# 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: $\infty$ 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 $\infty$ if the subset contains $\infty$ 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 $\infty$ 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 $\mu^*$ 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 $\mu^*$ is not assumed to be additive even if the collection $\{E_i\}$ is disjoint.

## 1.5. $\sigma$-algebras

A $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebras on $X$; they are the smallest and largest $\sigma$-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 $\sigma$-algebra $\mathcal{A}$ on $X$.

It is useful to compare the definition of a $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebra that contains $\mathcal{F}$, and an intersection of $\sigma$-algebras is a $\sigma$-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 $\sigma$-algebra on $X$, then $\sigma(\mathcal{F}) \subset \mathcal{A}$.*

Among the most important $\sigma$-algebras are the Borel $\sigma$-algebras on topological spaces.

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

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

is the $\sigma$-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 $\sigma$-algebra.

**Definition 1.9.** A measure $\mu$ 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 $\mu$ on a set $X$ is finite if $\mu(X) < \infty$, and $\sigma$-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 $\sigma$-algebra $\mathcal{A}$ on $X$, and a measure $\mu$ defined on $\mathcal{A}$. When $\mathcal{A}$ and $\mu$ 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 $\mu$ 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 $\nu$ is a measure, called counting measure on $X$. Every subset of $X$ is measurable with respect to $\nu$. Counting measure is finite if $X$ is finite and $\sigma$-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 $\mu$

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

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 $\mu$-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 $\mu$, not just the $\sigma$-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 $\mu$ to $\overline{\mathcal{A}}$.*

PROOF. The collection $\overline{\mathcal{A}}$ is a $\sigma$-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 $\mu$ to a measure on $\overline{\mathcal{A}}$, and that $(X, \overline{\mathcal{A}}, \overline{\mu})$ is complete. $\square$

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 $\mu$ 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 $\sigma$-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 $\sigma$-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 $\infty$. 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

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

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 $\epsilon$-trick' used in the example appears in various forms throughout measure theory.

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

**Theorem 2.4.** *Lebesgue outer measure $\mu^*$ has the following properties.*

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

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

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

The main property to prove is the countable subadditivity of $\mu^*$. 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

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

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 $R^\circ$, 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 $\{I_{i,j} \subset \mathbb{R} : j = 1, \ldots, N_i\}$,*

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

*Define the rectangles*

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

*Then*

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

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

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

□

**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*

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

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

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

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

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

where the $c_{i,j}$ are obtained by ordering the left and right $i$th 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 $\{R_1, R_2, \dots, R_N\}$ covers $R$, then there is a almost disjoint, finite collection of rectangles $\{S_1, S_2, \dots, S_M\}$ 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 $S_i^\circ$ 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 $\{S_1^\circ, S_2^\circ, \dots, S_N^\circ\}$. Then $\{S_1, S_2, \dots, S_N\}$ 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 $\mu^*$ be an outer measure on a set $X$. A subset $A \subset X$ is Carathéodory measurable with respect to $\mu^*$, 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 $\mu^*$ 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 $\mu^*$ is a $\sigma$-algebra, and the restriction of $\mu^*$ 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 $\sigma$-algebra, we only need to show that it is closed under countable unions. We will prove at the same time that $\mu^*$ is countably additive on measurable sets; since $\mu^*(\varnothing) = 0$, this will prove that the restriction of $\mu^*$ 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 $\mu^*$ 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 $\mu^*$ 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 $j \geq 2$),

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

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 $\mu^*$, 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 $\mu^*$ is countably additive on the $\sigma$-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 $\sigma$-algebra of Lebesgue measurable sets, the restriction of Lebesgue outer measure $\mu^*$ 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 $\mu$ 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 $\sigma$-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 $R^\circ$ 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, $R^\circ$ 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 $\mu^*$ 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 $\infty$, 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 $\infty$, 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 $\sigma$-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 $\sigma$-algebra $\mathcal{B}(\mathbb{R}^n)$ on $\mathbb{R}^n$ is the $\sigma$-algebra generated by the open sets, $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{T}(\mathbb{R}^n))$. A set that belongs to the Borel $\sigma$-algebra is called a Borel set.

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

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 $\sigma$-algebra, it follows that $\sigma(\mathcal{R}) \subset \mathcal{L}$, so $\mathcal{B} \subset \mathcal{L}$. $\square$

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 $\sigma$-algebra $\mathcal{B}$ is not complete and is strictly smaller than the Lebesgue $\sigma$-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 $\sigma$-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 $\mu^*(A) \leq \mu(G)$, so

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

and we just need to prove the reverse inequality,

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

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

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

Let $S_i$ be an rectangle whose interior $S_i^\circ$ contains $R_i$ such that

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

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 $G$,

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

and therefore

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

It follows that

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

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

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

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

Therefore, we just need to prove the reverse inequality,

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

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

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

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

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

It follows that

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

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 $\mu^*$ 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 $\mu^*$ 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. $\square$

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 $F_\sigma(\mathbb{R}^n)$.

