma(\mathcal{F})$.

Definition 1.6. If $\mathcal{F}$ is any collection of subsets of a set X, then the σ-algebra generated by $\mathcal{F}$ is

$$\sigma(\mathcal{F}) = \bigcap \{\mathcal{A} \subset \mathcal{P}(X) : \mathcal{A} \supset \mathcal{F} \text{ and } \mathcal{A} \text{ is a } \sigma\text{-algebra}\}.$$

This intersection is nonempty, since $\mathcal{P}(X)$ is a σ-algebra that contains $\mathcal{F}$, and an intersection of σ-algebras is a σ-algebra. An immediate consequence of the definition is the following result, which we will use repeatedly.

Proposition 1.7. *If $\mathcal{F}$ is a collection of subsets of a set X such that $\mathcal{F} \subset \mathcal{A}$ where $\mathcal{A}$ is a σ-algebra on X, then $\sigma(\mathcal{F}) \subset \mathcal{A}$.*

Among the most important σ-algebras are the Borel σ-algebras on topological spaces.

Definition 1.8. Let $(X, \mathcal{T})$ be a topological space. The Borel σ-algebra

$$\mathcal{B}(X) = \sigma(\mathcal{T})$$

is the σ-algebra generated by the collection $\mathcal{T}$ of open sets on X.

1.6. Measures

A measure is a countably additive, non-negative, extended real-valued function defined on a σ-algebra.

Definition 1.9. A measure μ on a measurable space $(X, \mathcal{A})$ is a function

$$\mu : \mathcal{A} \to [0, \infty]$$

such that

(a) $\mu(\varnothing) = 0$;
(b) if $\{A_i \in \mathcal{A} : i \in \mathbb{N}\}$ is a countable disjoint collection of sets in $\mathcal{A}$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

In comparison with an outer measure, a measure need not be defined on all subsets of a set, but it is countably additive rather than countably subadditive. A measure μ on a set X is finite if $\mu(X) < \infty$, and σ-finite if $X = \bigcup_{n=1}^{\infty} A_n$ is a countable union of measurable sets A_n with finite measure, $\mu(A_n) < \infty$. A probability measure is a finite measure with $\mu(X) = 1$.

A measure space $(X, \mathcal{A}, \mu)$ consists of a set X, a σ-algebra $\mathcal{A}$ on X, and a measure μ defined on $\mathcal{A}$. When $\mathcal{A}$ and μ are clear from the context, we will refer to the measure space X. We define subspaces of measure spaces in the natural way.

Definition 1.10. If $(X, \mathcal{A}, \mu)$ is a measure space and $E \subset X$ is a measurable subset, then the measure subspace $(E, \mathcal{A}|_E, \mu|_E)$ is defined by restricting μ to E:

$$\mathcal{A}|_E = \{A \cap E : A \in \mathcal{A}\}, \qquad \mu|_E (A \cap E) = \mu(A \cap E).$$

As we will see, the construction of nontrivial measures, such as Lebesgue measure, requires considerable effort. Nevertheless, there is at least one useful example of a measure that is simple to define.

Example 1.11. Let X be an arbitrary non-empty set. Define $\nu : \mathcal{P}(X) \to [0, \infty]$ by

$$\nu(E) = \text{number of elements in E},$$

where $\nu(\varnothing) = 0$ and $\nu(E) = \infty$ if E is not finite. Then ν is a measure, called counting measure on X. Every subset of X is measurable with respect to ν. Counting measure is finite if X is finite and σ-finite if X is countable.

A useful implication of the countable additivity of a measure is the following monotonicity result.

Proposition 1.12. *If $\{A_i : i \in \mathbb{N}\}$ is an increasing sequence of measurable sets, meaning that $A_{i+1} \supset A_i$, then*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i). \tag{1.1}$$

If $\{A_i : i \in \mathbb{N}\}$ is a decreasing sequence of measurable sets, meaning that $A_{i+1} \subset A_i$, and $\mu(A_1) < \infty$, then

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i). \tag{1.2}$$

PROOF. If $\{A_i : i \in \mathbb{N}\}$ is an increasing sequence of sets and $B_i = A_{i+1} \setminus A_i$, then $\{B_i : i \in \mathbb{N}\}$ is a disjoint sequence with the same union, so by the countable additivity of μ

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu\left(B_i\right).$$

Moreover, since $A_j = \bigcup_{i=1}^{j} B_i$,

$$\mu(A_j) = \sum_{i=1}^{j} \mu\left(B_i\right),$$

which implies that

$$\sum_{i=1}^{\infty} \mu\left(B_i\right) = \lim_{j\to\infty} \mu(A_j)$$

and the first result follows.

If $\mu(A_1) < \infty$ and $\{A_i\}$ is decreasing, then $\{B_i = A_1 \setminus A_i\}$ is increasing and

$$\mu(B_i) = \mu(A_1) - \mu(A_i).$$

It follows from the previous result that

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mu(B_i) = \mu(A_1) - \lim_{i\to\infty} \mu(A_i).$$

Since

$$\bigcup_{i=1}^{\infty} B_i = A_1 \setminus \bigcap_{i=1}^{\infty} A_i, \qquad \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu(A_1) - \mu\left(\bigcap_{i=1}^{\infty} A_i\right),$$

the result follows. □

Example 1.13. To illustrate the necessity of the condition $\mu(A_1) < \infty$ in the second part of the previous proposition, or more generally $\mu(A_n) < \infty$ for some $n \in \mathbb{N}$, consider counting measure $\nu : \mathcal{P}(\mathbb{N}) \to [0, \infty]$ on $\mathbb{N}$. If

$$A_n = \{k \in \mathbb{N} : k \geq n\},$$

then $\nu(A_n) = \infty$ for every $n \in \mathbb{N}$, so $\nu(A_n) \to \infty$ as $n \to \infty$, but

$$\bigcap_{n=1}^{\infty} A_n = \varnothing, \qquad \nu\left(\bigcap_{n=1}^{\infty} A_n\right) = 0.$$

1.7. Sets of measure zero

A set of measure zero, or a null set, is a measurable set N such that $\mu(N) = 0$. A property which holds for all $x \in X \setminus N$ where N is a set of measure zero is said to hold almost everywhere, or a.e. for short. If we want to emphasize the measure, we say μ-a.e. In general, a subset of a set of measure zero need not be measurable, but if it is, it must have measure zero.

It is frequently convenient to use measure spaces which are complete in the following sense. (This is, of course, a different sense of 'complete' than the one used in talking about complete metric spaces.)

Definition 1.14. A measure space $(X, \mathcal{A}, \mu)$ is complete if every subset of a set of measure zero is measurable.

Note that completeness depends on the measure μ, not just the σ-algebra $\mathcal{A}$. Any measure space $(X, \mathcal{A}, \mu)$ is contained in a uniquely defined completion $(X, \overline{\mathcal{A}}, \overline{\mu})$, which the smallest complete measure space that contains it and is given explicitly as follows.

Theorem 1.15. *If $(X, \mathcal{A}, \mu)$ is a measure space, define $(X, \overline{\mathcal{A}}, \overline{\mu})$ by*

$$\overline{\mathcal{A}} = \{A \cup M : A \in \mathcal{A},\ M \subset N \text{ where } N \in \mathcal{A} \text{ satisfies } \mu(N) = 0\}$$

with $\overline{\mu}(A \cup M) = \mu(A)$. Then $(X, \overline{\mathcal{A}}, \overline{\mu})$ is a complete measure space such that $\overline{\mathcal{A}} \supset \mathcal{A}$ and $\overline{\mu}$ is the unique extension of μ to $\overline{\mathcal{A}}$.

Proof. The collection $\overline{\mathcal{A}}$ is a σ-algebra. It is closed under complementation because, with the notation used in the definition,

$$(A \cup M)^c = A^c \cap M^c, \qquad M^c = N^c \cup (N \setminus M).$$

Therefore

$$(A \cup M)^c = (A^c \cap N^c) \cup (A^c \cap (N \setminus M)) \in \overline{\mathcal{A}},$$

since $A^c \cap N^c \in \mathcal{A}$ and $A^c \cap (N \setminus M) \subset N$. Moreover, $\overline{\mathcal{A}}$ is closed under countable unions because if $A_i \in \mathcal{A}$ and $M_i \subset N_i$ where $\mu(N_i) = 0$ for each $i \in \mathbb{N}$, then

$$\bigcup_{i=1}^{\infty} A_i \cup M_i = \left(\bigcup_{i=1}^{\infty} A_i\right) \cup \left(\bigcup_{i=1}^{\infty} M_i\right) \in \overline{\mathcal{A}},$$

since

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}, \qquad \bigcup_{i=1}^{\infty} M_i \subset \bigcup_{i=1}^{\infty} N_i, \quad \mu\left(\bigcup_{i=1}^{\infty} N_i\right) = 0.$$

It is straightforward to check that $\overline{\mu}$ is well-defined and is the unique extension of μ to a measure on $\overline{\mathcal{A}}$, and that $(X, \overline{\mathcal{A}}, \overline{\mu})$ is complete. □

CHAPTER 2

Lebesgue Measure on $\mathbb{R}^n$

Our goal is to construct a notion of the volume, or Lebesgue measure, of rather general subsets of $\mathbb{R}^n$ that reduces to the usual volume of elementary geometrical sets such as cubes or rectangles.

If $\mathcal{L}(\mathbb{R}^n)$ denotes the collection of Lebesgue measurable sets and

$$\mu : \mathcal{L}(\mathbb{R}^n) \to [0, \infty]$$

denotes Lebesgue measure, then we want $\mathcal{L}(\mathbb{R}^n)$ to contain all n-dimensional rectangles and $\mu(R)$ should be the usual volume of a rectangle R. Moreover, we want μ to be countably additive. That is, if

$$\{A_i \in \mathcal{L}(\mathbb{R}^n) : i \in \mathbb{N}\}$$

is a countable collection of disjoint measurable sets, then their union should be measurable and

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu\left(A_i\right).$$

The reason for requiring countable additivity is that finite additivity is too weak a property to allow the justification of any limiting processes, while uncountable additivity is too strong; for example, it would imply that if the measure of a set consisting of a single point is zero, then the measure of every subset of $\mathbb{R}^n$ would be zero.

It is not possible to define the Lebesgue measure of all subsets of $\mathbb{R}^n$ in a geometrically reasonable way. Hausdorff (1914) showed that for any dimension $n \geq 1$, there is no countably additive measure defined on all subsets of $\mathbb{R}^n$ that is invariant under isometries (translations and rotations) and assigns measure one to the unit cube. He further showed that if $n \geq 3$, there is no such finitely additive measure. This result is dramatized by the Banach-Tarski 'paradox': Banach and Tarski (1924) showed that if $n \geq 3$, one can cut up a ball in $\mathbb{R}^n$ into a finite number of pieces and use isometries to reassemble the pieces into a ball of any desired volume *e.g.* reassemble a pea into the sun. The 'construction' of these pieces requires the axiom of choice.[1] Banach (1923) also showed that if $n = 1$ or $n = 2$ there are finitely additive, isometrically invariant extensions of Lebesgue measure on $\mathbb{R}^n$ that are defined on all subsets of $\mathbb{R}^n$, but these extensions are not countably additive. For a detailed discussion of the Banach-Tarski paradox and related issues, see [**10**].

The moral of these results is that some subsets of $\mathbb{R}^n$ are too irregular to define their Lebesgue measure in a way that preserves countable additivity (or even finite additivity in $n \geq 3$ dimensions) together with the invariance of the measure under

[1] Solovay (1970) proved that one has to use the axiom of choice to obtain non-Lebesgue measurable sets.

isometries. We will show, however, that such a measure can be defined on a σ-algebra $\mathcal{L}(\mathbb{R}^n)$ of Lebesgue measurable sets which is large enough to include all set of 'practical' importance in analysis. Moreover, as we will see, it is possible to define an isometrically-invariant, countably *sub*-additive outer measure on all subsets of $\mathbb{R}^n$.

There are many ways to construct Lebesgue measure, all of which lead to the same result. We will follow an approach due to Carathéodory, which generalizes to other measures: We first construct an outer measure on all subsets of $\mathbb{R}^n$ by approximating them from the outside by countable unions of rectangles; we then restrict this outer measure to a σ-algebra of measurable subsets on which it is countably additive. This approach is somewhat asymmetrical in that we approximate sets (and their complements) from the outside by elementary sets, but we do not approximate them directly from the inside.

Jones [**5**], Stein and Shakarchi [**8**], and Wheeler and Zygmund [**11**] give detailed introductions to Lebesgue measure on $\mathbb{R}^n$. Cohn [**2**] gives a similar development to the one here, and Evans and Gariepy [**3**] discuss more advanced topics.

2.1. Lebesgue outer measure

We use rectangles as our elementary sets, defined as follows.

Definition 2.1. An n-dimensional, closed rectangle with sides oriented parallel to the coordinate axes, or rectangle for short, is a subset $R \subset \mathbb{R}^n$ of the form

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

where $-\infty < a_i \leq b_i < \infty$ for $i = 1, \dots, n$. The volume $\mu(R)$ of R is

$$\mu(R) = (b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n).$$

If $n = 1$ or $n = 2$, the volume of a rectangle is its length or area, respectively. We also consider the empty set to be a rectangle with $\mu(\varnothing) = 0$. We denote the collection of all n-dimensional rectangles by $\mathcal{R}(\mathbb{R}^n)$, or $\mathcal{R}$ when n is understood, and then $R \mapsto \mu(R)$ defines a map

$$\mu : \mathcal{R}(\mathbb{R}^n) \to [0, \infty).$$

The use of this particular class of elementary sets is for convenience. We could equally well use open or half-open rectangles, cubes, balls, or other suitable elementary sets; the result would be the same.

Definition 2.2. The outer Lebesgue measure $\mu^*(E)$ of a subset $E \subset \mathbb{R}^n$, or outer measure for short, is

$$\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(R_i) : E \subset \textstyle\bigcup_{i=1}^{\infty} R_i,\, R_i \in \mathcal{R}(\mathbb{R}^n) \right\} \tag{2.1}$$

where the infimum is taken over all countable collections of rectangles whose union contains E. The map

$$\mu^* : \mathcal{P}(\mathbb{R}^n) \to [0, \infty], \qquad \mu^* : E \mapsto \mu^*(E)$$

is called outer Lebesgue measure.

In this definition, a sum $\sum_{i=1}^{\infty} \mu(R_i)$ and $\mu^*(E)$ may take the value ∞. We do not require that the rectangles R_i are disjoint, so the same volume may contribute to multiple terms in the sum on the right-hand side of (2.1); this does not affect the value of the infimum.

Example 2.3. Let $E = \mathbb{Q} \cap [0,1]$ be the set of rational numbers between 0 and 1. Then E has outer measure zero. To prove this, let $\{q_i : i \in \mathbb{N}\}$ be an enumeration of the points in E. Given $\epsilon > 0$, let R_i be an interval of length $\epsilon/2^i$ which contains q_i. Then $E \subset \bigcup_{i=1}^{\infty} \mu(R_i)$ so

$$0 \le \mu^*(E) \le \sum_{i=1}^{\infty} \mu(R_i) = \epsilon.$$

Hence $\mu^*(E) = 0$ since $\epsilon > 0$ is arbitrary. The same argument shows that any countable set has outer measure zero. Note that if we cover E by a *finite* collection of intervals, then the union of the intervals would have to contain $[0,1]$ since E is dense in $[0,1]$ so their lengths sum to at least one.

The previous example illustrates why we need to use countably infinite collections of rectangles, not just finite collections, to define the outer measure.[2] The 'countable ϵ-trick' used in the example appears in various forms throughout measure theory.

Next, we prove that μ^* is an outer measure in the sense of Definition 1.2.

Theorem 2.4. *Lebesgue outer measure μ^* has the following properties.*

(a) $\mu^*(\varnothing) = 0$;
(b) if $E \subset F$, *then* $\mu^*(E) \le \mu^*(F)$;
(c) if $\{E_i \subset \mathbb{R}^n : i \in \mathbb{N}\}$ *is a countable collection of subsets of* $\mathbb{R}^n$, *then*

$$\mu^*\left(\bigcup_{i=1}^{\infty} E_i\right) \le \sum_{i=1}^{\infty} \mu^*\left(E_i\right).$$

Proof. It follows immediately from Definition 2.2 that $\mu^*(\varnothing) = 0$, since every collection of rectangles covers $\varnothing$, and that $\mu^*(E) \le \mu^*(F)$ if $E \subset F$ since any cover of F covers E.

The main property to prove is the countable subadditivity of μ^*. If $\mu^*(E_i) = \infty$ for some $i \in \mathbb{N}$, there is nothing to prove, so we may assume that $\mu^*(E_i)$ is finite for every $i \in \mathbb{N}$. If $\epsilon > 0$, there is a countable covering $\{R_{ij} : j \in \mathbb{N}\}$ of E_i by rectangles R_{ij} such that

$$\sum_{j=1}^{\infty} \mu(R_{ij}) \le \mu^*(E_i) + \frac{\epsilon}{2^i}, \qquad E_i \subset \bigcup_{j=1}^{\infty} R_{ij}.$$

Then $\{R_{ij} : i,j \in \mathbb{N}\}$ is a countable covering of

$$E = \bigcup_{i=1}^{\infty} E_i$$

[2]The use of finitely many intervals leads to the notion of the Jordan content of a set, introduced by Peano (1887) and Jordan (1892), which is closely related to the Riemann integral; Borel (1898) and Lebesgue (1902) generalized Jordan's approach to allow for countably many intervals, leading to Lebesgue measure and the Lebesgue integral.

and therefore

$$\mu^*(E) \leq \sum_{i,j=1}^{\infty} \mu(R_{ij}) \leq \sum_{i=1}^{\infty} \left\{ \mu^*(E_i) + \frac{\epsilon}{2^i} \right\} = \sum_{i=1}^{\infty} \mu^*(E_i) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, it follows that

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

which proves the result. □

2.2. Outer measure of rectangles

In this section, we prove the geometrically obvious, but not entirely trivial, fact that the outer measure of a rectangle is equal to its volume. The main point is to show that the volumes of a countable collection of rectangles that cover a rectangle R cannot sum to less than the volume of R.[3]

We begin with some combinatorial facts about finite covers of rectangles [**8**]. We denote the interior of a rectangle R by R°, and we say that rectangles R, S are almost disjoint if $R^\circ \cap S^\circ = \varnothing$, meaning that they intersect at most along their boundaries. The proofs of the following results are cumbersome to write out in detail (it's easier to draw a picture) but we briefly explain the argument.

Lemma 2.5. *Suppose that*

$$R = I_1 \times I_2 \times \cdots \times I_n$$

is an n-dimensional rectangle where each closed, bounded interval $I_i \subset \mathbb{R}$ is an almost disjoint union of closed, bounded intervals $\{I_{i,j} \subset \mathbb{R} : j = 1, \dots, N_i\}$,

$$I_i = \bigcup_{j=1}^{N_i} I_{i,j}.$$

Define the rectangles

$$S_{j_1 j_2 \dots j_n} = I_{1,j_1} \times I_{1,j_2} \times \cdots \times I_{j_n}. \tag{2.2}$$

Then

$$\mu(R) = \sum_{j_1=1}^{N_1} \cdots \sum_{j_n=1}^{N_n} \mu\left(S_{j_1 j_2 \dots j_n}\right).$$

Proof. Denoting the length of an interval I by $|I|$, using the fact that

$$|I_i| = \sum_{j=1}^{N_i} |I_{i,j}|,$$

[3] As a partial justification of the need to prove this fact, note that it would not be true if we allowed uncountable covers, since we could cover any rectangle by an uncountable collection of points all of whose volumes are zero.

and expanding the resulting product, we get that

$$\begin{aligned}
\mu(R) &= |I_1||I_2|\dots|I_n| \\
&= \left(\sum_{j_1=1}^{N_1} |I_{1,j_1}|\right)\left(\sum_{j_2=1}^{N_2} |I_{2,j_2}|\right)\dots\left(\sum_{j_n=1}^{N_n} |I_{n,j_n}|\right) \\
&= \sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2}\dots\sum_{j_n=1}^{N_n} |I_{1,j_1}||I_{2,j_2}|\dots|I_{n,j_n}| \\
&= \sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2}\dots\sum_{j_n=1}^{N_n} \mu\left(S_{j_1 j_2 \dots j_n}\right).
\end{aligned}$$

□

Proposition 2.6. *If a rectangle R is an almost disjoint, finite union of rectangles $\{R_1, R_2, \dots, R_N\}$, then*

$$\mu(R) = \sum_{i=1}^{N} \mu(R_i). \tag{2.3}$$

If R is covered by rectangles $\{R_1, R_2, \dots, R_N\}$, which need not be disjoint, then

$$\mu(R) \leq \sum_{i=1}^{N} \mu(R_i). \tag{2.4}$$

PROOF. Suppose that

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

is an almost disjoint union of the rectangles $\{R_1, R_2, \dots, R_N\}$. Then by 'extending the sides' of the R_i, we may decompose R into an almost disjoint collection of rectangles

$$\{S_{j_1 j_2 \dots j_n} : 1 \leq j_i \leq N_i \text{ for } 1 \leq i \leq n\}$$

that is obtained by taking products of subintervals of partitions of the coordinate intervals $[a_i, b_i]$ into unions of almost disjoint, closed subintervals. Explicitly, we partition $[a_i, b_i]$ into

$$a_i = c_{i,0} \leq c_{i,1} \leq \cdots \leq c_{i,N_i} = b_i, \qquad I_{i,j} = [c_{i,j-1}, c_{i,j}].$$

where the $c_{i,j}$ are obtained by ordering the left and right ith coordinates of all faces of rectangles in the collection $\{R_1, R_2, \dots, R_N\}$, and define rectangles $S_{j_1 j_2 \dots j_n}$ as in (2.2).

Each rectangle R_i in the collection is an almost disjoint union of rectangles $S_{j_1 j_2 \dots j_n}$, and their union contains all such products exactly once, so by applying Lemma 2.5 to each R_i and summing the results we see that

$$\sum_{i=1}^{N} \mu(R_i) = \sum_{j_1=1}^{N_1} \cdots \sum_{j_n=1}^{N_n} \mu\left(S_{j_1 j_2 \dots j_n}\right).$$

Similarly, R is an almost disjoint union of all the rectangles $S_{j_1 j_2 \dots j_n}$, so Lemma 2.5 implies that

$$\mu(R) = \sum_{j_1=1}^{N_1} \cdots \sum_{j_n=1}^{N_n} \mu\left(S_{j_1 j_2 \dots j_n}\right),$$

and (2.3) follows.

If a finite collection of rectangles $\{R_1, R_2, \ldots, R_N\}$ covers R, then there is a almost disjoint, finite collection of rectangles $\{S_1, S_2, \ldots, S_M\}$ such that

$$R = \bigcup_{i=1}^{M} S_i, \qquad \sum_{i=1}^{M} \mu(S_i) \leq \sum_{i=1}^{N} \mu(R_i).$$

To obtain the S_i, we replace R_i by the rectangle $R \cap R_i$, and then decompose these possibly non-disjoint rectangles into an almost disjoint, finite collection of sub-rectangles with the same union; we discard 'overlaps' which can only reduce the sum of the volumes. Then, using (2.3), we get

$$\mu(R) = \sum_{i=1}^{M} \mu(S_i) \leq \sum_{i=1}^{N} \mu(R_i),$$

which proves (2.4). □

The outer measure of a rectangle is defined in terms of countable covers. We reduce these to finite covers by using the topological properties of $\mathbb{R}^n$.

Proposition 2.7. *If R is a rectangle in $\mathbb{R}^n$, then $\mu^*(R) = \mu(R)$.*

PROOF. Since $\{R\}$ covers R, we have $\mu^*(R) \leq \mu(R)$, so we only need to prove the reverse inequality.

Suppose that $\{R_i : i \in \mathbb{N}\}$ is a countably infinite collection of rectangles that covers R. By enlarging R_i slightly we may obtain a rectangle S_i whose interior S_i° contains R_i such that

$$\mu(S_i) \leq \mu(R_i) + \frac{\epsilon}{2^i}.$$

Then $\{S_i^\circ : i \in \mathbb{N}\}$ is an open cover of the compact set R, so it contains a finite subcover, which we may label as $\{S_1^\circ, S_2^\circ, \ldots, S_N^\circ\}$. Then $\{S_1, S_2, \ldots, S_N\}$ covers R and, using (2.4), we find that

$$\mu(R) \leq \sum_{i=1}^{N} \mu(S_i) \leq \sum_{i=1}^{N} \left\{\mu(R_i) + \frac{\epsilon}{2^i}\right\} \leq \sum_{i=1}^{\infty} \mu(R_i) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have

$$\mu(R) \leq \sum_{i=1}^{\infty} \mu(R_i)$$

and it follows that $\mu(R) \leq \mu^*(R)$. □

2.3. Carathéodory measurability

We will obtain Lebesgue measure as the restriction of Lebesgue outer measure to Lebesgue measurable sets. The construction, due to Carathéodory, works for any outer measure, as given in Definition 1.2, so we temporarily consider general outer measures. We will return to Lebesgue measure on $\mathbb{R}^n$ at the end of this section.

The following is the Carathéodory definition of measurability.

Definition 2.8. Let μ^* be an outer measure on a set X. A subset $A \subset X$ is Carathéodory measurable with respect to μ^*, or measurable for short, if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{2.5}$$

for every subset $E \subset X$.

We also write $E \cap A^c$ as $E \setminus A$. Thus, a measurable set A splits any set E into disjoint pieces whose outer measures add up to the outer measure of E. Heuristically, this condition means that a set is measurable if it divides other sets in a 'nice' way. The regularity of the set E being divided is not important here.

Since μ^* is subadditive, we always have that

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Thus, in order to prove that $A \subset X$ is measurable, it is sufficient to show that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for every $E \subset X$, and then we have equality as in (2.5).

Definition 2.8 is perhaps not the most intuitive way to define the measurability of sets, but it leads directly to the following key result.

Theorem 2.9. *The collection of Carathéodory measurable sets with respect to an outer measure μ^* is a σ-algebra, and the restriction of μ^* to the measurable sets is a measure.*

Proof. It follows immediately from (2.5) that $\varnothing$ is measurable and the complement of a measurable set is measurable, so to prove that the collection of measurable sets is a σ-algebra, we only need to show that it is closed under countable unions. We will prove at the same time that μ^* is countably additive on measurable sets; since $\mu^*(\varnothing) = 0$, this will prove that the restriction of μ^* to the measurable sets is a measure.

First, we prove that the union of measurable sets is measurable. Suppose that A, B are measurable and $E \subset X$. The measurability of A and B implies that

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) \\ &\qquad + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c). \end{aligned} \tag{2.6}$$

Since $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$ and μ^* is subadditive, we have

$$\mu^*(E \cap (A \cup B)) \leq \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B).$$

The use of this inequality and the relation $A^c \cap B^c = (A \cup B)^c$ in (2.6) implies that

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

so $A \cup B$ is measurable.