$G_\delta$ and $F_\sigma$ 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 $G_\delta$ and $F_\sigma$ sets with the required properties. $\square$

In particular, since any measurable set can be approximated up to a set of measure zero by a $G_\delta$ or an $F_\sigma$, 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 $\sigma$-algebra is the completion of the Borel $\sigma$-algebra with respect to Lebesgue measure.

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

## 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 $\mu^*$ 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*

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

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

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

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 $N \in \mathbb{N}$

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

which implies that

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

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

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

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

**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

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

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

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

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

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

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 $\mu^*$ 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 $\Lambda E$ is Lebesgue measurable.*

Proof. The diagonal map $\Lambda$ does not change the orientation of a rectangle, so it maps a cover of $E$ by parallel rectangles to a cover of $\Lambda E$ by parallel rectangles, and conversely. Moreover, $\Lambda$ 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 $\Lambda E$ 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

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

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

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

or

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

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*

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

*for every $a < b$.*

The construction of $\mu_F$ 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 $\mu_F^*$ 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 $\mu_F$ 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 $\mu_F$ is the $\delta$-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 $\mu_F$ assigns measure one to the Cantor set, which has Lebesgue measure zero, and measure zero to its complement. Despite the fact that $\mu_F$ is supported on a set of Lebesgue measure zero, the $\mu_F$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebras in the previous proposition; this is required because we typically cannot use an explicit representation of sets in a $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebra, although if completeness is not required we may use the Borel $\sigma$-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 $\sigma$-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 $\sigma$-algebra generated by intervals of the form $(-\infty, b)$ is contained in the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$ since the intervals are open sets. Conversely, the $\sigma$-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 $\sigma$-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:*

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

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 $\sigma$-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 $\infty$ 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

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

so the supremum and infimum are measurable Moreover, since

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

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 $\chi_E$ 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 $\phi$ 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 $\mu$-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 $\chi_E$ are equal almost everywhere, but 0 is measurable and $\chi_E$ 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 $\mu$-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 $\phi$ with respect to $\mu$ is

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

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 $\infty$ is equal to 0. The integral may take the value $\infty$ (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:*

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

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

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

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:*

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

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 $\mu$ 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 $\phi$ that

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

This proves the theorem. $\square$

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 $\infty$, 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:*

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

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 $k \geq 0$, 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 $f \leq g$, then

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

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

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 $\mu$, 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 $\mu$-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 $f = 0$ $\mu$-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 $\mu$ 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. $\square$

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 $\mu$. 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

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

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*

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

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

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

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 $g_n \leq f_k$ for every $k \geq n$, we have

$$\int g_n \, d\mu \leq \inf_{k\geq n} \int f_k \, d\mu,$$

so that

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

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 $\infty$ 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 $\sigma$-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 $\theta$, 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

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

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 $\chi_A$ 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 $0 \le g \le 1$. 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

$$\|\chi_A - g\|_{L^1} = \int_{G \setminus K} |\chi_A - g| \, dx \le \mu(G \setminus K) < \epsilon,$$

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 $\sigma$-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 $\sigma$-algebras

We begin by describing product $\sigma$-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 $\sigma$-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 $\sigma$-algebra $\mathcal{A} \otimes \mathcal{B}$ is the $\sigma$-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 $\sigma$-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. $\square$

As an example, we consider the product of Borel $\sigma$-algebras on $\mathbb{R}^n$.

**Proposition 5.3.** *Suppose that $\mathbb{R}^m$, $\mathbb{R}^n$ are equipped with their Borel $\sigma$-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 $\sigma$-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 $\sigma$-algebra, since $\mathcal{B}(\mathbb{R}^{m+n})$ is a $\sigma$-algebra and

$$A^c \times \mathbb{R}^n = (A \times \mathbb{R}^n)^c, \qquad \left(\bigcup_{i=1}^{\infty} A_i\right) \times \mathbb{R}^n = \bigcup_{i=1}^{\infty} (A_i \times \mathbb{R}^n).$$

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

$$\mathcal{B}(\mathbb{R}^{m+n}) \supset \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

$\square$

By the repeated application of this result, we see that the Borel $\sigma$-algebra on $\mathbb{R}^n$ is the $n$-fold product of the Borel $\sigma$-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 $\sigma$-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 $\sigma$-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 $\lambda$ 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 $\lambda$ 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 $\lambda^*$ is an outer measure. Moreover, every set belonging to $\mathcal{E}$ is $\lambda^*$-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 $\lambda^*$ 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 $\lambda$ 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 $\lambda$ a premeasure on an algebra $\mathcal{E} \subset \mathcal{P}(X)$. Then $\lambda$ is $\sigma$-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 $\sigma$-finite premeasure on an algebra $\mathcal{E}$ and $\mathcal{A}$ is the $\sigma$-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 $\sigma$-algebra, but obviously it generates the same product $\sigma$-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 $\lambda$ 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 $\lambda$ 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 $\lambda$ 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 $\lambda$ is a premeasure as claimed. The outer product measure associated with $\lambda$, 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

$$(\mu\otimes\nu)^*(E)=\inf\left\{\sum_{i=1}^{\infty}\mu(A_i)\nu(B_i) : E\subset \textstyle\bigcup_{i=1}^{\infty}(A_i\times B_i) \text{ where } A_i\in\mathcal{A},\ B_i\in\mathcal{B}\right\}.$$

The product measure

$$(\mu\otimes\nu):\mathcal{A}\otimes\mathcal{B}\to[0,\infty],\qquad (\mu\otimes\nu)=\left.(\mu\otimes\nu)^*\right|_{\mathcal{A}\otimes\mathcal{B}}$$

is the restriction of the product outer measure to the product $\sigma$-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 $\sigma$-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 $\sigma$-algebra of Carathéodory measurable sets associated with $(\mu\otimes\nu)^*$ is strictly larger than the product $\sigma$-algebra. For example, if $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with Lebesgue measure defined on their Borel $\sigma$-algebras, then the Carathéodory $\sigma$-algebra on the product $\mathbb{R}^{m+n}=\mathbb{R}^m\times\mathbb{R}^n$ is the Lebesgue $\sigma$-algebra $\mathcal{L}(\mathbb{R}^{m+n})$, whereas the product $\sigma$-algebra is the Borel $\sigma$-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 $\sigma$-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 $\sigma$-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

$$E_1 \subset E_2 \subset \cdots \subset E_i \subset \ldots, \qquad F_1 \supset F_2 \supset \cdots \supset F_i \supset \ldots,$$

then

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{C}, \qquad \bigcap_{i=1}^{\infty} F_i \in \mathcal{C}.$$

Obviously, every $\sigma$-algebra is a monotone class, but not conversely. As with $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebra on $\mathbb{R}^{m+n}$ is the completion with respect to Lebesgue measure of the product of the Lebesgue $\sigma$-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 $L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

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 $\mu$ 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 $\mu$-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 $\mu$, $\nu$ 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 $\nu$ with respect to $\mu$. 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 $\nu$ with respect to $\mu$, subject to a $\sigma$-finiteness assumption: namely, that $\nu$ is absolutely continuous with respect to $\mu$.

## 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

$$\mu(K) \leq \frac{C}{t} \int_{\mathbb{R}^n} |f(y)|\, dy.$$

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 $\{B_1, B_2, \ldots, B_N\}$ from the open cover $\{B_x : x \in K\}$. By Lemma 6.1, there is a finite subfamily of disjoint balls $\{B_1', B_2', \ldots, B_M'\}$ such that

$$\begin{aligned}
\mu(K) &\leq \sum_{i=1}^{N} |B_i| \\
&\leq 3^n \sum_{j=1}^{M} |B_j'| \\
&\leq \frac{3^n}{t} \sum_{j=1}^{M} \int_{B_j'} |f|\, dx \\
&\leq \frac{3^n}{t} \int |f|\, dx,
\end{aligned}$$

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

$$\left|\frac{1}{|B_r(x)|} \int_{B_r(x)} f(y) - f(x)\, dy\right| \leq \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy,$$

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 $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ except on $\bigcup_{k=1}^\infty E_k$, which has measure zero.

Next, observe that since

$$|f(y) + g(y) - [f(x) + g(x)]| \le |f(y) - f(x)| + |g(y) - g(x)|$$

and $\limsup(A + B) \le \limsup A + \limsup B$, we have

$$(f + g)^* \le f^* + g^*.$$

Thus, if $f \in L^1(\mathbb{R}^n)$ and $g \in C_c(\mathbb{R}^n)$, we have

$$(f - g)^* \le f^* + g^* = f^*,$$
$$f^* = (f - g + g)^* \le (f - g)^* + g^* = (f - g)^*,$$

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 $0 < t < \infty$

$$\mu\left(\{x \in \mathbb{R}^n : f^*(x) > t\}\right) \le \frac{C}{t}\, \|f\|_{L^1}. \tag{6.7}$$

To show this, we estimate

$$\begin{aligned} f^*(x) &\le \sup_{r>0}\left[\frac{1}{|B_r(x)|}\int_{B_r(x)} |f(y) - f(x)|\, dy\right] \\ &\le \sup_{r>0}\left[\frac{1}{|B_r(x)|}\int_{B_r(x)} |f(y)|\, dy\right] + |f(x)| \\ &\le Mf(x) + |f(x)|. \end{aligned}$$