Moreover, if A is measurable and $A \cap B = \varnothing$, then by taking $E = A \cup B$ in (2.5), we see that

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Thus, the outer measure of the union of disjoint, measurable sets is the sum of their outer measures. The repeated application of this result implies that the finite union of measurable sets is measurable and μ^* is finitely additive on the collection of measurable sets.

Next, we we want to show that the countable union of measurable sets is measurable. It is sufficient to consider disjoint unions. To see this, note that if

$\{A_i : i \in \mathbb{N}\}$ is a countably infinite collection of measurable sets, then

$$B_j = \bigcup_{i=1}^{j} A_i, \qquad \text{for } j \geq 1$$

form an increasing sequence of measurable sets, and

$$C_j = B_j \setminus B_{j-1} \quad \text{for } j \geq 2, \qquad C_1 = B_1$$

form a disjoint measurable collection of sets. Moreover

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{j=1}^{\infty} C_j.$$

Suppose that $\{A_i : i \in \mathbb{N}\}$ is a countably infinite, disjoint collection of measurable sets, and define

$$B_j = \bigcup_{i=1}^{j} A_i, \qquad B = \bigcup_{i=1}^{\infty} A_i.$$

Let $E \subset X$. Since A_j is measurable and $B_j = A_j \cup B_{j-1}$ is a disjoint union (for $j \geq 2$),

$$\begin{aligned} \mu^*(E \cap B_j) &= \mu^*(E \cap B_j \cap A_j) + \mu^*(E \cap B_j \cap A_j^c), \\ &= \mu^*(E \cap A_j) + \mu^*(E \cap B_{j-1}). \end{aligned}$$

Also $\mu^*(E \cap B_1) = \mu^*(E \cap A_1)$. It follows by induction that

$$\mu^*(E \cap B_j) = \sum_{i=1}^{j} \mu^*(E \cap A_i).$$

Since B_j is a finite union of measurable sets, it is measurable, so

$$\mu^*(E) = \mu^*(E \cap B_j) + \mu^*(E \cap B_j^c),$$

and since $B_j^c \supset B^c$, we have

$$\mu^*(E \cap B_j^c) \geq \mu^*(E \cap B^c).$$

It follows that

$$\mu^*(E) \geq \sum_{i=1}^{j} \mu^*(E \cap A_i) + \mu^*(E \cap B^c).$$

Taking the limit of this inequality as $j \to \infty$ and using the subadditivity of μ^*, we get

$$\begin{aligned} \mu^*(E) &\geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap B^c) \\ &\geq \mu^*\left(\bigcup_{i=1}^{\infty} E \cap A_i\right) + \mu^*(E \cap B^c) \qquad (2.7) \\ &\geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \\ &\geq \mu^*(E). \end{aligned}$$

Therefore, we must have equality in (2.7), which shows that $B = \bigcup_{i=1}^{\infty} A_i$ is measurable. Moreover,

$$\mu^*\left(\bigcup_{i=1}^{\infty} E \cap A_i\right) = \sum_{i=1}^{\infty} \mu^*(E \cap A_i),$$

so taking $E = X$, we see that μ^* is countably additive on the σ-algebra of measurable sets. □

Returning to Lebesgue measure on $\mathbb{R}^n$, the preceding theorem shows that we get a measure on $\mathbb{R}^n$ by restricting Lebesgue outer measure to its Carathéodory-measurable sets, which are the Lebesgue measurable subsets of $\mathbb{R}^n$.

Definition 2.10. A subset $A \subset \mathbb{R}^n$ is Lebesgue measurable if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for every subset $E \subset \mathbb{R}^n$. If $\mathcal{L}(\mathbb{R}^n)$ denotes the σ-algebra of Lebesgue measurable sets, the restriction of Lebesgue outer measure μ^* to the Lebesgue measurable sets

$$\mu : \mathcal{L}(\mathbb{R}^n) \to [0, \infty], \qquad \mu = \mu^*|_{\mathcal{L}(\mathbb{R}^n)}$$

is called Lebesgue measure.

From Proposition 2.7, this notation is consistent with our previous use of μ to denote the volume of a rectangle. If $E \subset \mathbb{R}^n$ is any measurable subset of $\mathbb{R}^n$, then we define Lebesgue measure on E by restricting Lebesgue measure on $\mathbb{R}^n$ to E, as in Definition 1.10, and denote the corresponding σ-algebra of Lebesgue measurable subsets of E by $\mathcal{L}(E)$.

Next, we prove that all rectangles are measurable; this implies that $\mathcal{L}(\mathbb{R}^n)$ is a 'large' collection of subsets of $\mathbb{R}^n$. Not all subsets of $\mathbb{R}^n$ are Lebesgue measurable, however; *e.g.* see Example 2.17 below.

Proposition 2.11. *Every rectangle is Lebesgue measurable.*

PROOF. Let R be an n-dimensional rectangle and $E \subset \mathbb{R}^n$. Given $\epsilon > 0$, there is a cover $\{R_i : i \in \mathbb{N}\}$ of E by rectangles R_i such that

$$\mu^*(E) + \epsilon \geq \sum_{i=1}^{\infty} \mu(R_i).$$

We can decompose R_i into an almost disjoint, finite union of rectangles

$$\{\tilde{R}_i, S_{i,1}, \dots, S_{i,N}\}$$

such that

$$R_i = \tilde{R}_i + \bigcup_{j=1}^{N} S_{i,j}, \qquad \tilde{R}_i = R_i \cap R \subset R, \quad S_{i,j} \subset \overline{R^c}.$$

From (2.3),

$$\mu(R_i) = \mu(\tilde{R}_i) + \sum_{j=1}^{N} \mu(S_{i,j}).$$

Using this result in the previous sum, relabeling the $S_{i,j}$ as S_i, and rearranging the resulting sum, we get that

$$\mu^*(E) + \epsilon \geq \sum_{i=1}^{\infty} \mu(\tilde{R}_i) + \sum_{i=1}^{\infty} \mu(S_i).$$

Since the rectangles $\{\tilde{R}_i : i \in \mathbb{N}\}$ cover $E \cap R$ and the rectangles $\{S_i : i \in \mathbb{N}\}$ cover $E \cap R^c$, we have

$$\mu^*(E \cap R) \leq \sum_{i=1}^{\infty} \mu(\tilde{R}_i), \qquad \mu^*(E \cap R^c) \leq \sum_{i=1}^{\infty} \mu(S_i).$$

Hence,

$$\mu^*(E) + \epsilon \geq \mu^*(E \cap R) + \mu^*(E \cap R^c).$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\mu^*(E) \geq \mu^*(E \cap R) + \mu^*(E \cap R^c),$$

which proves the result. □

An open rectangle R° is a union of an increasing sequence of closed rectangles whose volumes approach $\mu(R)$; for example

$$\begin{aligned}&(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)\\ &\quad = \bigcup_{k=1}^{\infty} [a_1 + \frac{1}{k}, b_1 - \frac{1}{k}] \times [a_2 + \frac{1}{k}, b_2 - \frac{1}{k}] \times \cdots \times [a_n + \frac{1}{k}, b_n - \frac{1}{k}].\end{aligned}$$

Thus, R° is measurable and, from Proposition 1.12,

$$\mu(R^\circ) = \mu(R).$$

Moreover if $\partial R = R \setminus R^\circ$ denotes the boundary of R, then

$$\mu(\partial R) = \mu(R) - \mu(R^\circ) = 0.$$

2.4. Null sets and completeness

Sets of measure zero play a particularly important role in measure theory and integration. First, we show that all sets with outer Lebesgue measure zero are Lebesgue measurable.

Proposition 2.12. *If $N \subset \mathbb{R}^n$ and $\mu^*(N) = 0$, then N is Lebesgue measurable, and the measure space $(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \mu)$ is complete.*

Proof. If $N \subset \mathbb{R}^n$ has outer Lebesgue measure zero and $E \subset \mathbb{R}^n$, then

$$0 \leq \mu^*(E \cap N) \leq \mu^*(N) = 0,$$

so $\mu^*(E \cap N) = 0$. Therefore, since $E \supset E \cap N^c$,

$$\mu^*(E) \geq \mu^*(E \cap N^c) = \mu^*(E \cap N) + \mu^*(E \cap N^c),$$

which shows that N is measurable. If N is a measurable set with $\mu(N) = 0$ and $M \subset N$, then $\mu^*(M) = 0$, since $\mu^*(M) \leq \mu(N)$. Therefore M is measurable and $(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \mu)$ is complete. □

In view of the importance of sets of measure zero, we formulate their definition explicitly.

Definition 2.13. A subset $N \subset \mathbb{R}^n$ has Lebesgue measure zero if for every $\epsilon > 0$ there exists a countable collection of rectangles $\{R_i : i \in \mathbb{N}\}$ such that

$$N \subset \bigcup_{i=1}^{\infty} R_i, \qquad \sum_{i=1}^{\infty} \mu(R_i) < \epsilon.$$

The argument in Example 2.3 shows that every countable set has Lebesgue measure zero, but sets of measure zero may be uncountable; in fact the fine structure of sets of measure zero is, in general, very intricate.

Example 2.14. The standard Cantor set, obtained by removing 'middle thirds' from $[0, 1]$, is an uncountable set of zero one-dimensional Lebesgue measure.

Example 2.15. The x-axis in $\mathbb{R}^2$

$$A = \left\{(x, 0) \in \mathbb{R}^2 : x \in \mathbb{R}\right\}$$

has zero two-dimensional Lebesgue measure. More generally, any linear subspace of $\mathbb{R}^n$ with dimension strictly less than n has zero n-dimensional Lebesgue measure.

2.5. Translational invariance

An important geometric property of Lebesgue measure is its translational invariance. If $A \subset \mathbb{R}^n$ and $h \in \mathbb{R}^n$, let

$$A + h = \{x + h : x \in A\}$$

denote the translation of A by h.

Proposition 2.16. *If $A \subset \mathbb{R}^n$ and $h \in \mathbb{R}^n$, then*

$$\mu^*(A + h) = \mu^*(A),$$

and $A + h$ is measurable if and only if A is measurable.

Proof. The invariance of outer measure μ^* result is an immediate consequence of the definition, since $\{R_i + h : i \in \mathbb{N}\}$ is a cover of $A + h$ if and only if $\{R_i : i \in \mathbb{N}\}$ is a cover of A, and $\mu(R + h) = \mu(R)$ for every rectangle R. Moreover, the Carathéodory definition of measurability is invariant under translations since

$$(E + h) \cap (A + h) = (E \cap A) + h.$$

□

The space $\mathbb{R}^n$ is a locally compact topological (abelian) group with respect to translation, which is a continuous operation. More generally, there exists a (left or right) translation-invariant measure, called Haar measure, on any locally compact topological group; this measure is unique up to a scalar factor.

The following is the standard example of a non-Lebesgue measurable set, due to Vitali (1905).

Example 2.17. Define an equivalence relation $\sim$ on $\mathbb{R}$ by $x \sim y$ if $x - y \in \mathbb{Q}$. This relation has uncountably many equivalence classes, each of which contains a countably infinite number of points and is dense in $\mathbb{R}$. Let $E \subset [0, 1]$ be a set that contains exactly one element from each equivalence class, so that $\mathbb{R}$ is the disjoint union of the countable collection of rational translates of E. Then we claim that E is not Lebesgue measurable.

To show this, suppose for contradiction that E is measurable. Let $\{q_i : i \in \mathbb{N}\}$ be an enumeration of the rational numbers in the interval $[-1, 1]$ and let $E_i = E + q_i$ denote the translation of E by q_i. Then the sets E_i are disjoint and

$$[0, 1] \subset \bigcup_{i=1}^{\infty} E_i \subset [-1, 2].$$

The translational invariance of Lebesgue measure implies that each E_i is measurable with $\mu(E_i) = \mu(E)$, and the countable additivity of Lebesgue measure implies that

$$1 \leq \sum_{i=1}^{\infty} \mu(E_i) \leq 3.$$

But this is impossible, since $\sum_{i=1}^{\infty} \mu(E_i)$ is either 0 or ∞, depending on whether if $\mu(E) = 0$ or $\mu(E) > 0$.

The above example is geometrically simpler on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. When reduced modulo one, the sets $\{E_i : i \in \mathbb{N}\}$ partition $\mathbb{T}$ into a countable union of disjoint sets which are translations of each other. If the sets were measurable, their measures would be equal so they must sum to 0 or ∞, but the measure of $\mathbb{T}$ is one.

2.6. Borel sets

The relationship between measure and topology is not a simple one. In this section, we show that all open and closed sets in $\mathbb{R}^n$, and therefore all Borel sets (*i.e.* sets that belong to the σ-algebra generated by the open sets), are Lebesgue measurable.

Let $\mathcal{T}(\mathbb{R}^n) \subset \mathcal{P}(\mathbb{R}^n)$ denote the standard metric topology on $\mathbb{R}^n$ consisting of all open sets. That is, $G \subset \mathbb{R}^n$ belongs to $\mathcal{T}(\mathbb{R}^n)$ if for every $x \in G$ there exists $r > 0$ such that $B_r(x) \subset G$, where

$$B_r(x) = \{y \in \mathbb{R}^n : |x - y| < r\}$$

is the open ball of radius r centered at $x \in \mathbb{R}^n$ and $|\cdot|$ denotes the Euclidean norm.

Definition 2.18. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ on $\mathbb{R}^n$ is the σ-algebra generated by the open sets, $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{T}(\mathbb{R}^n))$. A set that belongs to the Borel σ-algebra is called a Borel set.

Since σ-algebras are closed under complementation, the Borel σ-algebra is also generated by the closed sets in $\mathbb{R}^n$. Moreover, since $\mathbb{R}^n$ is σ-compact (*i.e.* it is a countable union of compact sets) its Borel σ-algebra is generated by the compact sets.

Remark 2.19. This definition is not constructive, since we start with the power set of $\mathbb{R}^n$ and narrow it down until we obtain the smallest σ-algebra that contains the open sets. It is surprisingly complicated to obtain $\mathcal{B}(\mathbb{R}^n)$ by starting from the open or closed sets and taking successive complements, countable unions, and countable intersections. These operations give sequences of collections of sets in $\mathbb{R}^n$

$$G \subset G_\delta \subset G_{\delta\sigma} \subset G_{\delta\sigma\delta} \subset \dots, \qquad F \subset F_\sigma \subset F_{\sigma\delta} \subset F_{\delta\sigma\delta} \subset \dots, \tag{2.8}$$

where G denotes the open sets, F the closed sets, σ the operation of countable unions, and δ the operation of countable intersections. These collections contain each other; for example, $F_\sigma \supset G$ and $G_\delta \supset F$. This process, however, has to be repeated up to the first uncountable ordinal before we obtain $\mathcal{B}(\mathbb{R}^n)$. This is because if, for example, $\{A_i : i \in \mathbb{N}\}$ is a countable family of sets such that

$$A_1 \in G_\delta \setminus G, \quad A_2 \in G_{\delta\sigma} \setminus G_\delta, \quad A_3 \in G_{\delta\sigma\delta} \setminus G_{\delta\sigma}, \dots$$

and so on, then there is no guarantee that $\bigcup_{i=1}^{\infty} A_i$ or $\bigcap_{i=1}^{\infty} A_i$ belongs to any of the previously constructed families. In general, one only knows that they belong to the $\omega + 1$ iterates $G_{\delta\sigma\delta\dots\sigma}$ or $G_{\delta\sigma\delta\dots\delta}$, respectively, where ω is the ordinal number of $\mathbb{N}$. A similar argument shows that in order to obtain a family which is closed under countable intersections or unions, one has to continue this process until one has constructed an uncountable number of families.

To show that open sets are measurable, we will represent them as countable unions of rectangles. Every open set in $\mathbb{R}$ is a countable disjoint union of open intervals (one-dimensional open rectangles). When $n \geq 2$, it is not true that every open set in $\mathbb{R}^n$ is a countable disjoint union of open rectangles, but we have the following substitute.

Proposition 2.20. *Every open set in $\mathbb{R}^n$ is a countable union of almost disjoint rectangles.*

Proof. Let $G \subset \mathbb{R}^n$ be open. We construct a family of cubes (rectangles of equal sides) as follows. First, we bisect $\mathbb{R}^n$ into almost disjoint cubes $\{Q_i : i \in \mathbb{N}\}$ of side one with integer coordinates. If $Q_i \subset G$, we include Q_i in the family, and if Q_i is disjoint from G, we exclude it. Otherwise, we bisect the sides of Q_i to obtain 2^n almost disjoint cubes of side one-half and repeat the procedure. Iterating this process arbitrarily many times, we obtain a countable family of almost disjoint cubes.

The union of the cubes in this family is contained in G, since we only include cubes that are contained in G. Conversely, if $x \in G$, then since G is open some sufficiently small cube in the bisection procedure that contains x is entirely contained in G, and the largest such cube is included in the family. Hence the union of the family contains G, and is therefore equal to G. □

In fact, the proof shows that every open set is an almost disjoint union of dyadic cubes.

Proposition 2.21. *The Borel algebra $\mathcal{B}(\mathbb{R}^n)$ is generated by the collection of rectangles $\mathcal{R}(\mathbb{R}^n)$. Every Borel set is Lebesgue measurable.*

Proof. Since $\mathcal{R}$ is a subset of the closed sets, we have $\sigma(\mathcal{R}) \subset \mathcal{B}$. Conversely, by the previous proposition, $\sigma(\mathcal{R}) \supset \mathcal{T}$, so $\sigma(\mathcal{R}) \supset \sigma(\mathcal{T}) = \mathcal{B}$, and therefore $\mathcal{B} = \sigma(\mathcal{R})$. From Proposition 2.11, we have $\mathcal{R} \subset \mathcal{L}$. Since $\mathcal{L}$ is a σ-algebra, it follows that $\sigma(\mathcal{R}) \subset \mathcal{L}$, so $\mathcal{B} \subset \mathcal{L}$. □

Note that if

$$G = \bigcup_{i=1}^{\infty} R_i$$

is a decomposition of an open set G into an almost disjoint union of closed rectangles, then

$$G \supset \bigcup_{i=1}^{\infty} R_i^\circ$$

is a disjoint union, and therefore

$$\sum_{i=1}^{\infty} \mu(R_i^\circ) \leq \mu(G) \leq \sum_{i=1}^{\infty} \mu(R_i).$$

Since $\mu(R_i^\circ) = \mu(R_i)$, it follows that

$$\mu(G) = \sum_{i=1}^{\infty} \mu(R_i)$$

for any such decomposition and that the sum is independent of the way in which G is decomposed into almost disjoint rectangles.

The Borel σ-algebra $\mathcal{B}$ is not complete and is strictly smaller than the Lebesgue σ-algebra $\mathcal{L}$. In fact, one can show that the cardinality of $\mathcal{B}$ is equal to the cardinality $\mathfrak{c}$ of the real numbers, whereas the cardinality of $\mathcal{L}$ is equal to $2^{\mathfrak{c}}$. For example, the Cantor set is a set of measure zero with the same cardinality as $\mathbb{R}$ and every subset of the Cantor set is Lebesgue measurable.

We can obtain examples of sets that are Lebesgue measurable but not Borel measurable by considering subsets of sets of measure zero. In the following example of such a set in $\mathbb{R}$, we use some properties of measurable functions which will be proved later.

Example 2.22. Let $f : [0, 1] \to [0, 1]$ denote the standard Cantor function and define $g : [0, 1] \to [0, 1]$ by

$$g(y) = \inf \{x \in [0, 1] : f(x) = y\}.$$

Then g is an increasing, one-to-one function that maps $[0, 1]$ onto the Cantor set C. Since g is increasing it is Borel measurable, and the inverse image of a Borel set under g is Borel. Let $E \subset [0, 1]$ be a non-Lebesgue measurable set. Then $F = g(E) \subset C$ is Lebesgue measurable, since it is a subset of a set of measure zero, but F is not Borel measurable, since if it was $E = g^{-1}(F)$ would be Borel.

Other examples of Lebesgue measurable sets that are not Borel sets arise from the theory of product measures in $\mathbb{R}^n$ for $n \geq 2$. For example, let $N = E \times \{0\} \subset \mathbb{R}^2$ where $E \subset \mathbb{R}$ is a non-Lebesgue measurable set in $\mathbb{R}$. Then N is a subset of the x-axis, which has two-dimensional Lebesgue measure zero, so N belongs to $\mathcal{L}(\mathbb{R}^2)$ since Lebesgue measure is complete. One can show, however, that if a set belongs to $\mathcal{B}(\mathbb{R}^2)$ then every section with fixed x or y coordinate, belongs to $\mathcal{B}(\mathbb{R})$; thus, N cannot belong to $\mathcal{B}(\mathbb{R}^2)$ since the $y = 0$ section E is not Borel.

As we show below, $\mathcal{L}(\mathbb{R}^n)$ is the completion of $\mathcal{B}(\mathbb{R}^n)$ with respect to Lebesgue measure, meaning that we get all Lebesgue measurable sets by adjoining all subsets of Borel sets of measure zero to the Borel σ-algebra and taking unions of such sets.

2.7. Borel regularity

Regularity properties of measures refer to the possibility of approximating in measure one class of sets (for example, nonmeasurable sets) by another class of sets (for example, measurable sets). Lebesgue measure is Borel regular in the sense that Lebesgue measurable sets can be approximated in measure from the outside by open sets and from the inside by closed sets, and they can be approximated by Borel sets up to sets of measure zero. Moreover, there is a simple criterion for Lebesgue measurability in terms of open and closed sets.

The following theorem expresses a fundamental approximation property of Lebesgue measurable sets by open and compact sets. Equations (2.9) and (2.10) are called outer and inner regularity, respectively.

Theorem 2.23. *If $A \subset \mathbb{R}^n$, then*

$$\mu^*(A) = \inf \{\mu(G) : A \subset G,\ G \text{ open}\}, \tag{2.9}$$

and if A is Lebesgue measurable, then

$$\mu(A) = \sup \{\mu(K) : K \subset A,\ K \text{ compact}\}. \tag{2.10}$$

Proof. First, we prove (2.9). The result is immediate if $\mu^*(A) = \infty$, so we suppose that $\mu^*(A)$ is finite. If $A \subset G$, then $\mu^*(A) \le \mu(G)$, so

$$\mu^*(A) \le \inf \{\mu(G) : A \subset G,\, G \text{ open}\},$$

and we just need to prove the reverse inequality,

$$\mu^*(A) \ge \inf \{\mu(G) : A \subset G,\, G \text{ open}\}. \tag{2.11}$$

Let $\epsilon > 0$. There is a cover $\{R_i : i \in \mathbb{N}\}$ of A by rectangles R_i such that

$$\sum_{i=1}^{\infty} \mu(R_i) \le \mu^*(A) + \frac{\epsilon}{2}.$$

Let S_i be an rectangle whose interior S_i° contains R_i such that

$$\mu(S_i) \le \mu(R_i) + \frac{\epsilon}{2^{i+1}}.$$

Then the collection of open rectangles $\{S_i^\circ : i \in \mathbb{N}\}$ covers A and

$$G = \bigcup_{i=1}^{\infty} S_i^\circ$$

is an open set that contains A. Moreover, since $\{S_i : i \in \mathbb{N}\}$ covers G,

$$\mu(G) \le \sum_{i=1}^{\infty} \mu(S_i) \le \sum_{i=1}^{\infty} \mu(R_i) + \frac{\epsilon}{2},$$

and therefore

$$\mu(G) \le \mu^*(A) + \epsilon. \tag{2.12}$$

It follows that

$$\inf \{\mu(G) : A \subset G,\, G \text{ open}\} \le \mu^*(A) + \epsilon,$$

which proves (2.11) since $\epsilon > 0$ is arbitrary.

Next, we prove (2.10). If $K \subset A$, then $\mu(K) \le \mu(A)$, so

$$\sup \{\mu(K) : K \subset A,\, K \text{ compact}\} \le \mu(A).$$

Therefore, we just need to prove the reverse inequality,

$$\mu(A) \le \sup \{\mu(K) : K \subset A,\, K \text{ compact}\}. \tag{2.13}$$

To do this, we apply the previous result to A^c and use the measurability of A.

First, suppose that A is a bounded measurable set, in which case $\mu(A) < \infty$. Let $F \subset \mathbb{R}^n$ be a compact set that contains A. By the preceding result, for any $\epsilon > 0$, there is an open set $G \supset F \setminus A$ such that

$$\mu(G) \le \mu(F \setminus A) + \epsilon.$$

Then $K = F \setminus G$ is a compact set such that $K \subset A$. Moreover, $F \subset K \cup G$ and $F = A \cup (F \setminus A)$, so

$$\mu(F) \le \mu(K) + \mu(G), \qquad \mu(F) = \mu(A) + \mu(F \setminus A).$$

It follows that

$$\begin{aligned}
\mu(A) &= \mu(F) - \mu(F \setminus A) \\
&\le \mu(F) - \mu(G) + \epsilon \\
&\le \mu(K) + \epsilon,
\end{aligned}$$

which implies (2.13) and proves the result for bounded, measurable sets.

Now suppose that A is an unbounded measurable set, and define

$$A_k = \{x \in A : |x| \leq k\}. \tag{2.14}$$

Then $\{A_k : k \in \mathbb{N}\}$ is an increasing sequence of bounded measurable sets whose union is A, so

$$\mu(A_k) \uparrow \mu(A) \qquad \text{as } k \to \infty. \tag{2.15}$$

If $\mu(A) = \infty$, then $\mu(A_k) \to \infty$ as $k \to \infty$. By the previous result, we can find a compact set $K_k \subset A_k \subset A$ such that

$$\mu(K_k) + 1 \geq \mu(A_k)$$

so that $\mu(K_k) \to \infty$. Therefore

$$\sup\{\mu(K) : K \subset A,\, K \text{ compact}\} = \infty,$$

which proves the result in this case.

Finally, suppose that A is unbounded and $\mu(A) < \infty$. From (2.15), for any $\epsilon > 0$ we can choose $k \in \mathbb{N}$ such that

$$\mu(A) \leq \mu(A_k) + \frac{\epsilon}{2}.$$

Moreover, since A_k is bounded, there is a compact set $K \subset A_k$ such that

$$\mu(A_k) \leq \mu(K) + \frac{\epsilon}{2}.$$

Therefore, for every $\epsilon > 0$ there is a compact set $K \subset A$ such that

$$\mu(A) \leq \mu(K) + \epsilon,$$

which gives (2.13), and completes the proof. □

It follows that we may determine the Lebesgue measure of a measurable set in terms of the Lebesgue measure of open or compact sets by approximating the set from the outside by open sets or from the inside by compact sets.

The outer approximation in (2.9) does not require that A is measurable. Thus, for any set $A \subset \mathbb{R}^n$, given $\epsilon > 0$, we can find an open set $G \supset A$ such that $\mu(G) - \mu^*(A) < \epsilon$. If A is measurable, we can strengthen this condition to get that $\mu^*(G \setminus A) < \epsilon$; in fact, this gives a necessary and sufficient condition for measurability.