It follows that

$$\{f^* > t\} \subset \{Mf + |f| > t\} \subset \{Mf > t/2\} \cup \{|f| > t/2\}.$$

By the Hardy-Littlewood theorem,

$$\mu\left(\{x \in \mathbb{R}^n : Mf(x) > t/2\}\right) \le \frac{2 \cdot 3^n}{t}\|f\|_{L^1},$$

and by the Chebyshev inequality

$$\mu\left(\{x \in \mathbb{R}^n : |f(x)| > t/2\}\right) \le \frac{2}{t}\|f\|_{L^1}.$$

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

$$\begin{aligned}\mu\left(\{x \in \mathbb{R}^n : f^*(x) > t\}\right) &= \mu\left(\{x \in \mathbb{R}^n : (f-g)^*(x) > t\}\right)\\ &\leq \frac{C}{t}\|f-g\|_{L^1}\\ &\leq \frac{C\epsilon}{t}.\end{aligned}$$

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 $\nu$ on $X$ is a function $\nu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

(a) $\nu(\varnothing) = 0$;
(b) $\nu$ attains at most one of the values $\infty$, $-\infty$;
(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 $\mu$ 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 $\nu$ with respect to $\mu$. 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 $\nu$ of the form (6.8) must be absolutely continuous with respect to the measure $\mu$.

## 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 $\nu$ is a signed measure on a measurable space $X$. A set $A \subset X$ is positive for $\nu$ if it is measurable and $\nu(B) \geq 0$ for every measurable subset $B \subset A$. Similarly, $A$ is negative for $\nu$ if it is measurable and $\nu(B) \leq 0$ for every measurable subset $B \subset A$, and null for $\nu$ 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 $\nu$, nor does $\nu(A) = 0$ imply that $A$ is null for $\nu$. Nevertheless, as we show in the next result, if $\nu(A) > 0$, then $A$ contains a subset that is positive for $\nu$. 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 $\nu$ 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 $\nu$ 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 $\infty$. (Otherwise, consider $-\nu$.) Then $\nu(B) \neq \infty$; and if $B \subset A$, then the additivity of $\nu$ 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 $\delta_1$ 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 $\delta_i$ 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 $\nu$ is a signed measure on a measurable space $(X, \mathcal{A})$, then there is a positive set $P$ and a negative set $N$ for $\nu$ such that $P \cup N = X$ and $P \cap N = \varnothing$. These sets are unique up to $\nu$-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 $\nu$, 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 $\nu$. 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 $\nu$ and therefore null, and similarly for $P' \setminus P$. Thus, the decomposition is unique up to $\nu$-null sets. □

To describe the corresponding decomposition of the signed measure $\nu$ 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 $\mu$, $\nu$ 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 $\delta$-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 $\nu$ 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*

$$\nu = \nu^+ - \nu^- \qquad \textit{and } \nu^+ \perp \nu^-.$$

Proof. Let $X = P \cup N$ where $P$, $N$ are positive, negative sets for $\nu$. 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 $\nu$-null set, so the decomposition is unique. □

We call $\nu^+$ and $\nu^-$ the positive and negative parts of $\nu$, respectively. The total variation $|\nu|$ of $\nu$ is the measure

$$|\nu| = \nu^+ + \nu^-.$$

We say that the signed measure $\nu$ is $\sigma$-finite if $|\nu|$ is $\sigma$-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 $\nu$ singular with respect to a measure $\mu$, then it is supported on different sets from $\mu$, while if $\nu$ is absolutely continuous with respect to $\mu$, then it supported on on the same sets as $\mu$.

**Definition 6.22.** Let $\nu$ be a signed measure and $\mu$ a measure on a measurable space $(X, \mathcal{A})$. Then $\nu$ is absolutely continuous with respect to $\mu$, 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 $\mu$-null set is a $\nu$-null set. Unlike singularity, absolute continuity is not symmetric.

**Example 6.23.** If $\mu$ is Lebesgue measure and $\nu$ 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 $\mu$ 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 $\nu$ is absolutely continuous with respect to $\mu$.

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 $\nu$ is a finite signed measure and $\mu$ 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 $\nu$ is not absolutely continuous with respect to $\mu$. □

The Radon-Nikodym theorem provides a converse to Example 6.24 for absolutely continuous, $\sigma$-finite measures. As part of the proof, from [**4**], we also show that any signed measure $\nu$ can be decomposed into an absolutely continuous and singular part with respect to a measure $\mu$ (the Lebesgue decomposition of $\nu$). In the proof of the theorem, we will use the following lemma.