Theorem 2.24. *A subset $A \subset \mathbb{R}^n$ is Lebesgue measurable if and only if for every $\epsilon > 0$ there is an open set $G \supset A$ such that*

$$\mu^*(G \setminus A) < \epsilon. \tag{2.16}$$

Proof. First we assume that A is measurable and show that it satisfies the condition given in the theorem.

Suppose that $\mu(A) < \infty$ and let $\epsilon > 0$. From (2.12) there is an open set $G \supset A$ such that $\mu(G) < \mu^*(A) + \epsilon$. Then, since A is measurable,

$$\mu^*(G \setminus A) = \mu^*(G) - \mu^*(G \cap A) = \mu(G) - \mu^*(A) < \epsilon,$$

which proves the result when A has finite measure.

If $\mu(A) = \infty$, define $A_k \subset A$ as in (2.14), and let $\epsilon > 0$. Since A_k is measurable with finite measure, the argument above shows that for each $k \in \mathbb{N}$, there is an open set $G_k \supset A_k$ such that

$$\mu(G_k \setminus A_k) < \frac{\epsilon}{2^k}.$$

Then $G = \bigcup_{k=1}^{\infty} G_k$ is an open set that contains A, and

$$\mu^*(G \setminus A) = \mu^*\left(\bigcup_{k=1}^{\infty} G_k \setminus A\right) \leq \sum_{k=1}^{\infty} \mu^*(G_k \setminus A) \leq \sum_{k=1}^{\infty} \mu^*(G_k \setminus A_k) < \epsilon.$$

Conversely, suppose that $A \subset \mathbb{R}^n$ satisfies the condition in the theorem. Let $\epsilon > 0$, and choose an open set $G \supset A$ such that $\mu^*(G \setminus A) < \epsilon$. If $E \subset \mathbb{R}^n$, we have

$$E \cap A^c = (E \cap G^c) \cup (E \cap (G \setminus A)).$$

Hence, by the subadditivity and monotonicity of μ^* and the measurability of G,

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \cap A^c) &\leq \mu^*(E \cap A) + \mu^*(E \cap G^c) + \mu^*(E \cap (G \setminus A)) \\ &\leq \mu^*(E \cap G) + \mu^*(E \cap G^c) + \mu^*(G \setminus A) \\ &< \mu^*(E) + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

which proves that A is measurable. □

This theorem states that a set is Lebesgue measurable if and only if it can be approximated from the outside by an open set in such a way that the difference has arbitrarily small outer Lebesgue measure. This condition can be adopted as the definition of Lebesgue measurable sets, rather than the Carathéodory definition which we have used *c.f.* [**5, 8, 11**].

The following theorem gives another characterization of Lebesgue measurable sets, as ones that can be 'squeezed' between open and closed sets.

Theorem 2.25. *A subset $A \subset \mathbb{R}^n$ is Lebesgue measurable if and only if for every $\epsilon > 0$ there is an open set G and a closed set F such that $G \supset A \supset F$ and*

$$\mu(G \setminus F) < \epsilon. \tag{2.17}$$

If $\mu(A) < \infty$, then F may be chosen to be compact.

Proof. If A satisfies the condition in the theorem, then it follows from the monotonicity of μ^* that $\mu^*(G \setminus A) \leq \mu(G \setminus F) < \epsilon$, so A is measurable by Theorem 2.24.

Conversely, if A is measurable then A^c is measurable, and by Theorem 2.24 given $\epsilon > 0$, there are open sets $G \supset A$ and $H \supset A^c$ such that

$$\mu^*(G \setminus A) < \frac{\epsilon}{2}, \qquad \mu^*(H \setminus A^c) < \frac{\epsilon}{2}.$$

Then, defining the closed set $F = H^c$, we have $G \supset A \supset F$ and

$$\mu(G \setminus F) \leq \mu^*(G \setminus A) + \mu^*(A \setminus F) = \mu^*(G \setminus A) + \mu^*(H \setminus A^c) < \epsilon.$$

Finally, suppose that $\mu(A) < \infty$ and let $\epsilon > 0$. From Theorem 2.23, since A is measurable, there is a compact set $K \subset A$ such that $\mu(A) < \mu(K) + \epsilon/2$ and

$$\mu(A \setminus K) = \mu(A) - \mu(K) < \frac{\epsilon}{2}.$$

As before, from Theorem 2.24 there is an open set $G \supset A$ such that

$$\mu(G) < \mu(A) + \epsilon/2.$$

It follows that $G \supset A \supset K$ and

$$\mu(G \setminus K) = \mu(G \setminus A) + \mu(A \setminus K) < \epsilon,$$

which shows that we may take $F = K$ compact when A has finite measure. □

From the previous results, we can approximate measurable sets by open or closed sets, up to sets of arbitrarily small but, in general, nonzero measure. By taking countable intersections of open sets or countable unions of closed sets, we can approximate measurable sets by Borel sets, up to sets of measure zero

Definition 2.26. The collection of sets in $\mathbb{R}^n$ that are countable intersections of open sets is denoted by $G_\delta(\mathbb{R}^n)$, and the collection of sets in $\mathbb{R}^n$ that are countable unions of closed sets is denoted by $F_\sigma(\mathbb{R}^n)$.

G_δ and F_σ sets are Borel. Thus, it follows from the next result that every Lebesgue measurable set can be approximated up to a set of measure zero by a Borel set. This is the Borel regularity of Lebesgue measure.

Theorem 2.27. *Suppose that $A \subset \mathbb{R}^n$ is Lebesgue measurable. Then there exist sets $G \in G_\delta(\mathbb{R}^n)$ and $F \in F_\sigma(\mathbb{R}^n)$ such that*

$$G \supset A \supset F, \qquad \mu(G \setminus A) = \mu(A \setminus F) = 0.$$

Proof. For each $k \in \mathbb{N}$, choose an open set G_k and a closed set F_k such that $G_k \supset A \supset F_k$ and

$$\mu(G_k \setminus F_k) \le \frac{1}{k}$$

Then

$$G = \bigcap_{k=1}^{\infty} G_k, \qquad F = \bigcup_{k=1}^{\infty} F_k$$

are G_δ and F_σ sets with the required properties. □

In particular, since any measurable set can be approximated up to a set of measure zero by a G_δ or an F_σ, the complexity of the transfinite construction of general Borel sets illustrated in (2.8) is 'hidden' inside sets of Lebesgue measure zero.

As a corollary of this result, we get that the Lebesgue σ-algebra is the completion of the Borel σ-algebra with respect to Lebesgue measure.

Theorem 2.28. *The Lebesgue σ-algebra $\mathcal{L}(\mathbb{R}^n)$ is the completion of the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$.*

Proof. Lebesgue measure is complete from Proposition 2.12. By the previous theorem, if $A \subset \mathbb{R}^n$ is Lebesgue measurable, then there is a F_σ set $F \subset A$ such that $M = A \setminus F$ has Lebesgue measure zero. It follows by the approximation theorem that there is a Borel set $N \in G_\delta$ with $\mu(N) = 0$ and $M \subset N$. Thus, $A = F \cup M$ where $F \in \mathcal{B}$ and $M \subset N \in \mathcal{B}$ with $\mu(N) = 0$, which proves that $\mathcal{L}(\mathbb{R}^n)$ is the completion of $\mathcal{B}(\mathbb{R}^n)$ as given in Theorem 1.15. □

2.8. Linear transformations

The definition of Lebesgue measure is not rotationally invariant, since we used rectangles whose sides are parallel to the coordinate axes. In this section, we show that the resulting measure does not, in fact, depend upon the direction of the coordinate axes and is invariant under orthogonal transformations. We also show that Lebesgue measure transforms under a linear map by a factor equal to the absolute value of the determinant of the map.

As before, we use μ^* to denote Lebesgue outer measure defined using rectangles whose sides are parallel to the coordinate axes; a set is Lebesgue measurable if it satisfies the Carathéodory criterion (2.8) with respect to this outer measure. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map and $E \subset \mathbb{R}^n$, we denote the image of E under T by

$$TE = \{Tx \in \mathbb{R}^n : x \in E\}.$$

First, we consider the Lebesgue measure of rectangles whose sides are not parallel to the coordinate axes. We use a tilde to denote such rectangles by $\tilde{R}$; we denote closed rectangles whose sides are parallel to the coordinate axes by R as before. We refer to $\tilde{R}$ and R as oblique and parallel rectangles, respectively. We denote the volume of a rectangle $\tilde{R}$ by $v(\tilde{R})$, *i.e.* the product of the lengths of its sides, to avoid confusion with its Lebesgue measure $\mu(\tilde{R})$. We know that $\mu(R) = v(R)$ for parallel rectangles, and that $\tilde{R}$ is measurable since it is closed, but we have not yet shown that $\mu(\tilde{R}) = v(\tilde{R})$ for oblique rectangles.

More explicitly, we regard $\mathbb{R}^n$ as a Euclidean space equipped with the standard inner product,

$$(x, y) = \sum_{i=1}^{n} x_i y_i, \qquad x = (x_1, x_2, \dots, x_n), \quad y = (y_1, y_2, \dots, y_n).$$

If $\{e_1, e_2, \dots, e_n\}$ is the standard orthonormal basis of $\mathbb{R}^n$,

$$e_1 = (1, 0, \dots, 0), \quad e_2 = (0, 1, \dots, 0), \dots e_n = (0, 0, \dots, 1),$$

and $\{\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_n\}$ is another orthonormal basis, then we use R to denote rectangles whose sides are parallel to $\{e_i\}$ and $\tilde{R}$ to denote rectangles whose sides are parallel to $\{\tilde{e}_i\}$. The linear map $Q : \mathbb{R}^n \to \mathbb{R}^n$ defined by $Qe_i = \tilde{e}_i$ is orthogonal, meaning that $Q^T = Q^{-1}$ and

$$(Qx, Qy) = (x, y) \qquad \text{for all } x, y \in \mathbb{R}^n.$$

Since Q preserves lengths and angles, it maps a rectangle R to a rectangle $\tilde{R} = QR$ such that $v(\tilde{R}) = v(R)$.

We will use the following lemma.

Lemma 2.29. *If an oblique rectangle $\tilde{R}$ contains a finite almost disjoint collection of parallel rectangles $\{R_1, R_2, \dots, R_N\}$ then*

$$\sum_{i=1}^{N} v(R_i) \le v(\tilde{R}).$$

This result is geometrically obvious, but a formal proof seems to require a fuller discussion of the volume function on elementary geometrical sets, which is included in the theory of valuations in convex geometry. We omit the details.

Proposition 2.30. *If $\tilde{R}$ is an oblique rectangle, then given any $\epsilon > 0$ there is a collection of parallel rectangles $\{R_i : i \in \mathbb{N}\}$ that covers $\tilde{R}$ and satisfies*

$$\sum_{i=1}^{\infty} v(R_i) \leq v(\tilde{R}) + \epsilon.$$

Proof. Let $\tilde{S}$ be an oblique rectangle that contains $\tilde{R}$ in its interior such that

$$v(\tilde{S}) \leq v(\tilde{R}) + \epsilon.$$

Then, from Proposition 2.20, we may decompose the interior of S into an almost disjoint union of parallel rectangles

$$\tilde{S}^\circ = \bigcup_{i=1}^{\infty} R_i.$$

It follows from the previous lemma that for every $N \in \mathbb{N}$

$$\sum_{i=1}^{N} v(R_i) \leq v(\tilde{S}),$$

which implies that

$$\sum_{i=1}^{\infty} v(R_i) \leq v(\tilde{S}) \leq v(\tilde{R}) + \epsilon.$$

Moreover, the collection $\{R_i\}$ covers $\tilde{R}$ since its union is $\tilde{S}^\circ$, which contains $\tilde{R}$. □

Conversely, by reversing the roles of the axes, we see that if R is a parallel rectangle and $\epsilon > 0$, then there is a cover of R by oblique rectangles $\{\tilde{R}_i : i \in \mathbb{N}\}$ such that

$$\sum_{i=1}^{\infty} v(\tilde{R}_i) \leq v(R) + \epsilon. \tag{2.18}$$

Theorem 2.31. *If $E \subset \mathbb{R}^n$ and $Q : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation, then*

$$\mu^*(QE) = \mu^*(E),$$

and E is Lebesgue measurable if an only if QE is Lebesgue measurable.

Proof. Let $\tilde{E} = QE$. Given $\epsilon > 0$ there is a cover of $\tilde{E}$ by parallel rectangles $\{R_i : i \in \mathbb{N}\}$ such that

$$\sum_{i=1}^{\infty} v(R_i) \leq \mu^*(\tilde{E}) + \frac{\epsilon}{2}.$$

From (2.18), for each $i \in \mathbb{N}$ we can choose a cover $\{\tilde{R}_{i,j} : j \in \mathbb{N}\}$ of R_i by oblique rectangles such that

$$\sum_{i=1}^{\infty} v(\tilde{R}_{i,j}) \leq v(R_i) + \frac{\epsilon}{2^{i+1}}.$$

Then $\{\tilde{R}_{i,j} : i, j \in \mathbb{N}\}$ is a countable cover of $\tilde{E}$ by oblique rectangles, and

$$\sum_{i,j=1}^{\infty} v(\tilde{R}_{i,j}) \leq \sum_{i=1}^{\infty} v(R_i) + \frac{\epsilon}{2} \leq \mu^*(\tilde{E}) + \epsilon.$$

If $R_{i,j} = Q^T \tilde{R}_{i,j}$, then $\{R_{i,j} : j \in \mathbb{N}\}$ is a cover of E by parallel rectangles, so

$$\mu^*(E) \leq \sum_{i,j=1}^{\infty} v(R_{i,j}).$$

Moreover, since Q is orthogonal, we have $v(R_{i,j}) = v(\tilde{R}_{i,j})$. It follows that

$$\mu^*(E) \leq \sum_{i,j=1}^{\infty} v(R_{i,j}) = \sum_{i,j=1}^{\infty} v(\tilde{R}_{i,j}) \leq \mu^*(\tilde{E}) + \epsilon,$$

and since $\epsilon > 0$ is arbitrary, we conclude that

$$\mu^*(E) \leq \mu^*(\tilde{E}).$$

By applying the same argument to the inverse mapping $E = Q^T \tilde{E}$, we get the reverse inequality, and it follows that $\mu^*(E) = \mu^*(\tilde{E})$.

Since μ^* is invariant under Q, the Carathéodory criterion for measurability is invariant, and E is measurable if and only if QE is measurable. □

It follows from Theorem 2.31 that Lebesgue measure is invariant under rotations and reflections.[4] Since it is also invariant under translations, Lebesgue measure is invariant under all isometries of $\mathbb{R}^n$.

Next, we consider the effect of dilations on Lebesgue measure. Arbitrary linear maps may then be analyzed by decomposing them into rotations and dilations.

Proposition 2.32. *Suppose that* $\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ *is the linear transformation*

$$\Lambda : (x_1, x_2, \dots, x_n) \mapsto (\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_n x_n) \tag{2.19}$$

where the $\lambda_i > 0$ *are positive constants. Then*

$$\mu^*(\Lambda E) = (\det \Lambda)\mu^*(E),$$

and E *is Lebesgue measurable if and only if* ΛE *is Lebesgue measurable.*

Proof. The diagonal map Λ does not change the orientation of a rectangle, so it maps a cover of E by parallel rectangles to a cover of ΛE by parallel rectangles, and conversely. Moreover, Λ multiplies the volume of a rectangle by $\det \Lambda = \lambda_1 \dots \lambda_n$, so it immediate from the definition of outer measure that $\mu^*(\Lambda E) = (\det \Lambda)\mu^*(E)$, and E satisfies the Carathéodory criterion for measurability if and only if ΛE does. □

Theorem 2.33. *Suppose that* $T : \mathbb{R}^n \to \mathbb{R}^n$ *is a linear transformation and* $E \subset \mathbb{R}^n$*. Then*

$$\mu^*(TE) = |\det T| \, \mu^*(E),$$

and TE *is Lebesgue measurable if* E *is measurable*

Proof. If T is singular, then its range is a lower-dimensional subspace of $\mathbb{R}^n$, which has Lebesgue measure zero, and its determinant is zero, so the result holds.[5] We therefore assume that T is nonsingular.

[4] Unlike differential volume forms, Lebesgue measure does not depend on the orientation of $\mathbb{R}^n$; such measures are sometimes referred to as densities in differential geometry.

[5] In this case TE, is always Lebesgue measurable, with measure zero, even if E is not measurable.

In that case, according to the polar decomposition, the map T may be written as a composition

$$T = QU$$

of a positive definite, symmetric map $U = \sqrt{T^T T}$ and an orthogonal map Q. Any positive-definite, symmetric map U may be diagonalized by an orthogonal map O to get

$$U = O^T \Lambda O$$

where $\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ has the form (2.19). From Theorem 2.31, orthogonal mappings leave the Lebesgue measure of a set invariant, so from Proposition 2.32

$$\mu^*(TE) = \mu^*(\Lambda E) = (\det \Lambda)\mu^*(E).$$

Since $|\det Q| = 1$ for any orthogonal map Q, we have $\det \Lambda = |\det T|$, and it follows that $\mu^*(TE) = |\det T|\, \mu^*(E)$.

Finally, it is straightforward to see that TE is measurable if E is measurable. □

2.9. Lebesgue-Stieltjes measures

We briefly consider a generalization of one-dimensional Lebesgue measure, called Lebesgue-Stieltjes measures on $\mathbb{R}$. These measures are obtained from an increasing, right-continuous function $F : \mathbb{R} \to \mathbb{R}$, and assign to a half-open interval $(a, b]$ the measure

$$\mu_F\left((a, b]\right) = F(b) - F(a).$$

The use of half-open intervals is significant here because a Lebesgue-Stieltjes measure may assign nonzero measure to a single point. Thus, unlike Lebesgue measure, we need not have $\mu_F([a, b]) = \mu_F((a, b])$. Half-open intervals are also convenient because the complement of a half-open interval is a finite union of (possibly infinite) half-open intervals of the same type. Thus, the collection of finite unions of half-open intervals forms an algebra.

The right-continuity of F is consistent with the use of intervals that are half-open at the left, since

$$\bigcap_{i=1}^{\infty} (a, a + 1/i] = \varnothing,$$

so, from (1.2), if F is to define a measure we need

$$\lim_{i\to\infty} \mu_F\left((a, a + 1/i]\right) = 0$$

or

$$\lim_{i\to\infty} \left[F(a + 1/i) - F(a)\right] = \lim_{x\to a^+} F(x) - F(a) = 0.$$

Conversely, as we state in the next theorem, any such function F defines a Borel measure on $\mathbb{R}$.

Theorem 2.34. *Suppose that $F : \mathbb{R} \to \mathbb{R}$ is an increasing, right-continuous function. Then there is a unique Borel measure $\mu_F : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ such that*

$$\mu_F\left((a, b]\right) = F(b) - F(a)$$

for every $a < b$.

The construction of μ_F is similar to the construction of Lebesgue measure on $\mathbb{R}^n$. We define an outer measure $\mu_F^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ by

$$\mu_F^*(E) = \inf \left\{ \sum_{i=1}^{\infty} [F(b_i) - F(a_i)] : E \subset \bigcup_{i=1}^{\infty} (a_i, b_i] \right\},$$

and restrict μ_F^* to its Carathéodory measurable sets, which include the Borel sets. See *e.g.* Section 1.5 of Folland [**4**] for a detailed proof.

The following examples illustrate the three basic types of Lebesgue-Stieltjes measures.

Example 2.35. If $F(x) = x$, then μ_F is Lebesgue measure on $\mathbb{R}$ with

$$\mu_F\left((a, b]\right) = b - a.$$

Example 2.36. If

$$F(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

then μ_F is the δ-measure supported at 0,

$$\mu_F(A) = \begin{cases} 1 & \text{if } 0 \in A, \\ 0 & \text{if } 0 \notin A. \end{cases}$$

Example 2.37. If $F : \mathbb{R} \to \mathbb{R}$ is the Cantor function, then μ_F assigns measure one to the Cantor set, which has Lebesgue measure zero, and measure zero to its complement. Despite the fact that μ_F is supported on a set of Lebesgue measure zero, the μ_F-measure of any countable set is zero.

CHAPTER 3

Measurable functions

Measurable functions in measure theory are analogous to continuous functions in topology. A continuous function pulls back open sets to open sets, while a measurable function pulls back measurable sets to measurable sets.

3.1. Measurability

Most of the theory of measurable functions and integration does not depend on the specific features of the measure space on which the functions are defined, so we consider general spaces, although one should keep in mind the case of functions defined on $\mathbb{R}$ or $\mathbb{R}^n$ equipped with Lebesgue measure.

Definition 3.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A function $f : X \to Y$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$.

Note that the measurability of a function depends only on the σ-algebras; it is not necessary that any measures are defined.

In order to show that a function is measurable, it is sufficient to check the measurability of the inverse images of sets that generate the σ-algebra on the target space.

Proposition 3.2. *Suppose that $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces and $\mathcal{B} = \sigma(\mathcal{G})$ is generated by a family $\mathcal{G} \subset \mathcal{P}(Y)$. Then $f : X \to Y$ is measurable if and only if*

$$f^{-1}(G) \in \mathcal{A} \qquad \text{for every } G \in \mathcal{G}.$$

Proof. Set operations are natural under pull-backs, meaning that

$$f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$$

and

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i), \quad f^{-1}\left(\bigcap_{i=1}^{\infty} B_i\right) = \bigcap_{i=1}^{\infty} f^{-1}(B_i).$$

It follows that

$$\mathcal{M} = \left\{B \subset Y : f^{-1}(B) \in \mathcal{A}\right\}$$

is a σ-algebra on Y. By assumption, $\mathcal{M} \supset \mathcal{G}$ and therefore $\mathcal{M} \supset \sigma(\mathcal{G}) = \mathcal{B}$, which implies that f is measurable. □

It is worth noting the indirect nature of the proof of containment of σ-algebras in the previous proposition; this is required because we typically cannot use an explicit representation of sets in a σ-algebra. For example, the proof does not characterize $\mathcal{M}$, which may be strictly larger than $\mathcal{B}$.

If the target space Y is a topological space, then we always equip it with the Borel σ-algebra $\mathcal{B}(Y)$ generated by the open sets (unless stated explicitly otherwise).

In that case, it follows from Proposition 3.2 that $f : X \to Y$ is measurable if and only if $f^{-1}(G) \in \mathcal{A}$ is a measurable subset of X for every set G that is open in Y. In particular, every continuous function between topological spaces that are equipped with their Borel σ-algebras is measurable. The class of measurable function is, however, typically much larger than the class of continuous functions, since we only require that the inverse image of an open set is Borel; it need not be open.

3.2. Real-valued functions

We specialize to the case of real-valued functions

$$f : X \to \mathbb{R}$$

or extended real-valued functions

$$f : X \to \overline{\mathbb{R}}.$$

We will consider one case or the other as convenient, and comment on any differences. A positive extended real-valued function is a function

$$f : X \to [0, \infty].$$

Note that we allow a positive function to take the value zero.

We equip $\mathbb{R}$ and $\overline{\mathbb{R}}$ with their Borel σ-algebras $\mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\overline{\mathbb{R}})$. A Borel subset of $\overline{\mathbb{R}}$ has one of the forms

$$B, \qquad B \cup \{\infty\}, \qquad B \cup \{-\infty\}, \qquad B \cup \{-\infty, \infty\}$$

where B is a Borel subset of $\mathbb{R}$. As Example 2.22 shows, sets that are Lebesgue measurable but not Borel measurable need not be well-behaved under the inverse of even a monotone function, which helps explain why we do not include them in the range σ-algebra on $\mathbb{R}$ or $\overline{\mathbb{R}}$.

By contrast, when the domain of a function is a measure space it is often convenient to use a complete space. For example, if the domain is $\mathbb{R}^n$ we typically equip it with the Lebesgue σ-algebra, although if completeness is not required we may use the Borel σ-algebra. With this understanding, we get the following definitions. We state them for real-valued functions; the definitions for extended real-valued functions are completely analogous

Definition 3.3. If $(X, \mathcal{A})$ is a measurable space, then $f : X \to \mathbb{R}$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for every Borel set $B \in \mathcal{B}(\mathbb{R})$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue measurable if $f^{-1}(B)$ is a Lebesgue measurable subset of $\mathbb{R}^n$ for every Borel subset B of $\mathbb{R}$, and it is Borel measurable if $f^{-1}(B)$ is a Borel measurable subset of $\mathbb{R}^n$ for every Borel subset B of $\mathbb{R}$

This definition ensures that continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ are Borel measurable and functions that are equal a.e. to Borel measurable functions are Lebesgue measurable. If $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable and $g : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue (or Borel) measurable, then the composition $f \circ g$ is Lebesgue (or Borel) measurable since

$$(f \circ g)^{-1}(B) = g^{-1}\left(f^{-1}(B)\right).$$

Note that if f is Lebesgue measurable, then $f \circ g$ need not be measurable since $f^{-1}(B)$ need not be Borel even if B is Borel.

We can give more easily verifiable conditions for measurability in terms of generating families for Borel sets.

Proposition 3.4. *The Borel σ-algebra on $\mathbb{R}$ is generated by any of the following collections of intervals*

$$\{(-\infty, b) : b \in \mathbb{R}\}, \quad \{(-\infty, b] : b \in \mathbb{R}\}, \quad \{(a, \infty) : a \in \mathbb{R}\}, \quad \{[a, \infty) : a \in \mathbb{R}\}.$$

Proof. The σ-algebra generated by intervals of the form $(-\infty, b)$ is contained in the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ since the intervals are open sets. Conversely, the σ-algebra contains complementary closed intervals of the form $[a, \infty)$, half-open intersections $[a, b)$, and countable intersections

$$[a, b] = \bigcap_{n=1}^{\infty} [a, b + \frac{1}{n}).$$

From Proposition 2.20, the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by the collection of closed rectangles $[a, b]$, so

$$\sigma\left(\{(-\infty, b) : b \in \mathbb{R}\}\right) = \mathcal{B}(\mathbb{R}).$$

The proof for the other collections is similar. □

The properties given in the following proposition are sometimes taken as the definition of a measurable function.

Proposition 3.5. *If $(X, \mathcal{A})$ is a measurable space, then $f : X \to \mathbb{R}$ is measurable if and only if one of the following conditions holds:*

$$\begin{aligned} &\{x \in X : f(x) < b\} \in \mathcal{A} &&\text{for every } b \in \mathbb{R};\\ &\{x \in X : f(x) \le b\} \in \mathcal{A} &&\text{for every } b \in \mathbb{R};\\ &\{x \in X : f(x) > a\} \in \mathcal{A} &&\text{for every } a \in \mathbb{R};\\ &\{x \in X : f(x) \ge a\} \in \mathcal{A} &&\text{for every } a \in \mathbb{R}.\end{aligned}$$

Proof. Note that, for example,

$$\{x \in X : f(x) < b\} = f^{-1}\left((-\infty, b)\right)$$

and the result follows immediately from Propositions 3.2 and 3.4. □

If any one of these equivalent conditions holds, then $f^{-1}(B) \in \mathcal{A}$ for every set $B \in \mathcal{B}(\mathbb{R})$. We will often use a shorthand notation for sets, such as

$$\{f < b\} = \{x \in X : f(x) < b\}.$$

The Borel σ-algebra on $\overline{\mathbb{R}}$ is generated by intervals of the form $[-\infty, b)$, $[-\infty, b]$, $(a, \infty]$, or $[a, \infty]$ where $a, b \in \mathbb{R}$, and exactly the same conditions as the ones in Proposition 3.5 imply the measurability of an extended real-valued functions $f : X \to \overline{\mathbb{R}}$. In that case, we can allow $a, b \in \overline{\mathbb{R}}$ to be extended real numbers in Proposition 3.5, but it is not necessary to do so in order to imply that f is measurable.

Measurability is well-behaved with respect to algebraic operations.

Proposition 3.6. *If $f, g : X \to \mathbb{R}$ are real-valued measurable functions and $k \in \mathbb{R}$, then*

$$kf, f + g, \qquad fg, \qquad f/g$$

are measurable functions, where we assume that $g \neq 0$ in the case of f/g.

PROOF. If $k > 0$, then

$$\{kf < b\} = \{f < b/k\}$$

so kf is measurable, and similarly if $k < 0$ or $k = 0$. We have

$$\{f + g < b\} = \bigcup_{q+r<b;q,r\in\mathbb{Q}} \{f < q\} \cap \{g < r\}$$

so $f + g$ is measurable. The function f^2 is measurable since if $b \geq 0$

$$\left\{f^2 < b\right\} = \left\{-\sqrt{b} < f < \sqrt{b}\right\}.$$

It follows that

$$fg = \frac{1}{2}\left[(f + g)^2 - f^2 - g^2\right]$$

is measurable. Finally, if $g \neq 0$

$$\{1/g < b\} = \begin{cases} \{1/b < g < 0\} & \text{if } b < 0, \\ \{-\infty < g < 0\} & \text{if } b = 0, \\ \{-\infty < g < 0\} \cup \{1/b < g < \infty\} & \text{if } b > 0, \end{cases}$$

so $1/g$ is measurable and therefore f/g is measurable. □

An analogous result applies to extended real-valued functions provided that they are well-defined. For example, $f + g$ is measurable provided that $f(x)$, $g(x)$ are not simultaneously equal to ∞ and $-\infty$, and fg is is measurable provided that $f(x)$, $g(x)$ are not simultaneously equal to 0 and $\pm\infty$.

Proposition 3.7. *If $f, g : X \to \overline{\mathbb{R}}$ are extended real-valued measurable functions, then*

$$|f|, \qquad \max(f, g), \qquad \min(f, g)$$

are measurable functions.

PROOF. We have

$$\{\max(f, g) < b\} = \{f < b\} \cap \{g < b\},$$
$$\{\min(f, g) < b\} = \{f < b\} \cup \{g < b\},$$

and $|f| = \max(f, 0) - \min(f, 0)$, from which the result follows. □

3.3. Pointwise convergence

Crucially, measurability is preserved by limiting operations on sequences of functions. Operations in the following theorem are understood in a pointwise sense; for example,

$$\left(\sup_{n\in\mathbb{N}} f_n\right)(x) = \sup_{n\in\mathbb{N}} \{f_n(x)\}.$$

Theorem 3.8. *If $\{f_n : n \in \mathbb{N}\}$ is a sequence of measurable functions $f_n : X \to \overline{\mathbb{R}}$, then*

$$\sup_{n\in\mathbb{N}} f_n, \qquad \inf_{n\in\mathbb{N}} f_n, \qquad \limsup_{n\to\infty} f_n, \qquad \liminf_{n\to\infty} f_n$$

are measurable extended real-valued functions on X.

Proof. We have for every $b \in \mathbb{R}$ that

$$\left\{\sup_{n\in\mathbb{N}} f_n \leq b\right\} = \bigcap_{n=1}^{\infty} \{f_n \leq b\},$$
$$\left\{\inf_{n\in\mathbb{N}} f_n < b\right\} = \bigcup_{n=1}^{\infty} \{f_n < b\}$$

so the supremum and infimum are measurable Moreover, since

$$\limsup_{n\to\infty} f_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} f_k,$$
$$\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} \inf_{k\geq n} f_k$$

it follows that the limsup and liminf are measurable. □

Perhaps the most important way in which new functions arise from old ones is by pointwise convergence.

Definition 3.9. A sequence $\{f_n : n \in \mathbb{N}\}$ of functions $f_n : X \to \overline{\mathbb{R}}$ *converges pointwise* to a function $f : X \to \overline{\mathbb{R}}$ if $f_n(x) \to f(x)$ as $n \to \infty$ for every $x \in X$.

Pointwise convergence preserves measurability (unlike continuity, for example). This fact explains why the measurable functions form a sufficiently large class for the needs of analysis.

Theorem 3.10. *If $\{f_n : n \in \mathbb{N}\}$ is a sequence of measurable functions $f_n : X \to \overline{\mathbb{R}}$ and $f_n \to f$ pointwise as $n \to \infty$, then $f : X \to \overline{\mathbb{R}}$ is measurable.*

Proof. If $f_n \to f$ pointwise, then

$$f = \limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n$$

so the result follows from the previous proposition. □

3.4. Simple functions

The *characteristic function* (or *indicator function*) of a subset $E \subset X$ is the function $\chi_E : X \to \mathbb{R}$ defined by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

The function χ_E is measurable if and only if E is a measurable set.

Definition 3.11. A *simple function* $\phi : X \to \mathbb{R}$ on a measurable space $(X, \mathcal{A})$ is a function of the form

$$\phi(x) = \sum_{n=1}^{N} c_n \chi_{E_n}(x) \tag{3.1}$$

where $c_1, \dots, c_N \in \mathbb{R}$ and $E_1, \dots, E_N \in \mathcal{A}$.

Note that, according to this definition, a simple function is measurable. The representation of ϕ in (3.1) is not unique; we call it a standard representation if the constants c_n are distinct and the sets E_n are disjoint.

Theorem 3.12. *If $f : X \to [0, \infty]$ is a positive measurable function, then there is a monotone increasing sequence of positive simple functions $\phi_n : X \to [0, \infty)$ with $\phi_1 \le \phi_2 \le \dots \le \phi_n \le \dots$ such that $\phi_n \to f$ pointwise as $n \to \infty$. If f is bounded, then $\phi_n \to f$ uniformly.*

PROOF. For each $n \in \mathbb{N}$, we divide the interval $[0, 2^n]$ in the range of f into 2^{2n} subintervals of width 2^{-n},

$$I_{k,n} = (k2^{-n}, (k+1)2^{-n}], \qquad k = 0, 1, \dots, 2^{2n} - 1,$$

let $J_n = (2^n, \infty]$ be the remaining part of the range, and define

$$E_{k,n} = f^{-1}(I_{k,n}), \qquad F_n = f^{-1}(J_n).$$

Then the sequence of simple functions given by

$$\phi_n = \sum_{k=0}^{2^n-1} k2^{-n}\chi_{E_{k,n}} + 2^n\chi_{F_n}$$

has the required properties. □

In defining the Lebesgue integral of a measurable function, we will approximate it by simple functions. By contrast, in defining the Riemann integral of a function $f : [a, b] \to \mathbb{R}$, we partition the domain $[a, b]$ into subintervals and approximate f by step functions that are constant on these subintervals. This difference is sometime expressed by saying that in the Lebesgue integral we partition the range, and in the Riemann integral we partition the domain.

3.5. Properties that hold almost everywhere

Often, we want to consider functions or limits which are defined outside a set of measure zero. In that case, it is convenient to deal with complete measure spaces.

Proposition 3.13. *Let $(X, \mathcal{A}, \mu)$ be a complete measure space and $f, g : X \to \overline{\mathbb{R}}$. If $f = g$ pointwise μ-a.e. and f is measurable, then g is measurable.*

PROOF. Suppose that $f = g$ on N^c where N is a set of measure zero. Then

$$\{g < b\} = (\{f < b\} \cap N^c) \cup (\{g < b\} \cap N).$$

Each of these sets is measurable: $\{f < b\}$ is measurable since f is measurable; and $\{g < b\} \cap N$ is measurable since it is a subset of a set of measure zero and X is complete. □

The completeness of X is essential in this proposition. For example, if X is not complete and $E \subset N$ is a non-measurable subset of a set N of measure zero, then the functions 0 and χ_E are equal almost everywhere, but 0 is measurable and χ_E is not.

Proposition 3.14. *Let $(X, \mathcal{A}, \mu)$ be a complete measure space. If $\{f_n : n \in \mathbb{N}\}$ is a sequence of measurable functions $f_n : X \to \overline{\mathbb{R}}$ and $f_n \to f$ as $n \to \infty$ pointwise μ-a.e., then f is measurable.*

PROOF. Since f_n is measurable, $g = \limsup_{n\to\infty} f_n$ is measurable and $f = g$ pointwise a.e., so the result follows from the previous proposition. □

CHAPTER 4

Integration

In this Chapter, we define the integral of real-valued functions on an arbitrary measure space and derive some of its basic properties. We refer to this integral as the Lebesgue integral, whether or not the domain of the functions is subset of $\mathbb{R}^n$ equipped with Lebesgue measure. The Lebesgue integral applies to a much wider class of functions than the Riemann integral and is better behaved with respect to pointwise convergence. We carry out the definition in three steps: first for positive simple functions, then for positive measurable functions, and finally for extended real-valued measurable functions.

4.1. Simple functions

Suppose that $(X, \mathcal{A}, \mu)$ is a measure space.

Definition 4.1. If $\phi : X \to [0, \infty)$ is a positive simple function, given by

$$\phi = \sum_{i=1}^{N} c_i \chi_{E_i}$$

where $c_i \geq 0$ and $E_i \in \mathcal{A}$, then the integral of ϕ with respect to μ is

$$\int \phi \, d\mu = \sum_{i=1}^{N} c_i \mu(E_i) . \tag{4.1}$$

In (4.1), we use the convention that if $c_i = 0$ and $\mu(E_i) = \infty$, then $0 \cdot \infty = 0$, meaning that the integral of 0 over a set of measure ∞ is equal to 0. The integral may take the value ∞ (if $c_i > 0$ and $\mu(E_i) = \infty$ for some $1 \leq i \leq N$). One can verify that the value of the integral in (4.1) is independent of how the simple function is represented as a linear combination of characteristic functions.

Example 4.2. The characteristic function $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ of the rationals is not Riemann integrable on any compact interval of non-zero length, but it is Lebesgue integrable with

$$\int \chi_{\mathbb{Q}} \, d\mu = 1 \cdot \mu(\mathbb{Q}) = 0.$$

The integral of simple functions has the usual properties of an integral. In particular, it is linear, positive, and monotone.

Proposition 4.3. *If $\phi, \psi : X \to [0, \infty)$ are positive simple functions on a measure space X, then:*

$$\int k\phi \, d\mu = k \int \phi \, d\mu \qquad \text{if } k \in [0, \infty);$$
$$\int (\phi + \psi) \, d\mu = \int \phi \, d\mu + \int \psi \, d\mu;$$
$$0 \leq \int \phi \, d\mu \leq \int \psi \, d\mu \qquad \text{if } 0 \leq \phi \leq \psi.$$

Proof. These follow immediately from the definition. □

4.2. Positive functions

We define the integral of a measurable function by splitting it into positive and negative parts, so we begin by defining the integral of a positive function.

Definition 4.4. If $f : X \to [0, \infty]$ is a positive, measurable, extended real-valued function on a measure space X, then

$$\int f \, d\mu = \sup \left\{ \int \phi \, d\mu : 0 \leq \phi \leq f, \, \phi \text{ simple} \right\}.$$

A positive function $f : X \to [0, \infty]$ is integrable if it is measurable and

$$\int f \, d\mu < \infty.$$

In this definition, we approximate the function f from below by simple functions. In contrast with the definition of the Riemann integral, it is not necessary to approximate a measurable function from both above and below in order to define its integral.

If $A \subset X$ is a measurable set and $f : X \to [0, \infty]$ is measurable, we define

$$\int_A f \, d\mu = \int f \chi_A \, d\mu.$$

Unlike the Riemann integral, where the definition of the integral over non-rectangular subsets of $\mathbb{R}^2$ already presents problems, it is trivial to define the Lebesgue integral over arbitrary measurable subsets of a set on which it is already defined.

The following properties are an immediate consequence of the definition and the corresponding properties of simple functions.

Proposition 4.5. *If $f, g : X \to [0, \infty]$ are positive, measurable, extended real-valued function on a measure space X, then:*

$$\int kf \, d\mu = k \int f \, d\mu \qquad \text{if } k \in [0, \infty);$$
$$0 \leq \int f \, d\mu \leq \int g \, d\mu \qquad \text{if } 0 \leq f \leq g.$$

The integral is also linear, but this is not immediately obvious from the definition and it depends on the measurability of the functions. To show the linearity, we will first derive one of the fundamental convergence theorem for the Lebesgue integral, the monotone convergence theorem. We discuss this theorem and its applications in greater detail in Section 4.5.

Theorem 4.6 (Monotone Convergence Theorem)**.** *If $\{f_n : n \in \mathbb{N}\}$ is a monotone increasing sequence*

$$0 \le f_1 \le f_2 \le \dots \le f_n \le f_{n+1} \le \dots$$

of positive, measurable, extended real-valued functions $f_n : X \to [0, \infty]$ and

$$f = \lim_{n\to\infty} f_n,$$

then

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Proof. The pointwise limit $f : X \to [0, \infty]$ exists since the sequence $\{f_n\}$ is increasing. Moreover, by the monotonicity of the integral, the integrals are increasing, and

$$\int f_n \, d\mu \le \int f_{n+1} \, d\mu \le \int f \, d\mu,$$

so the limit of the integrals exists, and

$$\lim_{n\to\infty} \int f_n \, d\mu \le \int f \, d\mu.$$

To prove the reverse inequality, let $\phi : X \to [0, \infty)$ be a simple function with $0 \le \phi \le f$. Fix $0 < t < 1$. Then

$$A_n = \{x \in X : f_n(x) \ge t\phi(x)\}$$

is an increasing sequence of measurable sets $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$ whose union is X. It follows that

$$\int f_n \, d\mu \ge \int_{A_n} f_n \, d\mu \ge t \int_{A_n} \phi \, d\mu. \tag{4.2}$$

Moreover, if

$$\phi = \sum_{i=1}^{N} c_i \chi_{E_i}$$

we have from the monotonicity of μ in Proposition 1.12 that

$$\int_{A_n} \phi \, d\mu = \sum_{i=1}^{N} c_i \mu(E_i \cap A_n) \to \sum_{i=1}^{N} c_i \mu(E_i) = \int \phi \, d\mu$$

as $n \to \infty$. Taking the limit as $n \to \infty$ in (4.2), we therefore get

$$\lim_{n\to\infty} \int f_n \, d\mu \ge t \int \phi \, d\mu.$$

Since $0 < t < 1$ is arbitrary, we conclude that

$$\lim_{n\to\infty} \int f_n \, d\mu \ge \int \phi \, d\mu,$$

and since $\phi \le f$ is an arbitrary simple function, we get by taking the supremum over ϕ that

$$\lim_{n\to\infty} \int f_n \, d\mu \ge \int f \, d\mu.$$

This proves the theorem. □

In particular, this theorem implies that we can obtain the integral of a positive measurable function f as a limit of integrals of an increasing sequence of simple functions, not just as a supremum over all simple functions dominated by f as in Definition 4.4. As shown in Theorem 3.12, such a sequence of simple functions always exists.

Proposition 4.7. *If $f, g : X \to [0,\infty]$ are positive, measurable functions on a measure space X, then*

$$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu.$$

Proof. Let $\{\phi_n : n \in \mathbb{N}\}$ and $\{\psi_n : n \in \mathbb{N}\}$ be increasing sequences of positive simple functions such that $\phi_n \to f$ and $\psi_n \to g$ pointwise as $n \to \infty$. Then $\phi_n + \psi_n$ is an increasing sequence of positive simple functions such that $\phi_n + \psi_n \to f + g$. It follows from the monotone convergence theorem (Theorem 4.6) and the linearity of the integral on simple functions that

$$\begin{aligned}\int (f+g)\,d\mu &= \lim_{n\to\infty} \int (\phi_n + \psi_n)\,d\mu \\ &= \lim_{n\to\infty} \left(\int \phi_n\,d\mu + \int \psi_n\,d\mu\right) \\ &= \lim_{n\to\infty} \int \phi_n\,d\mu + \lim_{n\to\infty} \int \psi_n\,d\mu \\ &= \int f\,d\mu + \int g\,d\mu,\end{aligned}$$

which proves the result. □

4.3. Measurable functions

If $f : X \to \overline{\mathbb{R}}$ is an extended real-valued function, we define the positive and negative parts $f^+, f^- : X \to [0,\infty]$ of f by

$$f = f^+ - f^-, \qquad f^+ = \max\{f, 0\}, \quad f^- = \max\{-f, 0\}. \tag{4.3}$$

For this decomposition,

$$|f| = f^+ + f^-.$$

Note that f is measurable if and only if f^+ and f^- are measurable.

Definition 4.8. If $f : X \to \overline{\mathbb{R}}$ is a measurable function, then

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu,$$

provided that at least one of the integrals $\int f^+\,d\mu$, $\int f^-\,d\mu$ is finite. The function f is integrable if both $\int f^+\,d\mu$, $\int f^-\,d\mu$ are finite, which is the case if and only if

$$\int |f|\,d\mu < \infty.$$

Note that, according to Definition 4.8, the integral may take the values $-\infty$ or ∞, but it is not defined if both $\int f^+\,d\mu$, $\int f^-\,d\mu$ are infinite. Thus, although the integral of a positive measurable function always exists as an extended real number, the integral of a general, non-integrable real-valued measurable function may not exist.

This Lebesgue integral has all the usual properties of an integral. We restrict attention to integrable functions to avoid undefined expressions involving extended real numbers such as $\infty - \infty$.

Proposition 4.9. *If $f, g : X \to \mathbb{R}$ are integrable functions, then:*

$$\begin{aligned} &\int kf\, d\mu = k \int f\, d\mu \qquad \text{if } k \in \mathbb{R}; \\ &\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu; \\ &\int f\, d\mu \le \int g\, d\mu \qquad \text{if } f \le g; \\ &\left| \int f\, d\mu \right| \le \int |f|\, d\mu. \end{aligned}$$

Proof. These results follow by writing functions into their positive and negative parts, as in (4.3), and using the results for positive functions.

If $f = f^+ - f^-$ and $k \ge 0$, then $(kf)^+ = kf^+$ and $(kf)^- = kf^-$, so

$$\int kf\, d\mu = \int kf^+\, d\mu - \int kf^-\, d\mu = k \int f^+\, d\mu - k \int f^-\, d\mu = k \int f\, d\mu.$$

Similarly, $(-f)^+ = f^-$ and $(-f)^- = f^+$, so

$$\int (-f)\, d\mu = \int f^-\, d\mu - \int f^+\, d\mu = -\int f\, d\mu.$$

If $h = f + g$ and

$$f = f^+ - f^-, \qquad g = g^+ - g^-, \qquad h = h^+ - h^-$$

are the decompositions of f, g, h into their positive and negative parts, then

$$h^+ - h^- = f^+ - f^- + g^+ - g^-.$$

It need not be true that $h^+ = f^+ + g^+$, but we have

$$f^- + g^- + h^+ = f^+ + g^+ + h^-.$$

The linearity of the integral on positive functions gives

$$\int f^-\, d\mu + \int g^-\, d\mu + \int h^+\, d\mu = \int f^+\, d\mu + \int g^+\, d\mu + \int h^-\, d\mu,$$

which implies that

$$\int h^+\, d\mu - \int h^-\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu + \int g^+\, d\mu - \int g^-\, d\mu,$$

or $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$.

It follows that if $f \le g$, then

$$0 \le \int (g - f)\, d\mu = \int g\, d\mu - \int f\, d\mu,$$

so $\int f\, d\mu \le \int g\, d\mu$. The last result is then a consequence of the previous results and $-|f| \le f \le |f|$. □

Let us give two basic examples of the Lebesgue integral.

Example 4.10. Suppose that $X = \mathbb{N}$ and $\nu : \mathcal{P}(\mathbb{N}) \to [0, \infty]$ is counting measure on $\mathbb{N}$. If $f : \mathbb{N} \to \mathbb{R}$ and $f(n) = x_n$, then

$$\int_{\mathbb{N}} f \, d\nu = \sum_{n=1}^{\infty} x_n,$$

where the integral is finite if and only if the series is absolutely convergent. Thus, the theory of absolutely convergent series is a special case of the Lebesgue integral. Note that a conditionally convergent series, such as the alternating harmonic series, does not correspond to a Lebesgue integral, since both its positive and negative parts diverge.

Example 4.11. Suppose that $X = [a, b]$ is a compact interval and $\mu : \mathcal{L}([a, b]) \to \mathbb{R}$ is Lesbegue measure on $[a, b]$. We note in Section 4.8 that any Riemann integrable function $f : [a, b] \to \mathbb{R}$ is integrable with respect to Lebesgue measure μ, and its Riemann integral is equal to the Lebesgue integral,

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

Thus, all of the usual integrals from elementary calculus remain valid for the Lebesgue integral on $\mathbb{R}$. We will write an integral with respect to Lebesgue measure on $\mathbb{R}$, or $\mathbb{R}^n$, as

$$\int f \, dx.$$

Even though the class of Lebesgue integrable functions on an interval is wider than the class of Riemann integrable functions, some improper Riemann integrals may exist even though the Lebesegue integral does not.

Example 4.12. The integral

$$\int_0^1 \left(\frac{1}{x} \sin \frac{1}{x} + \cos \frac{1}{x} \right) \, dx$$

is not defined as a Lebesgue integral, although the improper Riemann integral

$$\lim_{\epsilon \to 0^+} \int_\epsilon^1 \left(\frac{1}{x} \sin \frac{1}{x} + \cos \frac{1}{x} \right) \, dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{d}{dx} \left[x \cos \frac{1}{x} \right] \, dx = \cos 1$$

exists.

Example 4.13. The integral

$$\int_{-1}^1 \frac{1}{x} \, dx$$

is not defined as a Lebesgue integral, although the principal value integral

$$\text{p.v.} \int_{-1}^1 \frac{1}{x} \, dx = \lim_{\epsilon \to 0^+} \left\{ \int_{-1}^{-\epsilon} \frac{1}{x} \, dx + \int_\epsilon^1 \frac{1}{x} \, dx \right\} = 0$$

exists. Note, however, that the Lebesgue integrals

$$\int_0^1 \frac{1}{x} \, dx = \infty, \qquad \int_{-1}^0 \frac{1}{x} \, dx = -\infty$$

are well-defined as extended real numbers.

The inability of the Lebesgue integral to deal directly with the cancelation between large positive and negative parts in oscillatory or singular integrals, such as the ones in the previous examples, is sometimes viewed as a defect (although the integrals above can still be defined as an appropriate limit of Lebesgue integrals). Other definitions of the integral such as the Henstock-Kurzweil integral, which is a generalization of the Riemann integral, avoid this defect but they have not proved to be as useful as the Lebesgue integral. Similar issues arise in connection with Feynman path integrals in quantum theory, where one would like to define the integral of highly oscillatory functionals on an infinite-dimensional function-space.

4.4. Absolute continuity

The following results show that a function with finite integral is finite a.e. and that the integral depends only on the pointwise a.e. values of a function.

Proposition 4.14. *If $f : X \to \overline{\mathbb{R}}$ is an integrable function, meaning that $\int |f|\, d\mu < \infty$, then f is finite μ-a.e. on X.*

Proof. We may assume that f is positive without loss of generality. Suppose that

$$E = \{x \in X : f = \infty\}$$

has nonzero measure. Then for any $t > 0$, we have $f > t\chi_E$, so

$$\int f\, d\mu \geq \int t\chi_E\, d\mu = t\mu(E),$$

which implies that $\int f\, d\mu = \infty$. □

Proposition 4.15. *Suppose that $f : X \to \overline{\mathbb{R}}$ is an extended real-valued measurable function. Then $\int |f|\, d\mu = 0$ if and only if $f = 0$ μ-a.e.*

Proof. By replacing f with $|f|$, we can assume that f is positive without loss of generality. Suppose that $f = 0$ a.e. If $0 \leq \phi \leq f$ is a simple function,

$$\phi = \sum_{i=1}^{N} c_i \chi_{E_i},$$

then $\phi = 0$ a.e., so $c_i = 0$ unless $\mu(E_i) = 0$. It follows that

$$\int \phi\, d\mu = \sum_{i=1}^{N} c_i \mu(E_i) = 0,$$

and Definition 4.4 implies that $\int f\, d\mu = 0$.

Conversely, suppose that $\int f\, d\mu = 0$. For $n \in \mathbb{N}$, let

$$E_n = \{x \in X : f(x) \geq 1/n\}.$$

Then $0 \leq (1/n)\chi_{E_n} \leq f$, so that

$$0 \leq \frac{1}{n}\mu(E_n) = \int \frac{1}{n}\chi_{E_n}\, d\mu \leq \int f\, d\mu = 0,$$

and hence $\mu(E_n) = 0$. Now observe that

$$\{x \in X : f(x) > 0\} = \bigcup_{n=1}^{\infty} E_n,$$

so it follows from the countable additivity of μ that $f = 0$ a.e. □

In particular, it follows that if $f : X \to \overline{\mathbb{R}}$ is any measurable function, then

$$\int_A f \, d\mu = 0 \qquad \text{if } \mu(A) = 0. \tag{4.4}$$

For integrable functions we can strengthen the previous result to get the following property, which is called the absolute continuity of the integral.

Proposition 4.16. *Suppose that $f : X \to \overline{\mathbb{R}}$ is an integrable function, meaning that $\int |f| \, d\mu < \infty$. Then, given any $\epsilon > 0$, there exists $\delta > 0$ such that*

$$0 \le \int_A |f| \, d\mu < \epsilon \tag{4.5}$$

whenever A is a measurable set with $\mu(A) < \delta$.