**Lemma 6.26.** *Suppose that $\mu$, $\nu$ 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 $\nu$ be a $\sigma$-finite signed measure and $\mu$ a $\sigma$-finite measure on a measurable space $(X, \mathcal{A})$. Then there exist unique $\sigma$-finite signed measures $\nu_a$, $\nu_s$ 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 $\mu$-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 $\nu$ 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 $\mu$, $\nu$ 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 $\nu$.

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

$$\int_A \max\{g, h\}\, d\mu = \int_B g\, d\mu + \int_C h\, d\mu \le \nu(B) + \nu(C) = \nu(A).$$

Let

$$m = \sup\left\{\int_X g\, d\mu : g \in \mathcal{F}\right\} \le \nu(X).$$

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 $\max\{g_1, g_2, \ldots, g_n\}$, 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

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A g_n\, d\mu \le \nu(A),$$

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 $\nu_s$ 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 $\nu_s \ge \epsilon\mu$ on $P$. It follows that for any $A \in \mathcal{A}$

$$\begin{aligned}
\nu(A) &= \int_A f\, d\mu + \nu_s(A) \\
&\ge \int_A f\, d\mu + \nu_s(A \cap P) \\
&\ge \int_A f\, d\mu + \epsilon\mu(A \cap P) \\
&\ge \int_A (f + \epsilon\chi_P)\, d\mu.
\end{aligned}$$

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 $\mu$ which implies that it is zero. Moreover, $f$ is determined uniquely by $\nu_a$ up to pointwise a.e. equivalence.

Finally, if $\mu$, $\nu$ are $\sigma$-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 $\nu$, 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 $\nu$ with respect to $\mu$, and denote it by

$$f = \frac{d\nu}{d\mu}.$$

Some hypothesis of $\sigma$-finiteness is essential in the theorem, as the following example shows.

**Example 6.28.** Let $\mathcal{B}$ be the Borel $\sigma$-algebra on $[0,1]$, $\mu$ Lebesgue measure, and $\nu$ counting measure on $\mathcal{B}$. Then $\mu$ is finite and $\mu \ll \nu$, but $\nu$ is not $\sigma$-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 $\sigma$-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 $\nu$ 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 $\nu$ be a complex measure and $\mu$ a $\sigma$-finite measure on a measurable space $(X, \mathcal{A})$. Then there exist unique complex measures $\nu_a$, $\nu_s$ 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* $\mu$*-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 $p$th 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 $\mu$-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 $\mu$ 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

$$\operatorname*{ess\,sup}_X f = \inf \{a \in \mathbb{R} : \mu\{x \in X : f(x) > a\} = 0\}.$$

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 $\mu$-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 $1 \leq p \leq \infty$. 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 $1 \leq p < \infty$. 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

$$|f - \phi_n|^p \leq |f|^p \in L^1(X).$$

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 $1 \leq p < \infty$ 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 $L^\infty$.

For suitable measures defined on topological spaces, Theorem 7.8 can be used to prove the density of continuous functions in $L^p$ for $1 \leq p < \infty$, 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 $1 \leq p < \infty$. 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 $L^\infty$, 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 $L^\infty$ as $j \to \infty$. This proves that $L^\infty$ 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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-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. $\square$

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 $\nu$ is a signed measure on $(X, \mathcal{A})$.

If $\mu(A) = 0$, then $\chi_A$ is equivalent to 0 in $L^p$ and therefore $\nu(A) = 0$ by the linearity of $F$. Thus, $\nu$ is absolutely continuous with respect to $\mu$. 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 $\phi$, 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 $L^\infty$, 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 $\sigma$-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 $\sigma$-finite, then for each $\sigma$-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 $\sigma$-finite, but we omit a detailed proof. On the other hand, the theorem does not hold if $p = \infty$. Thus $L^1$ and $L^\infty$ are not reflexive Banach spaces, except in trivial cases.

The following example illustrates a bounded linear functional on an $L^\infty$-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 $k$th 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 $\mu$ 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.

# LECTURE NOTES

# IN MEASURE THEORY

Christer Borell
Matematik
Chalmers och Göteborgs universitet
412 96 Göteborg
(Version: January 12)

# PREFACE

These are lecture notes on integration theory for a eight-week course at the Chalmers University of Technology and the Göteborg University. The parts defining the course essentially lead to the same results as the first three chapters in the Folland book [$F$] , which is used as a text book on the course. The proofs in the lecture notes sometimes differ from those given in [$F$] . Here is a brief description of the differences to simplify for the reader.

In Chapter 1 we introduce so called $\pi$-systems and $\sigma$-additive classes, which are substitutes for monotone classes of sets [$F$]. Besides we prefer to emphasize metric outer measures instead of so called premeasures. Throughout the course, a variety of important measures are obtained as image measures of the linear measure on the real line. In Section 1.6 positive measures in $\mathbf{R}$ induced by increasing right continuous mappings are constructed in this way.