Proof. Again, we can assume that f is positive. For $n \in \mathbb{N}$, define $f_n : X \to [0, \infty]$ by

$$f_n(x) = \begin{cases} n & \text{if } f(x) \ge n, \\ f(x) & \text{if } 0 \le f(x) < n. \end{cases}$$

Then $\{f_n\}$ is an increasing sequence of positive measurable functions that converges pointwise to f. We estimate the integral of f over A as follows:

$$\begin{aligned} \int_A f \, d\mu &= \int_A (f - f_n) \, d\mu + \int_A f_n \, d\mu \\ &\le \int_X (f - f_n) \, d\mu + n\mu(A). \end{aligned}$$

By the monotone convergence theorem,

$$\int_X f_n \, d\mu \to \int_X f \, d\mu < \infty$$

as $n \to \infty$. Therefore, given $\epsilon > 0$, we can choose n such that

$$0 \le \int_X (f - f_n) \, d\mu < \frac{\epsilon}{2},$$

and then choose

$$\delta = \frac{\epsilon}{2n}.$$

If $\mu(A) < \delta$, we get (4.5), which proves the result. □

Proposition 4.16 may fail if f is not integrable.

Example 4.17. Define $\nu : \mathcal{B}((0,1)) \to [0, \infty]$ by

$$\nu(A) = \int_A \frac{1}{x} \, dx,$$

where the integral is taken with respect to Lebesgue measure μ. Then $\nu(A) = 0$ if $\mu(A) = 0$, but (4.5) does not hold.

There is a converse to this theorem concerning the representation of absolutely continuous measures as integrals (the Radon-Nikodym theorem, stated in Theorem 6.27).

4.5. Convergence theorems

One of the most basic questions in integration theory is the following: If $f_n \to f$ pointwise, when can one say that

$$\int f_n \, d\mu \to \int f \, d\mu? \tag{4.6}$$

The Riemann integral is not sufficiently general to permit a satisfactory answer to this question.

Perhaps the simplest condition that guarantees the convergence of the integrals is that the functions $f_n : X \to \mathbb{R}$ converge uniformly to $f : X \to \mathbb{R}$ and X has finite measure. In that case

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| \leq \int |f_n - f| \, d\mu \leq \mu(X) \sup_X |f_n - f| \to 0$$

as $n \to \infty$. The assumption of uniform convergence is too strong for many purposes, and the Lebesgue integral allows the formulation of simple and widely applicable theorems for the convergence of integrals. The most important of these are the monotone convergence theorem (Theorem 4.6) and the Lebesgue dominated convergence theorem (Theorem 4.24). The utility of these results accounts, in large part, for the success of the Lebesgue integral.

Some conditions on the functions f_n in (4.6) are, however, necessary to ensure the convergence of the integrals, as can be seen from very simple examples. Roughly speaking, the convergence may fail because 'mass' can leak out to infinity in the limit.

Example 4.18. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ as $n \to \infty$ pointwise on $\mathbb{R}$, but

$$\int f_n \, dx = 1 \qquad \text{for every } n \in \mathbb{N}.$$

By modifying this example, and the following ones, we can obtain a sequence f_n that converges pointwise to zero but whose integrals converge to infinity; for example

$$f_n(x) = \begin{cases} n^2 & \text{if } 0 < x < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Example 4.19. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1/n & \text{if } 0 < x < n, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ as $n \to \infty$ pointwise on $\mathbb{R}$, and even uniformly, but

$$\int f_n \, dx = 1 \qquad \text{for every } n \in \mathbb{N}.$$

Example 4.20. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 & \text{if } n < x < n+1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ as $n \to \infty$ pointwise on $\mathbb{R}$, but

$$\int f_n \, dx = 1 \qquad \text{for every } n \in \mathbb{N}.$$

The monotone convergence theorem implies that a similar failure of convergence of the integrals cannot occur in an increasing sequence of functions, even if the convergence is not uniform or the domain space does not have finite measure. Note that the monotone convergence theorem does not hold for the Riemann integral; indeed, the pointwise limit of a monotone increasing, bounded sequence of Riemann integrable functions need not even be Riemann integrable.

Example 4.21. Let $\{q_i : i \in \mathbb{N}\}$ be an enumeration of the rational numbers in the interval $[0,1]$, and define $f_n : [0,1] \to [0,\infty)$ by

$$f_n(x) = \begin{cases} 1 & \text{if } x = q_i \text{ for some } 1 \le i \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\{f_n\}$ is a monotone increasing sequence of bounded, positive, Riemann integrable functions, each of which has zero integral. Nevertheless, as $n \to \infty$ the sequence converges pointwise to the characteristic function of the rationals in $[0,1]$, which is not Riemann integrable.

A useful generalization of the monotone convergence theorem is the following result, called Fatou's lemma.

Theorem 4.22. *Suppose that $\{f_n : n \in \mathbb{N}\}$ is sequence of positive measurable functions $f_n : X \to [0,\infty]$. Then*

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu. \tag{4.7}$$

Proof. For each $n \in \mathbb{N}$, let

$$g_n = \inf_{k \ge n} f_k.$$

Then $\{g_n\}$ is a monotone increasing sequence which converges pointwise to $\liminf f_n$ as $n \to \infty$, so by the monotone convergence theorem

$$\lim_{n\to\infty} \int g_n \, d\mu = \int \liminf_{n\to\infty} f_n \, d\mu. \tag{4.8}$$

Moreover, since $g_n \le f_k$ for every $k \ge n$, we have

$$\int g_n \, d\mu \le \inf_{k \ge n} \int f_k \, d\mu,$$

so that

$$\lim_{n\to\infty} \int g_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu.$$

Using (4.8) in this inequality, we get the result. □

We may have strict inequality in (4.7), as in the previous examples. The monotone convergence theorem and Fatou's Lemma enable us to determine the integrability of functions.

Example 4.23. For $\alpha \in \mathbb{R}$, consider the function $f : [0,1] \to [0,\infty]$ defined by

$$f(x) = \begin{cases} x^{-\alpha} & \text{if } 0 < x \leq 1, \\ \infty & \text{if } x = 0. \end{cases}$$

For $n \in \mathbb{N}$, let

$$f_n(x) = \begin{cases} x^{-\alpha} & \text{if } 1/n \leq x \leq 1, \\ n^{\alpha} & \text{if } 0 \leq x < 1/n. \end{cases}$$

Then $\{f_n\}$ is an increasing sequence of Lebesgue measurable functions (*e.g* since f_n is continuous) that converges pointwise to f. We denote the integral of f with respect to Lebesgue measure on $[0,1]$ by $\int_0^1 f(x)\,dx$. Then, by the monotone convergence theorem,

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \int_0^1 f_n(x)\,dx.$$

From elementary calculus,

$$\int_0^1 f_n(x)\,dx \to \frac{1}{1-\alpha}$$

as $n \to \infty$ if $\alpha < 1$, and to ∞ if $\alpha \geq 1$. Thus, f is integrable on $[0,1]$ if and only if $\alpha < 1$.

Perhaps the most frequently used convergence result is the following dominated convergence theorem, in which all the integrals are necessarily finite.

Theorem 4.24 (Lebesgue Dominated Convergence Theorem)**.** *If $\{f_n : n \in \mathbb{N}\}$ is a sequence of measurable functions $f_n : X \to \mathbb{R}$ such that $f_n \to f$ pointwise, and $|f_n| \leq g$ where $g : X \to [0,\infty]$ is an integrable function, meaning that $\int g\,d\mu < \infty$, then*

$$\int f_n\,d\mu \to \int f\,d\mu \qquad \text{as } n \to \infty.$$

Proof. Since $g + f_n \geq 0$ for every $n \in \mathbb{N}$, Fatou's lemma implies that

$$\int (g+f)\,d\mu \leq \liminf_{n\to\infty} \int (g+f_n)\,d\mu \leq \int g\,d\mu + \liminf_{n\to\infty} \int f_n\,d\mu,$$

which gives

$$\int f\,d\mu \leq \liminf_{n\to\infty} \int f_n\,d\mu.$$

Similarly, $g - f_n \geq 0$, so

$$\int (g-f)\,d\mu \leq \liminf_{n\to\infty} \int (g-f_n)\,d\mu \leq \int g\,d\mu - \limsup_{n\to\infty} \int f_n\,d\mu,$$

which gives

$$\int f\,d\mu \geq \limsup_{n\to\infty} \int f_n\,d\mu,$$

and the result follows. □

An alternative, and perhaps more illuminating, proof of the dominated convergence theorem may be obtained from Egoroff's theorem and the absolute continuity of the integral. Egoroff's theorem states that if a sequence $\{f_n\}$ of measurable functions, defined on a finite measure space $(X, \mathcal{A}, \mu)$, converges pointwise to a function f, then for every $\epsilon > 0$ there exists a measurable set $A \subset X$ such that $\{f_n\}$ converges uniformly to f on A and $\mu(X \setminus A) < \epsilon$. The uniform integrability of the functions and the absolute continuity of the integral imply that what happens off the set A may be made to have arbitrarily small effect on the integrals. Thus, the convergence theorems hold because of this 'almost' uniform convergence of pointwise-convergent sequences of measurable functions.

4.6. Complex-valued functions and a.e. convergence

In this section, we briefly indicate the generalization of the above results to complex-valued functions and sequences that converge pointwise almost everywhere. The required modifications are straightforward.

If $f : X \to \mathbb{C}$ is a complex valued function $f = g + ih$, then we say that f is measurable if and only if its real and imaginary parts $g, h : X \to \mathbb{R}$ are measurable, and integrable if and only if g, h are integrable. In that case, we define

$$\int f \, d\mu = \int g \, d\mu + i \int h \, d\mu.$$

Note that we do not allow extended real-valued functions or infinite integrals here. It follows from the discussion of product measures that $f : X \to \mathbb{C}$, where $\mathbb{C}$ is equipped with its Borel σ-algebra $\mathcal{B}(\mathbb{C})$, is measurable if and only if its real and imaginary parts are measurable, so this definition is consistent with our previous one.

The integral of complex-valued functions satisfies the properties given in Proposition 4.9, where we allow $k \in \mathbb{C}$ and the condition $f \leq g$ is only relevant for real-valued functions. For example, to show that $|\int f \, d\mu| \leq \int |f| \, d\mu$, we let

$$\int f \, d\mu = \left| \int f \, d\mu \right| e^{i\theta}$$

for a suitable argument θ, and then

$$\left| \int f \, d\mu \right| = e^{-i\theta} \int f \, d\mu = \int \Re[e^{-i\theta} f] \, d\mu \leq \int |\Re[e^{-i\theta} f]| \, d\mu \leq \int |f| \, d\mu.$$

Complex-valued functions also satisfy the properties given in Section 4.4.

The monotone convergence theorem holds for extended real-valued functions if $f_n \uparrow f$ pointwise a.e., and the Lebesgue dominated convergence theorem holds for complex-valued functions if $f_n \to f$ pointwise a.e. and $|f_n| \leq g$ pointwise a.e. where g is an integrable extended real-valued function. If the measure space is not complete, then we also need to assume that f is measurable. To prove these results, we replace the functions f_n, for example, by $f_n \chi_{N^c}$ where N is a null set off which pointwise convergence holds, and apply the previous theorems; the values of any integrals are unaffected.

4.7. L^1 spaces

We introduce here the space $L^1(X)$ of integrable functions on a measure space X; we will study its properties, and the properties of the closely related L^p spaces, in more detail later on.

Definition 4.25. If $(X, \mathcal{A}, \mu)$ is a measure space, then the space $L^1(X)$ consists of the integrable functions $f : X \to \mathbb{R}$ with norm

$$\|f\|_{L^1} = \int |f| \, d\mu < \infty,$$

where we identify functions that are equal a.e. A sequence of functions

$$\{f_n \in L^1(X)\}$$

converges in L^1, or in mean, to $f \in L^1(X)$ if

$$\|f - f_n\|_{L^1} = \int |f - f_n| \, d\mu \to 0 \qquad \text{as } n \to \infty.$$

We also denote the space of integrable complex-valued functions $f : X \to \mathbb{C}$ by $L^1(X)$. For definiteness, we consider real-valued functions unless stated otherwise; in most cases, the results generalize in an obvious way to complex-valued functions

Convergence in mean is not equivalent to pointwise a.e.-convergence. The sequences in Examples 4.18–4.20 converges to zero pointwise, but they do not converge in mean. The following is an example of a sequence that converges in mean but not pointwise a.e.

Example 4.26. Define $f_n : [0,1] \to \mathbb{R}$ by

$$f_1(x) = 1, \quad f_2(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1/2, \\ 0 & \text{if } 1/2 < x \le 1, \end{cases} \quad f_3(x) = \begin{cases} 0 & \text{if } 0 \le x < 1/2, \\ 1 & \text{if } 1/2 \le x \le 1, \end{cases}$$

$$f_4(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1/4, \\ 1 & \text{if } 1/4 < x \le 1, \end{cases} \qquad f_5(x) = \begin{cases} 0 & \text{if } 0 \le x < 1/4, \\ 1 & \text{if } 1/4 \le x \le 1/2, \\ 0 & \text{if } 1/2x < x \le 1, \end{cases}$$

and so on. That is, for $2^m \le n \le 2^m - 1$, the function f_n consists of a spike of height one and width 2^{-m} that sweeps across the interval $[0,1]$ as n increases from 2^m to $2^m - 1$. This sequence converges to zero in mean, but it does not converge pointwise as any point $x \in [0,1]$.

We will show, however, that a sequence which converges sufficiently rapidly in mean does converge pointwise a.e.; as a result, every sequence that converges in mean has a subsequence that converges pointwise a.e. (see Lemma 7.9 and Corollary 7.11).

Let us consider the particular case of $L^1(\mathbb{R}^n)$. As an application of the Borel regularity of Lebesgue measure, we prove that integrable functions on $\mathbb{R}^n$ may be approximated by continuous functions with compact support. This result means that $L^1(\mathbb{R}^n)$ is a concrete realization of the completion of $C_c(\mathbb{R}^n)$ with respect to the $L^1(\mathbb{R}^n)$-norm, where $C_c(\mathbb{R}^n)$ denotes the space of continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ with compact support. The support of f is defined by

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}.$$

Thus, f has compact support if and only if it vanishes outside a bounded set.

Theorem 4.27. *The space $C_c(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$. Explicitly, if $f \in L^1(\mathbb{R}^n)$, then for any $\epsilon > 0$ there exists a function $g \in C_c(\mathbb{R}^n)$ such that*

$$\|f - g\|_{L^1} < \epsilon.$$

Proof. Note first that by the dominated convergence theorem

$$\left\|f - f\chi_{B_R(0)}\right\|_{L^1} \to 0 \qquad \text{as } R \to \infty,$$

so we can assume that $f \in L^1(\mathbb{R}^n)$ has compact support. Decomposing $f = f^+ - f^-$ into positive and negative parts, we can also assume that f is positive. Then there is an increasing sequence of compactly supported simple functions that converges to f pointwise and hence, by the monotone (or dominated) convergence theorem, in mean. Since every simple function is a finite linear combination of characteristic functions, it is sufficient to prove the result for the characteristic function χ_A of a bounded, measurable set $A \subset \mathbb{R}^n$.

Given $\epsilon > 0$, by the Borel regularity of Lebesgue measure, there exists a bounded open set G and a compact set K such that $K \subset A \subset G$ and $\mu(G \setminus K) < \epsilon$. Let $g \in C_c(\mathbb{R}^n)$ be a Urysohn function such that $g = 1$ on K, $g = 0$ on G^c, and $0 \leq g \leq 1$. For example, we can define g explicitly by

$$g(x) = \frac{d(x, G^c)}{d(x, K) + d(x, G^c)}$$

where the distance function $d(\cdot, F) : \mathbb{R}^n \to \mathbb{R}$ from a subset $F \subset \mathbb{R}^n$ is defined by

$$d(x, F) = \inf \{|x - y| : y \in F\}.$$

If F is closed, then $d(\cdot, F)$ is continuous, so g is continuous.

We then have that

$$\|\chi_A - g\|_{L^1} = \int_{G \setminus K} |\chi_A - g| \, dx \leq \mu(G \setminus K) < \epsilon,$$

which proves the result. □

4.8. Riemann integral

Any Riemann integrable function $f : [a, b] \to \mathbb{R}$ is Lebesgue measurable, and in fact integrable since it must be bounded, but a Lebesgue integrable function need not be Riemann integrable. Using Lebesgue measure, we can give a necessary and sufficient condition for Riemann integrability.

Theorem 4.28. *If $f : [a, b] \to \mathbb{R}$ is Riemann integrable, then f is Lebesgue integrable on $[a, b]$ and its Riemann integral is equal to its Lebesgue integral. A Lebesgue measurable function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and the set of discontinuities $\{x \in [a, b] : f \text{ is discontinuous at } x\}$ has Lebesgue measure zero.*

For the proof, see *e.g.* Folland [**4**].

Example 4.29. The characteristic function of the rationals $\chi_{\mathbb{Q} \cap [0,1]}$ is discontinuous at every point and it is not Riemann integrable on $[0, 1]$. This function is, however, equal a.e. to the zero function which is continuous at every point and is Riemann integrable. (Note that being continuous a.e. is not the same thing as being equal a.e. to a continuous function.) Any function that is bounded and continuous except at countably many points is Riemann integrable, but these are not the only Riemann integrable functions. For example, the characteristic function of a Cantor set with zero measure is a Riemann integrable function that is discontinuous at uncountable many points.

4.9. Integrals of vector-valued functions

In Definition 4.4, we use the ordering properties of $\overline{\mathbb{R}}$ to define real-valued integrals as a supremum of integrals of simple functions. Finite integrals of complex-valued functions or vector-valued functions that take values in a finite-dimensional vector space are then defined componentwise.

An alternative method is to define the integral of a vector-valued function $f : X \to Y$ from a measure space X to a Banach space Y as a limit in norm of integrals of vector-valued simple functions. The integral of vector-valued simple functions is defined as in (4.1), assuming that $\mu(E_n) < \infty$; linear combinations of the values $c_n \in Y$ make sense since Y is a vector space. A function $f : X \to Y$ is integrable if there is a sequence of integrable simple functions $\{\phi_n : X \to Y\}$ such that $\phi_n \to f$ pointwise, where the convergence is with respect to the norm $\|\cdot\|$ on Y, and

$$\int \|f - \phi_n\| \, d\mu \to 0 \qquad \text{as } n \to \infty.$$

Then we define

$$\int f \, d\mu = \lim_{n\to\infty} \int \phi_n \, d\mu,$$

where the limit is the norm-limit in Y.

This definition of the integral agrees with the one used above for real-valued, integrable functions, and amounts to defining the integral of an integrable function by completion in the L^1-norm. We will not develop this definition here (see [**6**], for example, for a detailed account), but it is useful in defining the integral of functions that take values in an infinite-dimensional Banach space, when it leads to the Bochner integral. An alternative approach is to reduce vector-valued integrals to scalar-valued integrals by the use of continuous linear functionals belonging to the dual space of the Banach space.

CHAPTER 5

Product Measures

Given two measure spaces, we may construct a natural measure on their Cartesian product; the prototype is the construction of Lebesgue measure on $\mathbb{R}^2$ as the product of Lebesgue measures on $\mathbb{R}$. The integral of a measurable function on the product space may be evaluated as iterated integrals on the individual spaces provided that the function is positive or integrable (and the measure spaces are σ-finite). This result, called Fubini's theorem, is another one of the basic and most useful properties of the Lebesgue integral. We will not give complete proofs of all the results in this Chapter.

5.1. Product σ-algebras

We begin by describing product σ-algebras. If $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces, then a measurable rectangle is a subset $A \times B$ of $X \times Y$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$ are measurable subsets of X and Y, respectively. For example, if $\mathbb{R}$ is equipped with its Borel σ-algebra, then $\mathbb{Q} \times \mathbb{Q}$ is a measurable rectangle in $\mathbb{R} \times \mathbb{R}$. (Note that the 'sides' A, B of a measurable rectangle $A \times B \subset \mathbb{R} \times \mathbb{R}$ can be arbitrary measurable sets; they are not required to be intervals.)

Definition 5.1. Suppose that $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces. The product σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is the σ-algebra on $X \times Y$ generated by the collection of all measurable rectangles,

$$\mathcal{A} \otimes \mathcal{B} = \sigma\left(\{A \times B : A \in \mathcal{A},\ B \in \mathcal{B}\}\right).$$

The product of $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ is the measurable space $(X \times Y, \mathcal{A} \otimes \mathcal{B})$.

Suppose that $E \subset X \times Y$. For any $x \in X$ and $y \in Y$, we define the x-section $E_x \subset Y$ and the y-section $E^y \subset X$ of E by

$$E_x = \{y \in Y : (x, y) \in E\}, \qquad E^y = \{x \in X : (x, y) \in E\}.$$

As stated in the next proposition, all sections of a measurable set are measurable.

Proposition 5.2. *If $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces and $E \in \mathcal{A} \otimes \mathcal{B}$, then $E_x \in \mathcal{B}$ for every $x \in X$ and $E^y \in \mathcal{A}$ for every $y \in Y$.*

Proof. Let

$$\mathcal{M} = \{E \subset X \times Y : E_x \in \mathcal{B} \text{ for every } x \in X \text{ and } E^y \in \mathcal{A} \text{ for every } y \in Y\}.$$

Then $\mathcal{M}$ contains all measurable rectangles, since the x-sections of $A \times B$ are either $\varnothing$ or B and the y-sections are either $\varnothing$ or A. Moreover, $\mathcal{M}$ is a σ-algebra since, for example, if $E, E_i \subset X \times Y$ and $x \in X$, then

$$(E^c)_x = (E_x)^c, \qquad \left(\bigcup_{i=1}^{\infty} E_i\right)_x = \bigcup_{i=1}^{\infty} (E_i)_x.$$

It follows that $\mathcal{M} \supset \mathcal{A} \otimes \mathcal{B}$, which proves the proposition. □

As an example, we consider the product of Borel σ-algebras on $\mathbb{R}^n$.

Proposition 5.3. *Suppose that $\mathbb{R}^m$, $\mathbb{R}^n$ are equipped with their Borel σ-algebras $\mathcal{B}(\mathbb{R}^m)$, $\mathcal{B}(\mathbb{R}^n)$ and let $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$. Then*

$$\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

PROOF. Every $(m+n)$-dimensional rectangle, in the sense of Definition 2.1, is a product of an m-dimensional and an n-dimensional rectangle. Therefore

$$\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n) \supset \mathcal{R}(\mathbb{R}^{m+n})$$

where $\mathcal{R}(\mathbb{R}^{m+n})$ denotes the collection of rectangles in $\mathbb{R}^{m+n}$. From Proposition 2.21, the rectangles generate the Borel σ-algebra, and therefore

$$\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n) \supset \mathcal{B}(\mathbb{R}^{m+n}).$$

To prove the the reverse inclusion, let

$$\mathcal{M} = \left\{A \subset \mathbb{R}^m : A \times \mathbb{R}^n \in \mathcal{B}(\mathbb{R}^{m+n})\right\}.$$

Then $\mathcal{M}$ is a σ-algebra, since $\mathcal{B}(\mathbb{R}^{m+n})$ is a σ-algebra and

$$A^c \times \mathbb{R}^n = \left(A \times \mathbb{R}^n\right)^c, \qquad \left(\bigcup_{i=1}^{\infty} A_i\right) \times \mathbb{R}^n = \bigcup_{i=1}^{\infty} \left(A_i \times \mathbb{R}^n\right).$$

Moreover, $\mathcal{M}$ contains all open sets, since $G \times \mathbb{R}^n$ is open in $\mathbb{R}^{m+n}$ if G is open in $\mathbb{R}^m$. It follows that $\mathcal{M} \supset \mathcal{B}(\mathbb{R}^m)$, so $A \times \mathbb{R}^n \in \mathcal{B}(\mathbb{R}^{m+n})$ for every $A \in \mathcal{B}(\mathbb{R}^m)$, meaning that

$$\mathcal{B}(\mathbb{R}^{m+n}) \supset \left\{A \times \mathbb{R}^n : A \in \mathcal{B}(\mathbb{R}^m)\right\}.$$

Similarly, we have

$$\mathcal{B}(\mathbb{R}^{m+n}) \supset \left\{\mathbb{R}^n \times B : B \in \mathcal{B}(\mathbb{R}^n)\right\}.$$

Therefore, since $\mathcal{B}(\mathbb{R}^{m+n})$ is closed under intersections,

$$\mathcal{B}(\mathbb{R}^{m+n}) \supset \left\{A \times B : A \in \mathcal{B}(\mathbb{R}^m),\, B \in \mathcal{B}(\mathbb{R}^n)\right\},$$

which implies that

$$\mathcal{B}(\mathbb{R}^{m+n}) \supset \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

□

By the repeated application of this result, we see that the Borel σ-algebra on $\mathbb{R}^n$ is the n-fold product of the Borel σ-algebra on $\mathbb{R}$. This leads to an alternative method of constructing Lebesgue measure on $\mathbb{R}^n$ as a product of Lebesgue measures on $\mathbb{R}$, instead of the direct construction we gave earlier.

5.2. Premeasures

Premeasures provide a useful way to generate outer measures and measures, and we will use them to construct product measures. In this section, we derive some general results about premeasures and their associated measures that we use below. Premeasures are defined on algebras, rather than σ-algebras, but they are consistent with countable additivity.

Definition 5.4. An algebra on a set X is a collection of subsets of X that contains $\varnothing$ and X and is closed under complements, finite unions, and finite intersections.

If $\mathcal{F} \subset \mathcal{P}(X)$ is a family of subsets of a set X, then the algebra generated by $\mathcal{F}$ is the smallest algebra that contains $\mathcal{F}$. It is much easier to give an explicit description of the algebra generated by a family of sets $\mathcal{F}$ than the σ-algebra generated by $\mathcal{F}$. For example, if $\mathcal{F}$ has the property that for $A, B \in \mathcal{F}$, the intersection $A \cap B \in \mathcal{F}$ and the complement A^c is a finite union of sets belonging to $\mathcal{F}$, then the algebra generated by $\mathcal{F}$ is the collection of all finite unions of sets in $\mathcal{F}$.

Definition 5.5. Suppose that $\mathcal{E}$ is an algebra of subsets of a set X. A premeasure λ on $\mathcal{E}$, or on X if the algebra is understood, is a function $\lambda : \mathcal{E} \to [0, \infty]$ such that:

(a) $\lambda(\varnothing) = 0$;
(b) if $\{A_i \in \mathcal{E} : i \in \mathbb{N}\}$ is a countable collection of disjoint sets in $\mathcal{E}$ such that
$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{E},$$
then
$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda\left(A_i\right).$$

Note that a premeasure is finitely additive, since we may take $A_i = \varnothing$ for $i \geq N$, and monotone, since if $A \supset B$, then $\lambda(A) = \lambda(A \setminus B) + \lambda(B) \geq \lambda(B)$.

To define the outer measure associated with a premeasure, we use countable coverings by sets in the algebra.

Definition 5.6. Suppose that $\mathcal{E}$ is an algebra on a set X and $\lambda : \mathcal{E} \to [0, \infty]$ is a premeasure. The outer measure $\lambda^* : \mathcal{P}(X) \to [0, \infty]$ associated with λ is defined for $E \subset X$ by
$$\lambda^*(E) = \inf\left\{\sum_{i=1}^{\infty} \lambda(A_i) : E \subset \textstyle\bigcup_{i=1}^{\infty} A_i \text{ where } A_i \in \mathcal{E}\right\}.$$

As we observe next, the set-function λ^* is an outer measure. Moreover, every set belonging to $\mathcal{E}$ is λ^*-measurable and its outer measure is equal to its premeasure.

Proposition 5.7. *The set function $\lambda^* : \mathcal{P}(X) \to [0, \infty]$ given by Definition 5.6. is an outer measure on X. Every set $A \in \mathcal{E}$ is Carathéodory measurable and $\lambda^*(A) = \lambda(A)$.*

Proof. The proof that λ^* is an outer measure is identical to the proof of Theorem 2.4 for outer Lebesgue measure.

If $A \in \mathcal{E}$, then $\lambda^*(A) \leq \lambda(A)$ since A covers itself. To prove the reverse inequality, suppose that $\{A_i : i \in \mathbb{N}\}$ is a countable cover of A by sets $A_i \in \mathcal{E}$. Let $B_1 = A \cap A_1$ and
$$B_j = A \cap \left(A_j \setminus \bigcup_{i=1}^{j-1} A_i\right) \qquad \text{for } j \geq 2.$$
Then $B_j \in \mathcal{A}$ and A is the disjoint union of $\{B_j : j \in \mathbb{N}\}$. Since $B_j \subset A_j$, it follows that
$$\lambda(A) = \sum_{j=1}^{\infty} \lambda(B_j) \leq \sum_{j=1}^{\infty} \lambda(A_j),$$
which implies that $\lambda(A) \leq \lambda^*(A)$. Hence, $\lambda^*(A) = \lambda(A)$.

If $E \subset X$, $A \in \mathcal{E}$, and $\epsilon > 0$, then there is a cover $\{B_i \in \mathcal{E} : i \in \mathbb{N}\}$ of E such that

$$\lambda^*(E) + \epsilon \geq \sum_{i=1}^{\infty} \lambda(B_i).$$

Since λ is countably additive on $\mathcal{E}$,

$$\lambda^*(E) + \epsilon \geq \sum_{i=1}^{\infty} \lambda(B_i \cap A) + \sum_{i=1}^{\infty} \lambda(B_i \cap A^c) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c),$$

and since $\epsilon > 0$ is arbitrary, it follows that $\lambda^*(E) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$, which implies that A is measurable. □

Using Theorem 2.9, we see from the preceding results that every premeasure on an algebra $\mathcal{E}$ may be extended to a measure on $\sigma(\mathcal{E})$. A natural question is whether such an extension is unique. In general, the answer is no, but if the measure space is not 'too big,' in the following sense, then we do have uniqueness.

Definition 5.8. Let X be a set and λ a premeasure on an algebra $\mathcal{E} \subset \mathcal{P}(X)$. Then λ is σ-finite if $X = \bigcup_{i=1}^{\infty} A_i$ where $A_i \in \mathcal{E}$ and $\lambda(A_i) < \infty$.

Theorem 5.9. *If $\lambda : \mathcal{E} \to [0, \infty]$ is a σ-finite premeasure on an algebra $\mathcal{E}$ and $\mathcal{A}$ is the σ-algebra generated by $\mathcal{E}$, then there is a unique measure $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(A) = \lambda(A)$ for every $A \in \mathcal{E}$.*

5.3. Product measures

Next, we construct a product measure on the product of measure spaces that satisfies the natural condition that the measure of a measurable rectangle is the product of the measures of its 'sides.' To do this, we will use the Carathéodory method and first define an outer measure on the product of the measure spaces in terms of the natural premeasure defined on measurable rectangles. The procedure is essentially the same as the one we used to construct Lebesgue measure on $\mathbb{R}^n$.

Suppose that $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces. The intersection of measurable rectangles is a measurable rectangle

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D),$$

and the complement of a measurable rectangle is a finite union of measurable rectangles

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c).$$

Thus, the collection of finite unions of measurable rectangles in $X \times Y$ forms an algebra, which we denote by $\mathcal{E}$. This algebra is not, in general, a σ-algebra, but obviously it generates the same product σ-algebra as the measurable rectangles.

Every set $E \in \mathcal{E}$ may be represented as a finite disjoint union of measurable rectangles, though not necessarily in a unique way. To define a premeasure on $\mathcal{E}$, we first define the premeasure of measurable rectangles.

Definition 5.10. If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are measure spaces, then the product premeasure $\lambda(A \times B)$ of a measurable rectangle $A \times B \subset X \times Y$ is given by

$$\lambda(A \times B) = \mu(A)\nu(B)$$

where $0 \cdot \infty = 0$.

The premeasure λ is countably additive on rectangles. The simplest way to show this is to integrate the characteristic functions of the rectangles, which allows us to use the monotone convergence theorem.

Proposition 5.11. *If a measurable rectangle $A \times B$ is a countable disjoint union of measurable rectangles $\{A_i \times B_i : i \in \mathbb{N}\}$, then*

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

Proof. If

$$A \times B = \bigcup_{i=1}^{\infty} (A_i \times B_i)$$

is a disjoint union, then the characteristic function $\chi_{A\times B} : X \times Y \to [0, \infty)$ satisfies

$$\chi_{A\times B}(x, y) = \sum_{i=1}^{\infty} \chi_{A_i\times B_i}(x, y).$$

Therefore, since $\chi_{A\times B}(x, y) = \chi_A(x)\chi_B(y)$,

$$\chi_A(x)\chi_B(y) = \sum_{i=1}^{\infty} \chi_{A_i}(x)\chi_{B_i}(y).$$

Integrating this equation over Y for fixed $x \in X$ and using the monotone convergence theorem, we get

$$\chi_A(x)\nu(B) = \sum_{i=1}^{\infty} \chi_{A_i}(x)\nu(B_i).$$

Integrating again with respect to x, we get

$$\mu(A)\nu(B) = \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i),$$

which proves the result. □

In particular, it follows that λ is finitely additive on rectangles and therefore may be extended to a well-defined function on $\mathcal{E}$. To see this, note that any two representations of the same set as a finite disjoint union of rectangles may be decomposed into a common refinement such that each of the original rectangles is a disjoint union of rectangles in the refinement. The following definition therefore makes sense.

Definition 5.12. Suppose that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are measure spaces and $\mathcal{E}$ is the algebra generated by the measurable rectangles. The product premeasure $\lambda : \mathcal{E} \to [0, \infty]$ is given by

$$\lambda(E) = \sum_{i=1}^{N} \mu(A_i)\nu(B_i), \qquad E = \bigcup_{i=1}^{N} A_i \times B_i$$

where $E = \bigcup_{i=1}^{N} A_i \times B_i$ is any representation of $E \in \mathcal{E}$ as a disjoint union of measurable rectangles.

Proposition 5.11 implies that λ is countably additive on $\mathcal{E}$, since we may decompose any countable disjoint union of sets in $\mathcal{E}$ into a countable common disjoint refinement of rectangles, so λ is a premeasure as claimed. The outer product measure associated with λ, which we write as $(\mu \otimes \nu)^*$, is defined in terms of countable coverings by measurable rectangles. This gives the following.

Definition 5.13. Suppose that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are measure spaces. Then the product outer measure

$$(\mu \otimes \nu)^* : \mathcal{P}(X \times Y) \to [0, \infty]$$

on $X \times Y$ is defined for $E \subset X \times Y$ by

$$(\mu \otimes \nu)^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i) : E \subset \bigcup_{i=1}^{\infty}(A_i \times B_i) \text{ where } A_i \in \mathcal{A},\ B_i \in \mathcal{B} \right\}.$$

The product measure

$$(\mu \otimes \nu) : \mathcal{A} \otimes \mathcal{B} \to [0, \infty], \qquad (\mu \otimes \nu) = (\mu \otimes \nu)^*|_{\mathcal{A}\otimes\mathcal{B}}$$

is the restriction of the product outer measure to the product σ-algebra.

It follows from Proposition 5.7 that $(\mu \otimes \nu)^*$ is an outer measure and every measurable rectangle is $(\mu \otimes \nu)^*$-measurable with measure equal to its product premeasure. We summarize the conclusions of the Carátheodory theorem and Theorem 5.9 in the case of product measures as the following result.

Theorem 5.14. *If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are measure spaces, then*

$$(\mu \otimes \nu) : \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$$

is a measure on $X \times Y$ such that

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B) \qquad \text{for every } A \in \mathcal{A},\ B \in \mathcal{B}.$$

Moreover, if $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces, then $(\mu \otimes \nu)$ is the unique measure on $\mathcal{A} \otimes \mathcal{B}$ with this property.

Note that, in general, the σ-algebra of Carathéodory measurable sets associated with $(\mu \otimes \nu)^*$ is strictly larger than the product σ-algebra. For example, if $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with Lebesgue measure defined on their Borel σ-algebras, then the Carathéodory σ-algebra on the product $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$ is the Lebesgue σ-algebra $\mathcal{L}(\mathbb{R}^{m+n})$, whereas the product σ-algebra is the Borel σ-algebra $\mathcal{B}(\mathbb{R}^{m+n})$.

5.4. Measurable functions

If $f : X \times Y \to \mathbb{C}$ is a function of $(x, y) \in X \times Y$, then for each $x \in X$ we define the x-section $f_x : Y \to \mathbb{C}$ and for each $y \in Y$ we define the y-section $f^y : Y \to \mathbb{C}$ by

$$f_x(y) = f(x, y), \qquad f^y(x) = f(x, y).$$

Theorem 5.15. *If $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ are measure spaces and $f : X \times Y \to \mathbb{C}$ is a measurable function, then $f_x : Y \to \mathbb{C}$, $f^y : X \to \mathbb{C}$ are measurable for every $x \in X$, $y \in Y$. Moreover, if $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ are σ-finite, then the functions $g : X \to \mathbb{C}$, $h : Y \to \mathbb{C}$ defined by*

$$g(x) = \int f_x \, d\nu, \qquad h(y) = \int f^y \, d\mu$$

are measurable.

5.5. Monotone class theorem

We prove a general result about σ-algebras, called the monotone class theorem, which we will use in proving Fubini's theorem. A collection of subsets of a set is called a monotone class if it is closed under countable increasing unions and countable decreasing intersections.

Definition 5.16. A monotone class on a set X is a collection $\mathcal{C} \subset \mathcal{P}(X)$ of subsets of X such that if $E_i, F_i \in \mathcal{C}$ and

$$E_1 \subset E_2 \subset \dots \subset E_i \subset \dots, \qquad F_1 \supset F_2 \supset \dots \supset F_i \supset \dots,$$

then

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{C}, \qquad \bigcap_{i=1}^{\infty} F_i \in \mathcal{C}.$$

Obviously, every σ-algebra is a monotone class, but not conversely. As with σ-algebras, every family $\mathcal{F} \subset \mathcal{P}(X)$ of subsets of a set X is contained in a smallest monotone class, called the monotone class generated by $\mathcal{F}$, which is the intersection of all monotone classes on X that contain $\mathcal{F}$. As stated in the next theorem, if $\mathcal{F}$ is an algebra, then this monotone class is, in fact, a σ-algebra.

Theorem 5.17 (Monotone Class Theorem)**.** *If $\mathcal{F}$ is an algebra of sets, the monotone class generated by $\mathcal{F}$ coincides with the σ-algebra generated by $\mathcal{F}$.*

5.6. Fubini's theorem

Theorem 5.18 (Fubini's Theorem)**.** *Suppose that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces. A measurable function $f : X \times Y \to \mathbb{C}$ is integrable if and only if either one of the iterated integrals*

$$\int \left(\int |f^y| \, d\mu \right) d\nu, \qquad \int \left(\int |f_x| \, d\nu \right) d\mu$$

is finite. In that case

$$\int f \, d\mu \otimes d\nu = \int \left(\int f^y \, d\mu \right) d\nu = \int \left(\int f_x \, d\nu \right) d\mu.$$

Example 5.19. An application of Fubini's theorem to counting measure on $\mathbb{N} \times \mathbb{N}$ implies that if $\{a_{mn} \in \mathbb{C} \mid m, n \in \mathbb{N}\}$ is a doubly-indexed sequence of complex numbers such that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |a_{mn}| \right) < \infty$$

then

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn} \right) = \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{mn} \right).$$

5.7. Completion of product measures

The product of complete measure spaces is not necessarily complete.

Example 5.20. If $N \subset \mathbb{R}$ is a non-Lebesgue measurable subset of $\mathbb{R}$, then $\{0\} \times N$ does not belong to the product σ-algebra $\mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R})$ on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, since every section of a set in the product σ-algebra is measurable. It does, however, belong to $\mathcal{L}(\mathbb{R}^2)$, since it is a subset of the set $\{0\} \times \mathbb{R}$ of two-dimensional Lebesgue measure zero, and Lebesgue measure is complete. Instead one can show that the Lebesgue σ-algebra on $\mathbb{R}^{m+n}$ is the completion with respect to Lebesgue measure of the product of the Lebesgue σ-algebras on $\mathbb{R}^m$ and $\mathbb{R}^n$:

$$\mathcal{L}(\mathbb{R}^{m+n}) = \overline{\mathcal{L}(\mathbb{R}^m) \otimes \mathcal{L}(\mathbb{R}^n)}.$$

We state a version of Fubini's theorem for Lebesgue measurable functions on $\mathbb{R}^n$.

Theorem 5.21. *A Lebesgue measurable function $f : \mathbb{R}^{m+n} \to \mathbb{C}$ is integrable, meaning that*

$$\int_{\mathbb{R}^{m+n}} |f(x,y)|\, dxdy < \infty,$$

if and only if either one of the iterated integrals

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} |f(x,y)|\, dx \right) dy, \qquad \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} |f(x,y)|\, dy \right) dx$$

is finite. In that case,

$$\int_{\mathbb{R}^{m+n}} f(x,y)\, dxdy = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x,y)\, dx \right) dy = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x,y)\, dy \right) dx,$$

where all of the integrals are well-defined and finite a.e.

CHAPTER 6

Differentiation

The generalization from elementary calculus of differentiation in measure theory is less obvious than that of integration, and the methods of treating it are somewhat involved.

Consider the fundamental theorem of calculus (FTC) for smooth functions of a single variable. In one direction (FTC-I, say) it states that the derivative of the integral is the original function, meaning that

$$f(x) = \frac{d}{dx}\int_a^x f(y)\,dy. \tag{6.1}$$

In the other direction (FTC-II, say) it states that we recover the original function by integrating its derivative

$$F(x) = F(a) + \int_a^x f(y)\,dy, \qquad f = F'. \tag{6.2}$$

As we will see, (6.1) holds pointwise a.e. provided that f is locally integrable, which is needed to ensure that the right-hand side is well-defined. Equation (6.2), however, does not hold for all continuous functions F whose pointwise derivative is defined a.e. and integrable; we also need to require that F is absolutely continuous. The Cantor function is a counter-example.

First, we consider a generalization of (6.1) to locally integrable functions on $\mathbb{R}^n$, which leads to the Lebesgue differentiation theorem. We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is locally integrable if it is Lebesgue measurable and

$$\int_K |f|\,dx < \infty$$

for every compact subset $K \subset \mathbb{R}^n$; we denote the space of locally integrable functions by $L^1_{\text{loc}}(\mathbb{R}^n)$.

Let

$$B_r(x) = \{y \in \mathbb{R}^n : |y - x| < r\} \tag{6.3}$$

denote the open ball of radius r and center $x \in \mathbb{R}^n$. We denote Lebesgue measure on $\mathbb{R}^n$ by μ and the Lebesgue measure of a ball B by $\mu(B) = |B|$.

To motivate the statement of the Lebesgue differentiation theorem, observe that (6.1) may be written in terms of symmetric differences as

$$f(x) = \lim_{r\to 0^+} \frac{1}{2r}\int_{x-r}^{x+r} f(y)\,dy. \tag{6.4}$$

In other words, the value of f at a point x is the limit of local averages of f over intervals centered at x as their lengths approach zero. An n-dimensional version of

(6.4) is

$$f(x) = \lim_{r\to 0^+} \frac{1}{|B_r(x)|} \int_{B_r(x)} f(y)\, dy \tag{6.5}$$

where the integral is with respect n-dimensional Lebesgue measure. The Lebesgue differentiation theorem states that (6.5) holds pointwise μ-a.e. for any locally integrable function f.

To prove the theorem, we will introduce the maximal function of an integrable function, whose key property is that it is weak-L^1, as stated in the Hardy-Littlewood theorem. This property may be shown by the use of a simple covering lemma, which we begin by proving.

Second, we consider a generalization of (6.2) on the representation of a function as an integral. In defining integrals on a general measure space, it is natural to think of them as defined on sets rather than real numbers. For example, in (6.2), we would write $F(x) = \nu([a, x])$ where $\nu : \mathcal{B}([a, b]) \to \mathbb{R}$ is a signed measure. This interpretation leads to the following question: if μ, ν are measures on a measurable space X is there a function $f : X \to [0, \infty]$ such that

$$\nu(A) = \int_A f\, d\mu.$$

If so, we regard $f = d\nu/d\mu$ as the (Radon-Nikodym) derivative of ν with respect to μ. More generally, we may consider signed (or complex) measures, whose values are not restricted to positive numbers. The Radon-Nikodym theorem gives a necessary and sufficient condition for the differentiability of ν with respect to μ, subject to a σ-finiteness assumption: namely, that ν is absolutely continuous with respect to μ.

6.1. A covering lemma

We need only the following simple form of a covering lemma; there are many more sophisticated versions, such as the Vitali and Besicovitch covering theorems, which we do not consider here.

Lemma 6.1. *Let $\{B_1, B_2, \dots, B_N\}$ be a finite collection of open balls in $\mathbb{R}^n$. There is a disjoint subcollection $\{B_1', B_2', \dots, B_M'\}$ where $B_j' = B_{i_j}$, such that*

$$\mu\left(\bigcup_{i=1}^N B_i\right) \le 3^n \sum_{i=1}^M \left|B_j'\right|.$$

PROOF. If B is an open ball, let $\widehat{B}$ denote the open ball with the same center as B and three times the radius. Then

$$|\widehat{B}| = 3^n|B|.$$

Moreover, if B_1, B_2 are nondisjoint open balls and the radius of B_1 is greater than or equal to the radius of B_2, then $\widehat{B}_1 \supset B_2$.

We obtain the subfamily $\{B_j'\}$ by an iterative procedure. Choose B_1' to be a ball with the largest radius from the collection $\{B_1, B_2, \dots, B_N\}$. Delete from the collection all balls B_i that intersect B_1', and choose B_2' to be a ball with the largest radius from the remaining balls. Repeat this process until the balls are exhausted, which gives $M \le N$ balls, say.

By construction, the balls $\{B_1', B_2', \dots, B_M'\}$ are disjoint and

$$\bigcup_{i=1}^{N} B_i \subset \bigcup_{j=1}^{M} \widehat{B}_j'.$$

It follows that

$$\mu\left(\bigcup_{i=1}^{N} B_i\right) \le \sum_{i=1}^{M} \left|\widehat{B}_j'\right| = 3^n \sum_{i=1}^{M} \left|B_j'\right|,$$

which proves the result. □

6.2. Maximal functions

The maximal function of a locally integrable function is obtained by taking the supremum of averages of the absolute value of the function about a point. Maximal functions were introduced by Hardy and Littlewood (1930), and they are the key to proving pointwise properties of integrable functions. They also play a central role in harmonic analysis.

Definition 6.2. If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then the maximal function Mf of f is defined by

$$Mf(x) = \sup_{r>0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y)|\, dy.$$

The use of centered open balls to define the maximal function is for convenience. We could use non-centered balls or other sets, such as cubes, to define the maximal function. Some restriction on the shapes on the sets is, however, required; for example, we cannot use arbitrary rectangles, since averages over progressively longer and thinner rectangles about a point whose volumes shrink to zero do not, in general, converge to the value of the function at the point, even if the function is continuous.

Note that any two functions that are equal a.e. have the same maximal function.

Example 6.3. If $f : \mathbb{R} \to \mathbb{R}$ is the step function

$$f(x) = \begin{cases} 1 & \text{if } x \ge 0, \\ 0 & \text{if } x < 0, \end{cases}$$

then

$$Mf(x) = \begin{cases} 1 & \text{if } x > 0, \\ 1/2 & \text{if } x \le 0. \end{cases}$$

This example illustrates the following result.

Proposition 6.4. *If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then the maximal function Mf is lower semi-continuous and therefore Borel measurable.*

Proof. The function $Mf \ge 0$ is lower semi-continuous if

$$E_t = \{x : Mf(x) > t\}$$

is open for every $0 < t < \infty$. To prove that E_t is open, let $x \in E_t$. Then there exists $r > 0$ such that

$$\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y)|\, dy > t.$$

Choose $r' > r$ such that we still have

$$\frac{1}{|B_{r'}(x)|}\int_{B_r(x)} |f(y)|\,dy > t.$$

If $|x' - x| < r' - r$, then $B_r(x) \subset B_{r'}(x')$, so

$$t < \frac{1}{|B_{r'}(x)|}\int_{B_r(x)} |f(y)|\,dy \le \frac{1}{|B_{r'}(x')|}\int_{B_{r'}(x')} |f(y)|\,dy \le Mf(x'),$$

It follows that $x' \in E_t$, which proves that E_t is open. □

The maximal function of a non-zero function $f \in L^1(\mathbb{R}^n)$ is not in $L^1(\mathbb{R}^n)$ because it decays too slowly at infinity for its integral to converge. To show this, let $a > 0$ and suppose that $|x| \ge a$. Then, by considering the average of $|f|$ at x over a ball of radius $r = 2|x|$ and using the fact that $B_{2|x|}(x) \supset B_a(0)$, we see that

$$\begin{aligned} Mf(x) &\ge \frac{1}{|B_{2|x|}(x)|}\int_{B_{2|x|}(x)} |f(y)|\,dy \\ &\ge \frac{C}{|x|^n}\int_{B_a(0)} |f(y)|\,dy, \end{aligned}$$

where $C > 0$. The function $1/|x|^n$ is not integrable on $\mathbb{R}^n \setminus B_a(0)$, so if Mf is integrable then we must have

$$\int_{B_a(0)} |f(y)|\,dy = 0$$

for every $a > 0$, which implies that $f = 0$ a.e. in $\mathbb{R}^n$.

Moreover, as the following example shows, the maximal function of an integrable function need not even be locally integrable.

Example 6.5. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1/(x\log^2 x) & \text{if } 0 < x < 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

The change of variable $u = \log x$ implies that

$$\int_0^{1/2} \frac{1}{x|\log x|^n}\,dx$$

is finite if and only if $n > 1$. Thus $f \in L^1(\mathbb{R})$ and for $0 < x < 1/2$

$$\begin{aligned} Mf(x) &\ge \frac{1}{2x}\int_0^{2x} |f(y)|\,dy \\ &\ge \frac{1}{2x}\int_0^{x} \frac{1}{y\log^2 y}\,dy \\ &\ge \frac{1}{2x|\log x|} \end{aligned}$$

so $Mf \notin L^1_{\mathrm{loc}}(\mathbb{R})$.

6.3. Weak-L^1 spaces

Although the maximal function of an integrable function is not integrable, it is not much worse than an integrable function. As we show in the next section, it belongs to the space weak-L^1, which is defined as follows

Definition 6.6. The space weak-$L^1(\mathbb{R}^n)$ consists of measurable functions

$$f : \mathbb{R}^n \to \mathbb{R}$$

such that there exists a constant C, depending on f but not on t, with the property that for every $0 < t < \infty$

$$\mu \{x \in \mathbb{R}^n : |f(x)| > t\} \leq \frac{C}{t}.$$

An estimate of this form arises for integrable function from the following, almost trivial, Chebyshev inequality.

Theorem 6.7 (Chebyshev's inequality). *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space. If $f : X \to \mathbb{R}$ is integrable and $0 < t < \infty$, then*

$$\mu (\{x \in X : |f(x)| > t\}) \leq \frac{1}{t} \|f\|_{L^1} . \tag{6.6}$$

PROOF. Let $E_t = \{x \in X : |f(x)| > t\}$. Then

$$\int |f| \, d\mu \geq \int_{E_t} |f| \, d\mu \geq t\mu(E_t),$$

which proves the result. □

Chebyshev's inequality implies immediately that if f belongs to $L^1(\mathbb{R}^n)$, then f belongs to weak-$L^1(\mathbb{R}^n)$. The converse statement is, however, false.

Example 6.8. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{x}$$

for $x \neq 0$ satisfies

$$\mu \{x \in \mathbb{R} : |f(x)| > t\} = \frac{2}{t},$$

so f belongs to weak-$L^1(\mathbb{R})$, but f is not integrable or even locally integrable.

6.4. Hardy-Littlewood theorem

The following Hardy-Littlewood theorem states that the maximal function of an integrable function is weak-L^1.

Theorem 6.9 (Hardy-Littlewood). *If $f \in L^1(\mathbb{R}^n)$, there is a constant C such that for every $0 < t < \infty$*

$$\mu (\{x \in \mathbb{R}^n : Mf(x) > t\}) \leq \frac{C}{t} \|f\|_{L^1}$$

where $C = 3^n$ depends only on n.

Proof. Fix $t > 0$ and let

$$E_t = \{x \in \mathbb{R}^n : Mf(x) > t\}.$$

By the inner regularity of Lebesgue measure

$$\mu(E_t) = \sup\{\mu(K) : K \subset E_t \text{ is compact}\}$$

so it is enough to prove that

$$\mu(K) \le \frac{C}{t} \int_{\mathbb{R}^n} |f(y)|\, dy.$$

for every compact subset K of E_t.

If $x \in K$, then there is an open ball B_x centered at x such that

$$\frac{1}{|B_x|} \int_{B_x} |f(y)|\, dy > t.$$

Since K is compact, we may extract a finite subcover $\{B_1, B_2, \dots, B_N\}$ from the open cover $\{B_x : x \in K\}$. By Lemma 6.1, there is a finite subfamily of disjoint balls $\{B_1', B_2', \dots, B_M'\}$ such that

$$\begin{aligned}
\mu(K) &\le \sum_{i=1}^{N} |B_i| \\
&\le 3^n \sum_{j=1}^{M} |B_j'| \\
&\le \frac{3^n}{t} \sum_{j=1}^{M} \int_{B_j'} |f|\, dx \\
&\le \frac{3^n}{t} \int |f|\, dx,
\end{aligned}$$

which proves the result with $C = 3^n$. □

6.5. Lebesgue differentiation theorem

The maximal function provides the crucial estimate in the following proof.