Chapter 2 deals with integration and is very similar to [$F$] and most other texts.

Chapter 3 starts with some standard facts about metric spaces and relates the concepts to measure theory. For example Ulam's Theorem is included. The existence of product measures is based on properties of $\pi$-systems and $\sigma$-additive classes.

Chapter 4 deals with different modes of convergence and is mostly close to [$F$] . Here we include a section about orthogonality since many students have seen parts of this theory before.

The Lebesgue Decomposition Theorem and Radon-Nikodym Theorem in Chapter 5 are proved using the von Neumann beautiful $L^2$-proof.

To illustrate the power of abstract integration these notes contain several sections, which do not belong to the course but may help the student to a better understanding of measure theory. The corresponding parts are set between the symbols

$$\downarrow\downarrow\downarrow$$

and

$$\uparrow\uparrow\uparrow$$

respectively.

Finally I would like to express my deep gratitude to the students in my classes for suggesting a variety of improvements and a special thank to Jonatan Vasilis who has provided numerous comments and corrections in my original text.

Göteborg 2006
Christer Borell

# CONTENT

# CHAPTER 1

# MEASURES

## Introduction

The Riemann integral, dealt with in calculus courses, is well suited for computations but less suited for dealing with limit processes. In this course we will introduce the so called Lebesgue integral, which keeps the advantages of the Riemann integral and eliminates its drawbacks. At the same time we will develop a general measure theory which serves as the basis of contemporary analysis and probability.

In this introductory chapter we set forth some basic concepts of measure theory, which will open for abstract Lebesgue integration.

## 1.1. $\sigma$-Algebras and Measures

Throughout this course

$\mathbf{N} = \{0, 1, 2, ...\}$ (the set of natural numbers)
$\mathbf{Z} = \{..., -2, -1, 0, 1, , 2, ...\}$ (the set of integers)
$\mathbf{Q}$ = the set of rational numbers
$\mathbf{R}$ = the set of real numbers
$\mathbf{C}$ = the set of complex numbers.

If $A \subseteq \mathbf{R}$, $A_+$ is the set of all strictly positive elements in $A$.

If $f$ is a function from a set $A$ into a set $B$, this means that to every $x \in A$ there corresponds a point $f(x) \in B$ and we write $f : A \to B$. A function is often called a map or a mapping. The function $f$ is injective if

$$(x \neq y) \Rightarrow (f(x) \neq f(y))$$

and surjective if to each $y \in B$, there exists an $x \in A$ such that $f(x) = y$. An injective and surjective function is said to be bijective.

A set $A$ is finite if either $A$ is empty or there exist an $n \in \mathbf{N}_+$ and a bijection $f : \{1, ..., n\} \to A$. The empty set is denoted by $\phi$. A set $A$ is said to be denumerable if there exists a bijection $f : \mathbf{N}_+ \to A$. A subset of a denumerable set is said to be at most denumerable.

Let $X$ be a set. For any $A \subseteq X$, the indicator function $\chi_A$ of $A$ relative to $X$ is defined by the equation

$$\chi_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \in A^c. \end{cases}$$

The indicator function $\chi_A$ is sometimes written $1_A$. We have the following relations:

$$\chi_{A^c} = 1 - \chi_A$$

$$\chi_{A \cap B} = \min(\chi_A, \chi_B) = \chi_A \chi_B$$

and

$$\chi_{A \cup B} = \max(\chi_A, \chi_B) = \chi_A + \chi_B - \chi_A \chi_{B.}$$

**Definition 1.1.1.** Let $X$ be a set.

a) A collection $\mathcal{A}$ of subsets of $X$ is said to be an algebra in $X$ if $\mathcal{A}$ has the following properties:

(i) $X \in \mathcal{A}$.
(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, where $A^c$ is the complement of $A$ relative to $X$.
(iii) If $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$.

(b) A collection $\mathcal{M}$ of subsets of $X$ is said to be a $\sigma$-algebra in $X$ if $\mathcal{M}$ is an algebra with the following property:

If $A_n \in \mathcal{M}$ for all $n \in \mathbf{N}_+$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{M}$.

If $\mathcal{M}$ is a $\sigma$-algebra in $X$, $(X, \mathcal{M})$ is called a measurable space and the members of $\mathcal{M}$ are called measurable sets. The so called power set $\mathcal{P}(X)$, that is the collection of all subsets of $X$, is a $\sigma$-algebra in $X$. It is simple to prove that the intersection of any family of $\sigma$-algebras in $X$ is a $\sigma$-algebra. It follows that if $\mathcal{E}$ is any subset of $\mathcal{P}(X)$, there is a unique smallest $\sigma$-algebra $\sigma(\mathcal{E})$ containing $\mathcal{E}$, namely the intersection of all $\sigma$-algebras containing $\mathcal{E}$.

The $\sigma$-algebra $\sigma(\mathcal{E})$ is called the $\sigma$-algebra generated by $\mathcal{E}$. The $\sigma$-algebra generated by all open intervals in $\mathbf{R}$ is denoted by $\mathcal{R}$. It is readily seen that the $\sigma$-algebra $\mathcal{R}$ contains every subinterval of $\mathbf{R}$. Before we proceed, recall that a subset $E$ of $\mathbf{R}$ is open if to each $x \in E$ there exists an open subinterval of $\mathbf{R}$ contained in $E$ and containing $x$; the complement of an open set is said to be closed. We claim that $\mathcal{R}$ contains every open subset $U$ of $\mathbf{R}$. To see this suppose $x \in U$ and let $x \in \, ]a, b[ \, \subseteq U$, where $-\infty < a < b < \infty$. Now pick $r, s \in \mathbf{Q}$ such that $a < r < x < s < b$. Then $x \in \, ]r, s[ \, \subseteq U$ and it follows that $U$ is the union of all bounded open intervals with rational boundary points contained in $U$. Since this family of intervals is at most denumberable we conclude that $U \in \mathcal{R}$. In addition, any closed set belongs to $\mathcal{R}$ since its complements is open. It is by no means simple to grasp the definition of $\mathcal{R}$ at this stage but the reader will successively see that the $\sigma$-algebra $\mathcal{R}$ has very nice properties. At the very end of Section 1.3, using the so called Axiom of Choice, we will exemplify a subset of the real line which does not belong to $\mathcal{R}$. In fact, an example of this type can be constructed without the Axiom of Choice (see Dudley's book $[D]$).

In measure theory, inevitably one encounters $\infty$. For example the real line has infinite length. Below $[0, \infty] = [0, \infty[ \cup \{\infty\}$. The inequalities $x \leq y$ and $x < y$ have their usual meanings if $x, y \in [0, \infty[$. Furthermore, $x \leq \infty$ if $x \in [0, \infty]$ and $x < \infty$ if $x \in [0, \infty[$. We define $x + \infty = \infty + x = \infty$ if $x, y \in [0, \infty]$, and

$$x \cdot \infty = \infty \cdot x = \begin{cases} 0 & \text{if } x = 0 \\ \infty & \text{if } 0 < x \leq \infty. \end{cases}$$

Sums and multiplications of real numbers are defined in the usual way.

If $A_n \subseteq X$, $n \in \mathbf{N}_+$, and $A_k \cap A_n = \phi$ if $k \neq n$, the sequence $(A_n)_{n \in \mathbf{N}_+}$ is called a disjoint denumerable collection. If $(X, \mathcal{M})$ is a measurable space, the collection is called a denumerable measurable partition of $A$ if $A = \cup_{n=1}^{\infty} A_n$ and $A_n \in \mathcal{M}$ for every $n \in \mathbf{N}_+$. Some authors call a denumerable collection of sets a countable collection of sets.

**Definition 1.1.2.** (a) Let $\mathcal{A}$ be an algebra of subsets of $X$. A function $\mu : \mathcal{A} \to [0, \infty]$ is called a content if

(i) $\mu(\phi) = 0$
(ii) $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A, B \in \mathcal{A}$ and $A \cap B = \phi$.

(b) If $(X, \mathcal{M})$ is a measurable space a content $\mu$ defined on the $\sigma$-algebra $\mathcal{M}$ is called a positive measure if it has the following property:

For any disjoint denumerable collection $(A_n)_{n \in \mathbf{N}_+}$ of members of $\mathcal{M}$

$$\mu(\cup_{n=1}^{\infty} A_n) = \Sigma_{n=1}^{\infty} \mu(A_n).$$

If $(X, \mathcal{M})$ is a measurable space and the function $\mu : \mathcal{M} \to [0, \infty]$ is a positive measure, $(X, \mathcal{M}, \mu)$ is called a positive measure space. The quantity $\mu(A)$ is called the $\mu$-measure of $A$ or simply the measure of $A$ if there is no ambiguity. Here $(X, \mathcal{M}, \mu)$ is called a probability space if $\mu(X) = 1$, a finite positive measure space if $\mu(X) < \infty$, and a $\sigma$-finite positive measure space if $X$ is a denumerable union of measurable sets with finite $\mu$-measure. The measure $\mu$ is called a probability measure, finite measure, and $\sigma$-finite measure, if $(X, \mathcal{M}, \mu)$ is a probability space, a finite positive measure space, and a $\sigma$-finite positive measure space, respectively. A probability space is often denoted by $(\Omega, \mathcal{F}, P)$. A member $A$ of $\mathcal{F}$ is called an event.