Theorem 6.10. *If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then for a.e. $x \in \mathbb{R}^n$*

$$\lim_{r \to 0^+} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} f(y)\, dy \right] = f(x).$$

Moreover, for a.e. $x \in \mathbb{R}^n$

$$\lim_{r \to 0^+} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy \right] = 0.$$

Proof. Since

$$\left| \frac{1}{|B_r(x)|} \int_{B_r(x)} f(y) - f(x)\, dy \right| \le \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy,$$

we just need to prove the second result. We define $f^* : \mathbb{R}^n \to [0, \infty]$ by

$$f^*(x) = \limsup_{r \to 0^+} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy \right].$$

We want to show that $f^* = 0$ pointwise a.e.

If $g \in C_c(\mathbb{R}^n)$ is continuous, then given any $\epsilon > 0$ there exists $\delta > 0$ such that $|g(x) - g(y)| < \epsilon$ whenever $|x - y| < \delta$. Hence if $r < \delta$

$$\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy < \epsilon,$$

which implies that $g^* = 0$. We prove the result for general f by approximation with a continuous function.

First, note that we can assume that $f \in L^1(\mathbb{R}^n)$ is integrable without loss of generality; for example, if the result holds for $f\chi_{B_k(0)} \in L^1(\mathbb{R}^n)$ for each $k \in \mathbb{N}$ except on a set E_k of measure zero, then it holds for $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ except on $\bigcup_{k=1}^{\infty} E_k$, which has measure zero.

Next, observe that since

$$|f(y) + g(y) - [f(x) + g(x)]| \leq |f(y) - f(x)| + |g(y) - g(x)|$$

and $\limsup(A + B) \leq \limsup A + \limsup B$, we have

$$(f + g)^* \leq f^* + g^*.$$

Thus, if $f \in L^1(\mathbb{R}^n)$ and $g \in C_c(\mathbb{R}^n)$, we have

$$(f - g)^* \leq f^* + g^* = f^*,$$
$$f^* = (f - g + g)^* \leq (f - g)^* + g^* = (f - g)^*,$$

which shows that $(f - g)^* = f^*$.

If $f \in L^1(\mathbb{R}^n)$, then we claim that there is a constant C, depending only on n, such that for every $0 < t < \infty$

$$\mu\left(\{x \in \mathbb{R}^n : f^*(x) > t\}\right) \leq \frac{C}{t} \|f\|_{L^1}. \tag{6.7}$$

To show this, we estimate

$$\begin{aligned} f^*(x) &\leq \sup_{r>0} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy\right] \\ &\leq \sup_{r>0} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y)|\, dy\right] + |f(x)| \\ &\leq Mf(x) + |f(x)|. \end{aligned}$$

It follows that

$$\{f^* > t\} \subset \{Mf + |f| > t\} \subset \{Mf > t/2\} \cup \{|f| > t/2\}.$$

By the Hardy-Littlewood theorem,

$$\mu\left(\{x \in \mathbb{R}^n : Mf(x) > t/2\}\right) \leq \frac{2 \cdot 3^n}{t} \|f\|_{L^1},$$

and by the Chebyshev inequality

$$\mu\left(\{x \in \mathbb{R}^n : |f(x)| > t/2\}\right) \leq \frac{2}{t} \|f\|_{L^1}.$$

Combining these estimates, we conclude that (6.7) holds with $C = 2\,(3^n + 1)$.

Finally suppose that $f \in L^1(\mathbb{R}^n)$ and $0 < t < \infty$. From Theorem 4.27, for any $\epsilon > 0$, there exists $g \in C_c(\mathbb{R}^n)$ such that $\|f - g\|_{L^1} < \epsilon$. Then

$$\begin{aligned}\mu\left(\{x \in \mathbb{R}^n : f^*(x) > t\}\right) &= \mu\left(\{x \in \mathbb{R}^n : (f-g)^*(x) > t\}\right) \\ &\le \frac{C}{t}\|f - g\|_{L^1} \\ &\le \frac{C\epsilon}{t}.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\mu\left(\{x \in \mathbb{R}^n : f^*(x) > t\}\right) = 0,$$

and hence since

$$\{x \in \mathbb{R}^n : f^*(x) > 0\} = \bigcup_{k=1}^{\infty} \{x \in \mathbb{R}^n : f^*(x) > 1/k\}$$

that

$$\mu\left(\{x \in \mathbb{R}^n : f^*(x) > 0\}\right) = 0.$$

This proves the result. □

The set of points x for which the limits in Theorem 6.10 exist for a suitable definition of $f(x)$ is called the Lebesgue set of f.

Definition 6.11. If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then a point $x \in \mathbb{R}^n$ belongs to the Lebesgue set of f if there exists a constant $c \in \mathbb{R}$ such that

$$\lim_{r \to 0^+} \left[\frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - c|\, dy\right] = 0.$$

If such a constant c exists, then it is unique. Moreover, its value depends only on the equivalence class of f with respect to pointwise a.e. equality. Thus, we can use this definition to give a canonical pointwise a.e. representative of a function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ that is defined on its Lebesgue set.

Example 6.12. The Lebesgue set of the step function f in Example 6.3 is $\mathbb{R} \setminus \{0\}$. The point 0 does not belong to the Lebesgue set, since

$$\lim_{r \to 0^+} \left[\frac{1}{2r} \int_{-r}^{r} |f(y) - c|\, dy\right] = \frac{1}{2}\left(|c| + |1 - c|\right)$$

is nonzero for every $c \in \mathbb{R}$. Note that the existence of the limit

$$\lim_{r \to 0^+} \left[\frac{1}{2r} \int_{-r}^{r} f(y)\, dy\right] = \frac{1}{2}$$

is not sufficient to imply that 0 belongs to the Lebesgue set of f.

6.6. Signed measures

A signed measure is a countably additive, extended real-valued set function whose values are not required to be positive. Measures may be thought of as a generalization of volume or mass, and signed measures may be thought of as a generalization of charge, or a similar quantity. We allow a signed measure to take infinite values, but to avoid undefined expressions of the form $\infty - \infty$, it should not take both positive and negative infinite values.

Definition 6.13. Let $(X, \mathcal{A})$ be a measurable space. A signed measure ν on X is a function $\nu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

(a) $\nu(\varnothing) = 0$;
(b) ν attains at most one of the values ∞, $-\infty$;
(c) if $\{A_i \in \mathcal{A} : i \in \mathbb{N}\}$ is a disjoint collection of measurable sets, then

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i).$$

We say that a signed measure is finite if it takes only finite values. Note that since $\nu\left(\bigcup_{i=1}^{\infty} A_i\right)$ does not depend on the order of the A_i, the sum $\sum_{i=1}^{\infty} \nu(A_i)$ converges unconditionally if it is finite, and therefore it is absolutely convergent. Signed measures have the same monotonicity property (1.1) as measures, with essentially the same proof. We will always refer to signed measures explicitly, and 'measure' will always refer to a positive measure.

Example 6.14. If $(X, \mathcal{A}, \mu)$ is a measure space and $\nu^+, \nu^- : \mathcal{A} \to [0, \infty]$ are measures, one of which is finite, then $\nu = \nu^+ - \nu^-$ is a signed measure.

Example 6.15. If $(X, \mathcal{A}, \mu)$ is a measure space and $f : X \to \overline{\mathbb{R}}$ is an $\mathcal{A}$-measurable function whose integral with respect to μ is defined as an extended real number, then $\nu : \mathcal{A} \to \overline{\mathbb{R}}$ defined by

$$\nu(A) = \int_A f \, d\mu \tag{6.8}$$

is a signed measure on X. As we describe below, we interpret f as the derivative $d\nu/d\mu$ of ν with respect to μ. If $f = f^+ - f^-$ is the decomposition of f into positive and negative parts then $\nu = \nu^+ - \nu^-$, where the measures $\nu^+, \nu^- : \mathcal{A} \to [0, \infty]$ are defined by

$$\nu^+(A) = \int_A f^+ \, d\mu, \qquad \nu^-(A) = \int_A f^- \, d\mu.$$

We will show that any signed measure can be decomposed into a difference of singular measures, called its Jordan decomposition. Thus, Example 6.14 includes all signed measures. Not all signed measures have the form given in Example 6.15. As we discuss this further in connection with the Radon-Nikodym theorem, a signed measure ν of the form (6.8) must be absolutely continuous with respect to the measure μ.

6.7. Hahn and Jordan decompositions

To prove the Jordan decomposition of a signed measure, we first show that a measure space can be decomposed into disjoint subsets on which a signed measure is positive or negative, respectively. This is called the Hahn decomposition.

Definition 6.16. Suppose that ν is a signed measure on a measurable space X. A set $A \subset X$ is positive for ν if it is measurable and $\nu(B) \geq 0$ for every measurable subset $B \subset A$. Similarly, A is negative for ν if it is measurable and $\nu(B) \leq 0$ for every measurable subset $B \subset A$, and null for ν if it is measurable and $\nu(B) = 0$ for every measurable subset $B \subset A$.

Because of the possible cancelation between the positive and negative signed measure of subsets, $\nu(A) > 0$ does not imply that A is positive for ν, nor does $\nu(A) = 0$ imply that A is null for ν. Nevertheless, as we show in the next result, if $\nu(A) > 0$, then A contains a subset that is positive for ν. The idea of the (slightly tricky) proof is to remove subsets of A with negative signed measure until only a positive subset is left.

Lemma 6.17. *Suppose that ν is a signed measure on a measurable space $(X, \mathcal{A})$. If $A \in \mathcal{A}$ and $0 < \nu(A) < \infty$, then there exists a positive subset $P \subset A$ such that $\nu(P) > 0$.*

PROOF. First, we show that if $A \in \mathcal{A}$ is a measurable set with $|\nu(A)| < \infty$, then $|\nu(B)| < \infty$ for every measurable subset $B \subset A$. This is because ν takes at most one infinite value, so there is no possibility of canceling an infinite signed measure to give a finite measure. In more detail, we may suppose without loss of generality that $\nu : \mathcal{A} \to [-\infty, \infty)$ does not take the value ∞. (Otherwise, consider $-\nu$.) Then $\nu(B) \neq \infty$; and if $B \subset A$, then the additivity of ν implies that

$$\nu(B) = \nu(A) - \nu(A \setminus B) \neq -\infty$$

since $\nu(A)$ is finite and $\nu(A \setminus B) \neq \infty$.

Now suppose that $0 < \nu(A) < \infty$. Let

$$\delta_1 = \inf \{\nu(E) : E \in \mathcal{A} \text{ and } E \subset A\}.$$

Then $-\infty \leq \delta_1 \leq 0$, since $\varnothing \subset A$. Choose $A_1 \subset A$ such that $\delta_1 \leq \nu(A_1) \leq \delta_1/2$ if δ_1 is finite, or $\mu(A_1) \leq -1$ if $\delta_1 = -\infty$. Define a disjoint sequence of subsets $\{A_i \subset A : i \in \mathbb{N}\}$ inductively by setting

$$\delta_i = \inf \left\{\nu(E) : E \in \mathcal{A} \text{ and } E \subset A \setminus \left(\bigcup_{j=1}^{i-1} A_j\right)\right\}$$

and choosing $A_i \subset A \setminus \left(\bigcup_{j=1}^{i-1} A_j\right)$ such that

$$\delta_i \leq \nu(A_i) \leq \frac{1}{2}\delta_i$$

if $-\infty < \delta_i \leq 0$, or $\nu(A_i) \leq -1$ if $\delta_i = -\infty$.

Let

$$B = \bigcup_{i=1}^{\infty} A_i, \qquad P = A \setminus B.$$

Then, since the A_i are disjoint, we have

$$\nu(B) = \sum_{i=1}^{\infty} \nu(A_i).$$

As proved above, $\nu(B)$ is finite, so this negative sum must converge. It follows that $\nu(A_i) \leq -1$ for only finitely many i, and therefore δ_i is infinite for at most finitely many i. For the remaining i, we have

$$\sum \nu(A_i) \leq \frac{1}{2} \sum \delta_i \leq 0,$$

so $\sum \delta_i$ converges and therefore $\delta_i \to 0$ as $i \to \infty$.

If $E \subset P$, then by construction $\nu(E) \geq \delta_i$ for every sufficiently large $i \in \mathbb{N}$. Hence, taking the limit as $i \to \infty$, we see that $\nu(E) \geq 0$, which implies that P is positive. The proof also shows that, since $\nu(B) \leq 0$, we have

$$\nu(P) = \nu(A) - \nu(B) \geq \nu(A) > 0,$$

which proves that P has strictly positive signed measure. □

The Hahn decomposition follows from this result in a straightforward way.

Theorem 6.18 (Hahn decomposition)**.** *If ν is a signed measure on a measurable space $(X, \mathcal{A})$, then there is a positive set P and a negative set N for ν such that $P \cup N = X$ and $P \cap N = \varnothing$. These sets are unique up to ν-null sets.*

Proof. Suppose, without loss of generality, that $\nu(A) < \infty$ for every $A \in \mathcal{A}$. (Otherwise, consider $-\nu$.) Let

$$m = \sup\{\nu(A) : A \in \mathcal{A} \text{ such that } A \text{ is positive for } \nu\},$$

and choose a sequence $\{A_i : i \in \mathbb{N}\}$ of positive sets such that $\nu(A_i) \to m$ as $i \to \infty$. Then, since the union of positive sets is positive,

$$P = \bigcup_{i=1}^{\infty} A_i$$

is a positive set. Moreover, by the monotonicity of of ν, we have $\nu(P) = m$. Since $\nu(P) \neq \infty$, it follows that $m \geq 0$ is finite.

Let $N = X \setminus P$. Then we claim that N is negative for ν. If not, there is a subset $A' \subset N$ such that $\nu(A') > 0$, so by Lemma 6.17 there is a positive set $P' \subset A'$ with $\nu(P') > 0$. But then $P \cup P'$ is a positive set with $\nu(P \cup P') > m$, which contradicts the definition of m.

Finally, if P', N' is another such pair of positive and negative sets, then

$$P \setminus P' \subset P \cap N',$$

so $P \setminus P'$ is both positive and negative for ν and therefore null, and similarly for $P' \setminus P$. Thus, the decomposition is unique up to ν-null sets. □

To describe the corresponding decomposition of the signed measure ν into the difference of measures, we introduce the notion of singular measures, which are measures that are supported on disjoint sets.

Definition 6.19. Two measures μ, ν on a measurable space $(X, \mathcal{A})$ are singular, written $\mu \perp \nu$, if there exist sets $M, N \in \mathcal{A}$ such that $M \cap N = \varnothing$, $M \cup N = X$ and $\mu(M) = 0$, $\nu(N) = 0$.

Example 6.20. The δ-measure in Example 2.36 and the Cantor measure in Example 2.37 are singular with respect to Lebesgue measure on $\mathbb{R}$ (and conversely, since the relation is symmetric).

Theorem 6.21 (Jordan decomposition)**.** *If ν is a signed measure on a measurable space $(X, \mathcal{A})$, then there exist unique measures $\nu^+, \nu^- : \mathcal{A} \to [0, \infty]$, one of which is finite, such that*

$$\nu = \nu^+ - \nu^- \qquad \text{and } \nu^+ \perp \nu^-.$$

Proof. Let $X = P \cup N$ where P, N are positive, negative sets for ν. Then

$$\nu^+(A) = \nu(A \cap P), \qquad \nu^-(A) = -\nu(A \cap N)$$

is the required decomposition. The values of $\nu^\pm$ are independent of the choice of P, N up to a ν-null set, so the decomposition is unique. □

We call ν^+ and ν^- the positive and negative parts of ν, respectively. The total variation $|\nu|$ of ν is the measure

$$|\nu| = \nu^+ + \nu^-.$$

We say that the signed measure ν is σ-finite if $|\nu|$ is σ-finite.

6.8. Radon-Nikodym theorem

The absolute continuity of measures is in some sense the opposite relationship to the singularity of measures. If a measure ν singular with respect to a measure μ, then it is supported on different sets from μ, while if ν is absolutely continuous with respect to μ, then it supported on on the same sets as μ.

Definition 6.22. Let ν be a signed measure and μ a measure on a measurable space $(X, \mathcal{A})$. Then ν is absolutely continuous with respect to μ, written $\nu \ll \mu$, if $\nu(A) = 0$ for every set $A \in \mathcal{A}$ such that $\mu(A) = 0$.

Equivalently, $\nu \ll \mu$ if every μ-null set is a ν-null set. Unlike singularity, absolute continuity is not symmetric.

Example 6.23. If μ is Lebesgue measure and ν is counting measure on $\mathcal{B}(\mathbb{R})$, then $\mu \ll \nu$, but $\nu \not\ll \mu$.

Example 6.24. If $f : X \to \overline{\mathbb{R}}$ is a measurable function on a measure space $(X, \mathcal{A}, \mu)$ whose integral with respect μ is well-defined as an extended real number and the signed measure $\nu : \mathcal{A} \to \overline{\mathbb{R}}$ is defined by

$$\nu(A) = \int_A f \, d\mu,$$

then (4.4) shows that ν is absolutely continuous with respect to μ.

The next result clarifies the relation between Definition 6.22 and the absolute continuity property of integrable functions proved in Proposition 4.16.

Proposition 6.25. *If ν is a finite signed measure and μ is a measure, then $\nu \ll \mu$ if and only if for every $\epsilon > 0$, there exists $\delta > 0$ such that $|\nu(A)| < \epsilon$ whenever $\mu(A) < \delta$.*

Proof. Suppose that the given condition holds. If $\mu(A) = 0$, then $|\nu(A)| < \epsilon$ for every $\epsilon > 0$, so $\nu(A) = 0$, which shows that $\nu \ll \mu$.

Conversely, suppose that the given condition does not hold. Then there exists $\epsilon > 0$ such that for every $k \in \mathbb{N}$ there exists a measurable set A_k with $|\nu|(A_k) \geq \epsilon$ and $\mu(A_k) < 1/2^k$. Defining

$$B = \bigcap_{k=1}^\infty \bigcup_{j=k}^\infty A_j,$$

we see that $\mu(B) = 0$ but $|\nu|(B) \geq \epsilon$, so ν is not absolutely continuous with respect to μ. □

The Radon-Nikodym theorem provides a converse to Example 6.24 for absolutely continuous, σ-finite measures. As part of the proof, from [**4**], we also show that any signed measure ν can be decomposed into an absolutely continuous and singular part with respect to a measure μ (the Lebesgue decomposition of ν). In the proof of the theorem, we will use the following lemma.

Lemma 6.26. *Suppose that μ, ν are finite measures on a measurable space $(X, \mathcal{A})$. Then either $\mu \perp \nu$, or there exists $\epsilon > 0$ and a set P such that $\mu(P) > 0$ and P is a positive set for the signed measure $\nu - \epsilon\mu$.*

Proof. For each $n \in \mathbb{N}$, let $X = P_n \cup N_n$ be a Hahn decomposition of X for the signed measure $\nu - \frac{1}{n}\mu$. If

$$P = \bigcup_{n=1}^{\infty} P_n \qquad N = \bigcap_{n=1}^{\infty} N_n,$$

then $X = P \cup N$ is a disjoint union, and

$$0 \leq \nu(N) \leq \frac{1}{n}\mu(N)$$

for every $n \in \mathbb{N}$, so $\nu(N) = 0$. Thus, either $\mu(P) = 0$, when $\nu \perp \mu$, or $\mu(P_n) > 0$ for some $n \in \mathbb{N}$, which proves the result with $\epsilon = 1/n$. □

Theorem 6.27 (Lebesgue-Radon-Nikodym theorem)**.** *Let ν be a σ-finite signed measure and μ a σ-finite measure on a measurable space $(X, \mathcal{A})$. Then there exist unique σ-finite signed measures ν_a, ν_s such that*

$$\nu = \nu_a + \nu_s \qquad \text{where } \nu_a \ll \mu \text{ and } \nu_s \perp \mu.$$

Moreover, there exists a measurable function $f : X \to \overline{\mathbb{R}}$, uniquely defined up to μ-a.e. equivalence, such that

$$\nu_a(A) = \int_A f \, d\mu$$

for every $A \in \mathcal{A}$, where the integral is well-defined as an extended real number.

Proof. It is enough to prove the result when ν is a measure, since we may decompose a signed measure into its positive and negative parts and apply the result to each part.

First, we assume that μ, ν are finite. We will construct a function f and an absolutely continuous signed measure $\nu_a \ll \mu$ such that

$$\nu_a(A) = \int_A f \, d\mu \qquad \text{for all } A \in \mathcal{A}.$$

We write this equation as $d\nu_a = f \, d\mu$ for short. The remainder $\nu_s = \nu - \nu_a$ is the singular part of ν.

Let $\mathcal{F}$ be the set of all $\mathcal{A}$-measurable functions $g : X \to [0, \infty]$ such that

$$\int_A g \, d\mu \leq \nu(A) \qquad \text{for every } A \in \mathcal{A}.$$

We obtain f by taking a supremum of functions from $\mathcal{F}$. If $g, h \in \mathcal{F}$, then $\max\{g, h\} \in \mathcal{F}$. To see this, note that if $A \in \mathcal{A}$, then we may write $A = B \cup C$ where

$$B = A \cap \{x \in X : g(x) > h(x)\}, \qquad C = A \cap \{x \in X : g(x) \leq h(x)\},$$

and therefore

$$\int_A \max\{g, h\}\, d\mu = \int_B g\, d\mu + \int_C h\, d\mu \leq \nu(B) + \nu(C) = \nu(A).$$

Let

$$m = \sup \left\{ \int_X g\, d\mu : g \in \mathcal{F} \right\} \leq \nu(X).$$

Choose a sequence $\{g_n \in \mathcal{F} : n \in \mathbb{N}\}$ such that

$$\lim_{n\to\infty} \int_X g_n\, d\mu = m.$$

By replacing g_n with $\max\{g_1, g_2, \dots, g_n\}$, we may assume that $\{g_n\}$ is an increasing sequence of functions in $\mathcal{F}$. Let

$$f = \lim_{n\to\infty} g_n.$$

Then, by the monotone convergence theorem, for every $A \in \mathcal{A}$ we have

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A g_n\, d\mu \leq \nu(A),$$

so $f \in \mathcal{F}$ and

$$\int_X f\, d\mu = m.$$

Define $\nu_s : \mathcal{A} \to [0, \infty)$ by

$$\nu_s(A) = \nu(A) - \int_A f\, d\mu.$$

Then ν_s is a positive measure on X. We claim that $\nu_s \perp \mu$, which proves the result in this case. Suppose not. Then, by Lemma 6.26, there exists $\epsilon > 0$ and a set P with $\mu(P) > 0$ such that $\nu_s \geq \epsilon\mu$ on P. It follows that for any $A \in \mathcal{A}$

$$\begin{aligned}
\nu(A) &= \int_A f\, d\mu + \nu_s(A) \\
&\geq \int_A f\, d\mu + \nu_s(A \cap P) \\
&\geq \int_A f\, d\mu + \epsilon\mu(A \cap P) \\
&\geq \int_A (f + \epsilon\chi_P)\, d\mu.
\end{aligned}$$

It follows that $f + \epsilon\chi_P \in \mathcal{F}$ but

$$\int_X (f + \epsilon\chi_P)\, d\mu = m + \epsilon\mu(P) > m,$$

which contradicts the definition of m. Hence $\nu_s \perp \mu$.

If $\nu = \nu_a + \nu_s$ and $\nu = \nu_a' + \nu_s'$ are two such decompositions, then $\nu_a - \nu_a' = \nu_s' - \nu_s$ is both absolutely continuous and singular with respect to μ which implies that it is zero. Moreover, f is determined uniquely by ν_a up to pointwise a.e. equivalence.

Finally, if μ, ν are σ-finite measures, then we may decompose

$$X = \bigcup_{i=1}^{\infty} A_i$$

into a countable disjoint union of sets with $\mu(A_i) < \infty$ and $\nu(A_i) < \infty$. We decompose the finite measure $\nu_i = \nu|_{A_i}$ as

$$\nu_i = \nu_{ia} + \nu_{is} \qquad \text{where } \nu_{ia} \ll \mu_i \text{ and } \nu_{is} \perp \mu_i.$$

Then $\nu = \nu_a + \nu_s$ is the required decomposition with

$$\nu_a = \sum_{i=1}^{\infty} \nu_{ia}, \qquad \nu_s = \sum_{i=1}^{\infty} \nu_{ia}$$

is the required decomposition. □

The decomposition $\nu = \nu_a + \nu_s$ is called the Lebesgue decomposition of ν, and the representation of an absolutely continuous signed measure $\nu \ll \mu$ as $d\nu = f\, d\mu$ is the Radon-Nikodym theorem. We call the function f here the Radon-Nikodym derivative of ν with respect to μ, and denote it by

$$f = \frac{d\nu}{d\mu}.$$

Some hypothesis of σ-finiteness is essential in the theorem, as the following example shows.

Example 6.28. Let $\mathcal{B}$ be the Borel σ-algebra on $[0, 1]$, μ Lebesgue measure, and ν counting measure on $\mathcal{B}$. Then μ is finite and $\mu \ll \nu$, but ν is not σ-finite. There is no function $f : [0, 1] \to [0, \infty]$ such that

$$\mu(A) = \int_A f\, d\nu = \sum_{x \in A} f(x).$$

There are generalizations of the Radon-Nikodym theorem which apply to measures that are not σ-finite, but we will not consider them here.

6.9. Complex measures

Complex measures are defined analogously to signed measures, except that they are only permitted to take finite complex values.

Definition 6.29. Let $(X, \mathcal{A})$ be a measurable space. A complex measure ν on X is a function $\nu : \mathcal{A} \to \mathbb{C}$ such that:

(a) $\nu(\varnothing) = 0$;
(b) if $\{A_i \in \mathcal{A} : i \in \mathbb{N}\}$ is a disjoint collection of measurable sets, then

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i).$$

There is an analogous Radon-Nikodym theorems for complex measures. The Radon-Nikodym derivative of a complex measure is necessarily integrable, since the measure is finite.

Theorem 6.30 (Lebesgue-Radon-Nikodym theorem)**.** *Let ν be a complex measure and μ a σ-finite measure on a measurable space $(X, \mathcal{A})$. Then there exist unique complex measures ν_a, ν_s such that*

$$\nu = \nu_a + \nu_s \qquad \textit{where } \nu_a \ll \mu \textit{ and } \nu_s \perp \mu.$$

Moreover, there exists an integrable function $f : X \to \mathbb{C}$*, uniquely defined up to* μ*-a.e. equivalence, such that*

$$\nu_a(A) = \int_A f \, d\mu$$

for every $A \in \mathcal{A}$*.*

To prove the result, we decompose a complex measure into its real and imaginary parts, which are finite signed measures, and apply the corresponding theorem for signed measures.

CHAPTER 7

L^p spaces

In this Chapter we consider L^p-spaces of functions whose pth powers are integrable. We will not develop the full theory of such spaces here, but consider only those properties that are directly related to measure theory — in particular, density, completeness, and duality results. The fact that spaces of Lebesgue integrable functions are complete, and therefore Banach spaces, is another crucial reason for the success of the Lebesgue integral. The L^p-spaces are perhaps the most useful and important examples of Banach spaces.

7.1. L^p spaces

For definiteness, we consider real-valued functions. Analogous results apply to complex-valued functions.

Definition 7.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 \leq p < \infty$. The space $L^p(X)$ consists of equivalence classes of measurable functions $f : X \to \mathbb{R}$ such that

$$\int |f|^p \, d\mu < \infty,$$

where two measurable functions are equivalent if they are equal μ-a.e. The L^p-norm of $f \in L^p(X)$ is defined by

$$\|f\|_{L^p} = \left(\int |f|^p \, d\mu\right)^{1/p}.$$

The notation $L^p(X)$ assumes that the measure μ on X is understood. We say that $f_n \to f$ in L^p if $\|f - f_n\|_{L^p} \to 0$. The reason to regard functions that are equal a.e. as equivalent is so that $\|f\|_{L^p} = 0$ implies that $f = 0$. For example, the characteristic function $\chi_{\mathbb{Q}}$ of the rationals on $\mathbb{R}$ is equivalent to 0 in $L^p(\mathbb{R})$. We will not worry about the distinction between a function and its equivalence class, except when the precise pointwise values of a representative function are significant.

Example 7.2. If $\mathbb{N}$ is equipped with counting measure, then $L^p(\mathbb{N})$ consists of all sequences $\{x_n \in \mathbb{R} : n \in \mathbb{N}\}$ such that

$$\sum_{n=1}^{\infty} |x_n|^p < \infty.$$

We write this sequence space as $\ell^p(\mathbb{N})$, with norm

$$\|\{x_n\}\|_{\ell^p} = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}.$$

The space $L^\infty(X)$ is defined in a slightly different way. First, we introduce the notion of esssential supremum.

Definition 7.3. Let $f : X \to \mathbb{R}$ be a measurable function on a measure space $(X, \mathcal{A}, \mu)$. The essential supremum of f on X is

$$\operatorname*{ess\,sup}_X f = \inf\left\{a \in \mathbb{R} : \mu\{x \in X : f(x) > a\} = 0\right\}.$$

Equivalently,

$$\operatorname*{ess\,sup}_X f = \inf\left\{\sup_X g : g = f \text{ pointwise a.e.}\right\}.$$

Thus, the essential supremum of a function depends only on its μ-a.e. equivalence class. We say that f is essentially bounded on X if

$$\operatorname*{ess\,sup}_X |f| < \infty.$$

Definition 7.4. Let $(X, \mathcal{A}, \mu)$ be a measure space. The space $L^\infty(X)$ consists of pointwise a.e.-equivalence classes of essentially bounded measurable functions $f : X \to \mathbb{R}$ with norm

$$\|f\|_{L^\infty} = \operatorname*{ess\,sup}_X |f|.$$

In future, we will write

$$\operatorname{ess\,sup} f = \sup f.$$

We rarely want to use the supremum instead of the essential supremum when the two have different values, so this notation should not lead to any confusion.

7.2. Minkowski and Hölder inequalities

We state without proof two fundamental inequalities.

Theorem 7.5 (Minkowski inequality)**.** *If $f, g \in L^p(X)$, where $1 \le p \le \infty$, then $f + g \in L^p(X)$ and*

$$\|f + g\|_{L^p} \le \|f\|_{L^p} + \|f\|_{L^p}.$$

This inequality means, as stated previously, that $\|\cdot\|_{L^p}$ is a norm on $L^p(X)$ for $1 \le p \le \infty$. If $0 < p < 1$, then the reverse inequality holds

$$\|f\|_{L^p} + \|g\|_{L^p} \le \|f + g\|_{L^p},$$

so $\|\cdot\|_{L^p}$ is not a norm in that case. Nevertheless, for $0 < p < 1$ we have

$$|f + g|^p \le |f|^p + |g|^p,$$

so $L^p(X)$ is a linear space in that case also.

To state the second inequality, we define the Hölder conjugate of an exponent.

Definition 7.6. Let $1 \le p \le \infty$. The Hölder conjugate p' of p is defined by

$$\frac{1}{p} + \frac{1}{p'} = 1 \qquad \text{if } 1 < p < \infty,$$

and $1' = \infty$, $\infty' = 1$.

Note that $1 \le p' \le \infty$, and the Hölder conjugate of p' is p.

Theorem 7.7 (Hölder's inequality)**.** *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and $1 \le p \le \infty$. If $f \in L^p(X)$ and $g \in L^{p'}(X)$, then $fg \in L^1(X)$ and*

$$\int |fg|\, d\mu \le \|f\|_{L^p} \|g\|_{L^{p'}}.$$

For $p = p' = 2$, this is the Cauchy-Schwartz inequality.

7.3. Density

Density theorems enable us to prove properties of L^p functions by proving them for functions in a dense subspace and then extending the result by continuity. For general measure spaces, the simple functions are dense in L^p.

Theorem 7.8. *Suppose that $(X, \mathcal{A}, \nu)$ is a measure space and $1 \le p \le \infty$. Then the simple functions that belong to $L^p(X)$ are dense in $L^p(X)$.*

Proof. It is sufficient to prove that we can approximate a positive function $f : X \to [0, \infty)$ by simple functions, since a general function may be decomposed into its positive and negative parts.

First suppose that $f \in L^p(X)$ where $1 \le p < \infty$. Then, from Theorem 3.12, there is an increasing sequence of simple functions $\{\phi_n\}$ such that $\phi_n \uparrow f$ pointwise. These simple functions belong to L^p, and

$$|f - \phi_n|^p \le |f|^p \in L^1(X).$$

Hence, the dominated convergence theorem implies that

$$\int |f - \phi_n|^p \, d\mu \to 0 \qquad \text{as } n \to \infty,$$

which proves the result in this case.

If $f \in L^\infty(X)$, then we may choose a representative of f that is bounded. According to Theorem 3.12, there is a sequence of simple functions that converges uniformly to f, and therefore in $L^\infty(X)$. □

Note that a simple function

$$\phi = \sum_{i=1}^{n} c_i \chi_{A_i}$$

belongs to L^p for $1 \le p < \infty$ if and only if $\mu(A_i) < \infty$ for every A_i such that $c_i \neq 0$, meaning that its support has finite measure. On the other hand, every simple function belongs to L^∞.

For suitable measures defined on topological spaces, Theorem 7.8 can be used to prove the density of continuous functions in L^p for $1 \le p < \infty$, as in Theorem 4.27 for Lebesgue measure on $\mathbb{R}^n$. We will not consider extensions of that result to more general measures or topological spaces here.

7.4. Completeness

In proving the completeness of $L^p(X)$, we will use the following Lemma.

Lemma 7.9. *Suppose that X is a measure space and $1 \le p < \infty$. If*

$$\{g_k \in L^p(X) : k \in \mathbb{N}\}$$

is a sequence of L^p-functions such that

$$\sum_{k=1}^{\infty} \|g_k\|_{L^p} < \infty,$$

then there exists a function $f \in L^p(X)$ such that

$$\sum_{k=1}^{\infty} g_k = f$$

where the sum converges pointwise a.e. and in L^p.

Proof. Define $h_n, h : X \to [0, \infty]$ by

$$h_n = \sum_{k=1}^{n} |g_k|, \qquad h = \sum_{k=1}^{\infty} |g_k|.$$

Then $\{h_n\}$ is an increasing sequence of functions that converges pointwise to h, so the monotone convergence theorem implies that

$$\int h^p \, d\mu = \lim_{n\to\infty} \int h_n^p \, d\mu.$$

By Minkowski's inequality, we have for each $n \in \mathbb{N}$ that

$$\|h_n\|_{L^p} \le \sum_{k=1}^{n} \|g_k\|_{L^p} \le M$$

where $\sum_{k=1}^{\infty} \|g_k\|_{L^p} = M$. It follows that $h \in L^p(X)$ with $\|h\|_{L^p} \le M$, and in particular that h is finite pointwise a.e. Moreover, the sum $\sum_{k=1}^{\infty} g_k$ is absolutely convergent pointwise a.e., so it converges pointwise a.e. to a function $f \in L^p(X)$ with $|f| \le h$. Since

$$\left| f - \sum_{k=1}^{n} g_k \right|^p \le \left(|f| + \sum_{k=1}^{n} |g_k| \right)^p \le (2h)^p \in L^1(X),$$

the dominated convergence theorem implies that

$$\int \left| f - \sum_{k=1}^{n} g_k \right|^p d\mu \to 0 \qquad \text{as } n \to \infty,$$

meaning that $\sum_{k=1}^{\infty} g_k$ converges to f in L^p. □

The following theorem implies that $L^p(X)$ equipped with the L^p-norm is a Banach space.

Theorem 7.10 (Riesz-Fischer theorem)**.** *If X is a measure space and $1 \le p \le \infty$, then $L^p(X)$ is complete.*

Proof. First, suppose that $1 \le p < \infty$. If $\{f_k : k \in \mathbb{N}\}$ is a Cauchy sequence in $L^p(X)$, then we can choose a subsequence $\{f_{k_j} : j \in \mathbb{N}\}$ such that

$$\left\| f_{k_{j+1}} - f_{k_j} \right\|_{L^p} \le \frac{1}{2^j}.$$

Writing $g_j = f_{k_{j+1}} - f_{k_j}$, we have

$$\sum_{j=1}^{\infty} \|g_j\|_{L^p} < \infty,$$

so by Lemma 7.9, the sum

$$f_{k_1} + \sum_{j=1}^{\infty} g_j$$

converges pointwise a.e. and in L^p to a function $f \in L^p$. Hence, the limit of the subsequence

$$\lim_{j\to\infty} f_{k_j} = \lim_{j\to\infty}\left(f_{k_1} + \sum_{i=1}^{j-1} g_i\right) = f_{k_1} + \sum_{j=1}^{\infty} g_j = f$$

exists in L^p. Since the original sequence is Cauchy, it follows that

$$\lim_{k\to\infty} f_k = f$$

in L^p. Therefore every Cauchy sequence converges, and $L^p(X)$ is complete when $1 \le p < \infty$.

Second, suppose that $p = \infty$. If $\{f_k\}$ is Cauchy in L^∞, then for every $m \in \mathbb{N}$ there exists an integer $n \in \mathbb{N}$ such that we have

$$|f_j(x) - f_k(x)| < \frac{1}{m} \qquad \text{for all } j, k \ge n \text{ and } x \in N^c_{j,k,m} \tag{7.1}$$

where $N_{j,k,m}$ is a null set. Let

$$N = \bigcup_{j,k,m\in\mathbb{N}} N_{j,k,m}.$$

Then N is a null set, and for every $x \in N^c$ the sequence $\{f_k(x) : k \in \mathbb{N}\}$ is Cauchy in $\mathbb{R}$. We define a measurable function $f : X \to \mathbb{R}$, unique up to pointwise a.e. equivalence, by

$$f(x) = \lim_{k\to\infty} f_k(x) \qquad \text{for } x \in N^c.$$

Letting $k \to \infty$ in (7.1), we find that for every $m \in \mathbb{N}$ there exists an integer $n \in \mathbb{N}$ such that

$$|f_j(x) - f(x)| \le \frac{1}{m} \qquad \text{for } j \ge n \text{ and } x \in N^c.$$

It follows that f is essentially bounded and $f_j \to f$ in L^∞ as $j \to \infty$. This proves that L^∞ is complete. □

One useful consequence of this proof is worth stating explicitly.

Corollary 7.11. *Suppose that X is a measure space and $1 \le p < \infty$. If $\{f_k\}$ is a sequence in $L^p(X)$ that converges in L^p to f, then there is a subsequence $\{f_{k_j}\}$ that converges pointwise a.e. to f.*

As Example 4.26 shows, the full sequence need not converge pointwise a.e.

7.5. Duality

The dual space of a Banach space consists of all bounded linear functionals on the space.

Definition 7.12. If X is a real Banach space, the dual space of X^* consists of all bounded linear functionals $F : X \to \mathbb{R}$, with norm

$$\|F\|_{X^*} = \sup_{x\in X\setminus\{0\}} \left[\frac{|F(x)|}{\|x\|_X}\right] < \infty.$$

A linear functional is bounded if and only if it is continuous. For L^p spaces, we will use the Radon-Nikodym theorem to show that $L^p(X)^*$ may be identified with $L^{p'}(X)$ for $1 < p < \infty$. Under a σ-finiteness assumption, it is also true that $L^1(X)^* = L^\infty(X)$, but in general $L^\infty(X)^* \neq L^1(X)$.

Hölder's inequality implies that functions in $L^{p'}$ define bounded linear functionals on L^p with the same norm, as stated in the following proposition.

Proposition 7.13. *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and $1 < p \leq \infty$. If $f \in L^{p'}(X)$, then*

$$F(g) = \int fg \, d\mu$$

defines a bounded linear functional $F : L^p(X) \to \mathbb{R}$, and

$$\|F\|_{L^{p*}} = \|f\|_{L^{p'}}.$$

If X is σ-finite, then the same result holds for $p = 1$.

Proof. From Hölder's inequality, we have for $1 \leq p \leq \infty$ that

$$|F(g)| \leq \|f\|_{L^{p'}} \|g\|_{L^p},$$

which implies that F is a bounded linear functional on L^p with

$$\|F\|_{L^{p*}} \leq \|f\|_{L^{p'}}.$$

In proving the reverse inequality, we may assume that $f \neq 0$ (otherwise the result is trivial).

First, suppose that $1 < p < \infty$. Let

$$g = (\operatorname{sgn} f) \left(\frac{|f|}{\|f\|_{L^{p'}}} \right)^{p'/p}.$$

Then $g \in L^p$, since $f \in L^{p'}$, and $\|g\|_{L^p} = 1$. Also, since $p'/p = p' - 1$,

$$\begin{aligned} F(g) &= \int (\operatorname{sgn} f) f \left(\frac{|f|}{\|f\|_{L^{p'}}} \right)^{p'-1} d\mu \\ &= \|f\|_{L^{p'}}. \end{aligned}$$

Since $\|g\|_{L^p} = 1$, we have $\|F\|_{L^{p*}} \geq |F(g)|$, so that

$$\|F\|_{L^{p*}} \geq \|f\|_{L^{p'}}.$$

If $p = \infty$, we get the same conclusion by taking $g = \operatorname{sgn} f \in L^\infty$. Thus, in these cases the supremum defining $\|F\|_{L^{p*}}$ is actually attained for a suitable function g.

Second, suppose that $p = 1$ and X is σ-finite. For $\epsilon > 0$, let

$$A = \{x \in X : |f(x)| > \|f\|_{L^\infty} - \epsilon\}.$$

Then $0 < \mu(A) \leq \infty$. Moreover, since X is σ-finite, there is an increasing sequence of sets A_n of finite measure whose union is A such that $\mu(A_n) \to \mu(A)$, so we can find a subset $B \subset A$ such that $0 < \mu(B) < \infty$. Let

$$g = (\operatorname{sgn} f) \frac{\chi_B}{\mu(B)}.$$

Then $g \in L^1(X)$ with $\|g\|_{L^1} = 1$, and

$$F(g) = \frac{1}{\mu(B)} \int_B |f| \, d\mu \geq \|f\|_{L^\infty} - \epsilon.$$

It follows that

$$\|F\|_{L^{1*}} \geq \|f\|_{L^\infty} - \epsilon,$$

and therefore $\|F\|_{L^{1*}} \geq \|f\|_{L^\infty}$ since $\epsilon > 0$ is arbitrary. □

This proposition shows that the map $F = J(f)$ defined by

$$J : L^{p'}(X) \to L^p(X)^*, \qquad J(f) : g \mapsto \int fg\, d\mu, \tag{7.2}$$

is an isometry from $L^{p'}$ into L^{p*}. The main part of the following result is that J is onto when $1 < p < \infty$, meaning that every bounded linear functional on L^p arises in this way from an $L^{p'}$-function.

The proof is based on the idea that if $F : L^p(X) \to \mathbb{R}$ is a bounded linear functional on $L^p(X)$, then $\nu(E) = F(\chi_E)$ defines an absolutely continuous measure on $(X, \mathcal{A}, \mu)$, and its Radon-Nikodym derivative $f = d\nu/d\mu$ is the element of $L^{p'}$ corresponding to F.

Theorem 7.14 (Dual space of L^p)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $1 < p < \infty$, then (7.2) defines an isometric isomorphism of $L^{p'}(X)$ onto the dual space of $L^p(X)$.*

Proof. We just have to show that the map J defined in (7.2) is onto, meaning that every $F \in L^p(X)^*$ is given by $J(f)$ for some $f \in L^{p'}(X)$.

First, suppose that X has finite measure, and let

$$F : L^p(X) \to \mathbb{R}$$

be a bounded linear functional on $L^p(X)$. If $A \in \mathcal{A}$, then $\chi_A \in L^p(X)$, since X has finite measure, and we may define $\nu : \mathcal{A} \to \mathbb{R}$ by

$$\nu(A) = F(\chi_A).$$

If $A = \bigcup_{i=1}^\infty A_i$ is a disjoint union of measurable sets, then

$$\chi_A = \sum_{i=1}^\infty \chi_{A_i},$$

and the dominated convergence theorem implies that

$$\left\| \chi_A - \sum_{i=1}^n \chi_{A_i} \right\|_{L^p} \to 0$$

as $n \to \infty$. Hence, since F is a continuous linear functional on L^p,

$$\nu(A) = F(\chi_A) = F\left(\sum_{i=1}^\infty \chi_{A_i}\right) = \sum_{i=1}^\infty F(\chi_{A_i}) = \sum_{i=1}^\infty \nu(A_i),$$

meaning that ν is a signed measure on $(X, \mathcal{A})$.

If $\mu(A) = 0$, then χ_A is equivalent to 0 in L^p and therefore $\nu(A) = 0$ by the linearity of F. Thus, ν is absolutely continuous with respect to μ. By the Radon-Nikodym theorem, there is a function $f : X \to \mathbb{R}$ such that $d\nu = f d\mu$ and

$$F(\chi_A) = \int f\chi_A\, d\mu \qquad \text{for every} A \in \mathcal{A}.$$

Hence, by the linearity and boundedness of F,

$$F(\phi) = \int f\phi \, d\mu$$

for all simple functions ϕ, and

$$\left| \int f\phi \, d\mu \right| \leq M\|\phi\|_{L^p}$$

where $M = \|F\|_{L^{p*}}$.

Taking $\phi = \operatorname{sgn} f$, which is a simple function, we see that $f \in L^1(X)$. We may then extend the integral of f against bounded functions by continuity. Explicitly, if $g \in L^\infty(X)$, then from Theorem 7.8 there is a sequence of simple functions $\{\phi_n\}$ with $|\phi_n| \leq |g|$ such that $\phi_n \to g$ in L^∞, and therefore also in L^p. Since

$$|f\phi_n| \leq \|g\|_{L^\infty}|f| \in L^1(X),$$

the dominated convergence theorem and the continuity of F imply that

$$F(g) = \lim_{n\to\infty} F(\phi_n) = \lim_{n\to\infty} \int f\phi_n \, d\mu = \int fg \, d\mu,$$

and that

$$\left| \int fg \, d\mu \right| \leq M\|g\|_{L^p} \qquad \text{for every } g \in L^\infty(X). \tag{7.3}$$

Next we prove that $f \in L^{p'}(X)$. We will estimate the $L^{p'}$ norm of f by a similar argument to the one used in the proof of Proposition 7.13. However, we need to apply the argument to a suitable approximation of f, since we do not know *a priori* that $f \in L^{p'}$.

Let $\{\phi_n\}$ be a sequence of simple functions such that

$$\phi_n \to f \qquad \text{pointwise a.e. as } n \to \infty$$

and $|\phi_n| \leq |f|$. Define

$$g_n = (\operatorname{sgn} f) \left(\frac{|\phi_n|}{\|\phi_n\|_{L^{p'}}} \right)^{p'/p}.$$

Then $g_n \in L^\infty(X)$ and $\|g_n\|_{L^p} = 1$. Moreover, $fg_n = |fg_n|$ and

$$\int |\phi_n g_n| \, d\mu = \|\phi_n\|_{L^{p'}}.$$

It follows from these equalities, Fatou's lemma, the inequality $|\phi_n| \leq |f|$, and (7.3) that

$$\begin{aligned}
\|f\|_{L^{p'}} &\leq \liminf_{n\to\infty} \|\phi_n\|_{L^{p'}} \\
&\leq \liminf_{n\to\infty} \int |\phi_n g_n| \, d\mu \\
&\leq \liminf_{n\to\infty} \int |fg_n| \, d\mu \\
&\leq M.
\end{aligned}$$

Thus, $f \in L^{p'}$. Since the simple functions are dense in L^p and $g \mapsto \int fg\, d\mu$ is a continuous functional on L^p when $f \in L^{p'}$, it follows that $F(g) = \int fg\, d\mu$ for every $g \in L^p(X)$. Proposition 7.13 then implies that

$$\|F\|_{L^{p*}} = \|f\|_{L^{p'}},$$

which proves the result when X has finite measure.

The extension to non-finite measure spaces is straightforward, and we only outline the proof. If X is σ-finite, then there is an increasing sequence $\{A_n\}$ of sets with finite measure whose union is X. By the previous result, there is a unique function $f_n \in L^{p'}(A_n)$ such that

$$F(g) = \int_{A_n} f_n g\, d\mu \qquad \text{for all } g \in L^p(A_n).$$

If $m \geq n$, the functions f_m, f_n are equal pointwise a.e. on A_n, and the dominated convergence theorem implies that $f = \lim_{n\to\infty} f_n \in L^{p'}(X)$ is the required function.

Finally, if X is not σ-finite, then for each σ-finite subset $A \subset X$, let $f_A \in L^{p'}(A)$ be the function such that $F(g) = \int_A f_A g\, d\mu$ for every $g \in L^p(A)$. Define

$$M' = \sup\left\{\|f_A\|_{L^{p'}(A)} : A \subset X \text{ is } \sigma\text{-finite}\right\} \leq \|F\|_{L^p(X)^*},$$

and choose an increasing sequence of sets A_n such that

$$\|f_{A_n}\|_{L^{p'}(A_n)} \to M' \qquad \text{as } n \to \infty.$$

Defining $B = \bigcup_{n=1}^{\infty} A_n$, one may verify that f_B is the required function. □

A Banach space X is reflexive if its bi-dual X^{**} is equal to the original space X under the natural identification

$$\iota : X \to X^{**} \qquad \text{where } \iota(x)(F) = F(x) \text{ for every } F \in X^*,$$

meaning that x acting on F is equal to F acting on x. Reflexive Banach spaces are generally better-behaved than non-reflexive ones, and an immediate corollary of Theorem 7.14 is the following.

Corollary 7.15. *If X is a measure space and $1 < p < \infty$, then $L^p(X)$ is reflexive.*

Theorem 7.14 also holds if $p = 1$ provided that X is σ-finite, but we omit a detailed proof. On the other hand, the theorem does not hold if $p = \infty$. Thus L^1 and L^∞ are not reflexive Banach spaces, except in trivial cases.

The following example illustrates a bounded linear functional on an L^∞-space that does not arise from an element of L^1.

Example 7.16. Consider the sequence space $\ell^\infty(\mathbb{N})$. For

$$x = \{x_i : i \in \mathbb{N}\} \in \ell^\infty(\mathbb{N}), \qquad \|x\|_{\ell^\infty} = \sup_{i\in\mathbb{N}} |x_i| < \infty,$$

define $F_n \in \ell^\infty(\mathbb{N})^*$ by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

meaning that F_n maps a sequence to the mean of its first n terms. Then

$$\|F_n\|_{\ell^{\infty*}} = 1$$

for every $n \in \mathbb{N}$, so by the Alaoglu theorem on the weak-$*$ compactness of the unit ball, there exists a subsequence $\{F_{n_j} : j \in \mathbb{N}\}$ and an element $F \in \ell^\infty(\mathbb{N})^*$ with $\|F\|_{\ell^{\infty *}} \le 1$ such that $F_{n_j} \overset{*}{\rightharpoonup} F$ in the weak-$*$ topology on $\ell^{\infty *}$.

If $u \in \ell^\infty$ is the unit sequence with $u_i = 1$ for every $i \in \mathbb{N}$, then $F_n(u) = 1$ for every $n \in \mathbb{N}$, and hence

$$F(u) = \lim_{j\to\infty} F_{n_j}(u) = 1,$$

so $F \neq 0$; in fact, $\|F\|_{\ell^\infty} = 1$. Now suppose that there were $y = \{y_i\} \in \ell^1(\mathbb{N})$ such that

$$F(x) = \sum_{i=1}^{\infty} x_i y_i \qquad \text{for every } x \in \ell^\infty.$$

Then, denoting by $e_k \in \ell^\infty$ the sequence with kth component equal to 1 and all other components equal to 0, we have

$$y_k = F(e_k) = \lim_{j\to\infty} F_{n_j}(e_k) = \lim_{j\to\infty} \frac{1}{n_j} = 0$$

so $y = 0$, which is a contradiction. Thus, $\ell^\infty(\mathbb{N})^*$ is strictly larger than $\ell^1(\mathbb{N})$.

We remark that if a sequence $x = \{x_i\} \in \ell^\infty$ has a limit $L = \lim_{i\to\infty} x_i$, then $F(x) = L$, so F defines a generalized limit of arbitrary bounded sequences in terms of their Cesàro sums. Such bounded linear functionals on $\ell^\infty(\mathbb{N})$ are called Banach limits.

It is possible to characterize the dual of $L^\infty(X)$ as a space $\mathrm{ba}(X)$ of bounded, finitely additive, signed measures that are absolutely continuous with respect to the measure μ on X. This result is rarely useful, however, since finitely additive measures are not easy to work with. Thus, for example, instead of using the weak topology on $L^\infty(X)$, we can regard $L^\infty(X)$ as the dual space of $L^1(X)$ and use the corresponding weak-$*$ topology.

Bibliography

[1] V. I. Bogachev, *Measure Theory*, Vol I. and II., Springer-Verlag, Heidelberg, 2007.
[2] D. L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.
[3] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, 1992.
[4] G. B. Folland, *Real Analysis*, 2nd ed., Wiley, New York, 1999.
[5] F. Jones, *Lebesgue Integration on Euclidean Space*, Revised Ed., Jones and Bartlett, Sudberry, 2001.
[6] S. Lang, *Real and Functional Analysis*, Springer-Verlag, 1993.
[7] E. H. Lieb and M. Loss, *Analysis*, AMS 1997.
[8] E. M. Stein and R. Shakarchi, *Real Analysis*, Princeton University Press, 2005.
[9] M. E. Taylor, *Measure Theory and Integration*, American Mathematical Society, Providence, 2006.
[10] S. Wagon, *The Banach-Tarski Paradox*, Encyclopedia of Mathematics and its Applications, Vol. 24 Cambridge University Press, 1985.
[11] R. L. Wheeler and Z. Zygmund, *Measure and Integral*, Marcel Dekker, 1977.