As soon as we have a positive measure space $(X, \mathcal{M}, \mu)$, it turns out to be a fairly simple task to define a so called $\mu$-integral

$$\int_X f(x) d\mu(x)$$

as will be seen in Chapter 2.

The class of all finite unions of subintervals of $\mathbf{R}$ is an algebra which is denoted by $\mathcal{R}_0$. If $A \in \mathcal{R}_0$ we denote by $l(A)$ the Riemann integral

$$\int_{-\infty}^{\infty} \chi_A(x)dx$$

and it follows from courses in calculus that the function $l : \mathcal{R}_0 \to [0, \infty]$ is a content. The algebra $\mathcal{R}_0$ is called the Riemann algebra and $l$ the Riemann content. If $I$ is a subinterval of $\mathbf{R}$, $l(I)$ is called the length of $I$. Below we follow the convention that the empty set is an interval.

If $A \in \mathcal{P}(X)$, $c_X(A)$ equals the number of elements in $A$, when $A$ is a finite set, and $c_X(A) = \infty$ otherwise. Clearly, $c_X$ is a positive measure. The measure $c_X$ is called the counting measure on $X$.

Given $a \in X$, the probability measure $\delta_a$ defined by the equation $\delta_a(A) = \chi_A(a)$, if $A \in \mathcal{P}(X)$, is called the Dirac measure at the point $a$. Sometimes we write $\delta_a = \delta_{X,a}$ to emphasize the set $X$.

If $\mu$ and $\nu$ are positive measures defined on the same $\sigma$-algebra $\mathcal{M}$, the sum $\mu + \nu$ is a positive measure on $\mathcal{M}$. More generally, $\alpha\mu + \beta\nu$ is a positive measure for all real $\alpha$, $\beta \geq 0$. Furthermore, if $E \in \mathcal{M}$, the function $\lambda(A) = \mu(A \cap E)$, $A \in \mathcal{M}$, is a positive measure. Below this measure $\lambda$ will be denoted by $\mu^E$ and we say that $\mu^E$ is concentrated on $E$. If $E \in \mathcal{M}$, the class $\mathcal{M}_E = \{A \in \mathcal{M};\ A \subseteq E\}$ is a $\sigma$-algebra of subsets of $E$ and the function $\theta(A) = \mu(A)$, $A \in \mathcal{M}_E$, is a positive measure. Below this measure $\theta$ will be denoted by $\mu_{|E}$ and is called the restriction of $\mu$ to $\mathcal{M}_E$.

Let $I_1, ..., I_n$ be subintervals of the real line. The set

$$I_1 \times ... \times I_n = \{(x_1, ..., x_n) \in \mathbf{R}^n;\ x_k \in I_k,\ k = 1, ..., n\}$$

is called an $n$-cell in $\mathbf{R}^n$; its volume $\text{vol}(I_1 \times ... \times I_n)$ is, by definition, equal to

$$\text{vol}(I_1 \times ... \times I_n) = \Pi_{k=1}^n l(I_k).$$

If $I_1, ..., I_n$ are open subintervals of the real line, the $n$-cell $I_1 \times ... \times I_n$ is called an open $n$-cell. The $\sigma$-algebra generated by all open $n$-cells in $\mathbf{R}^n$ is denoted by $\mathcal{R}_n$. In particular, $\mathcal{R}_1 = \mathcal{R}$. A basic theorem in measure theory states that there exists a unique positive measure $v_n$ defined on $\mathcal{R}_n$ such that the measure of any $n$-cell is equal to its volume. The measure $v_n$ is called the volume measure on $\mathcal{R}_n$ or the volume measure on $\mathbf{R}^n$. Clearly, $v_n$ is $\sigma$-finite. The measure $v_2$ is called the area measure on $\mathbf{R}^2$ and $v_1$ the linear measure on $\mathbf{R}$.

**Theorem 1.1.1.** *The volume measure on* $\mathbf{R}^n$ *exists.*

Theorem 1.1.1 will be proved in Section 1.5 in the special case $n = 1$. The general case then follows from the existence of product measures in Section 3.4. An alternative proof of Theorem 1.1.1 will be given in Section 3.2. As soon as the existence of volume measure is established a variety of interesting measures can be introduced.

Next we prove some results of general interest for positive measures.

**Theorem 1.1.2.** *Let* $\mathcal{A}$ *be an algebra of subsets of* $X$ *and* $\mu$ *a content defined on* $\mathcal{A}$. *Then,*

(a) $\mu$ *is finitely additive, that is*

$$\mu(A_1 \cup ... \cup A_n) = \mu(A_1) + ... + \mu(A_n)$$

*if* $A_1, ..., A_n$ *are pairwise disjoint members of* $\mathcal{A}$.

(b) *if* $A, B \in \mathcal{A}$,

$$\mu(A) = \mu(A \setminus B) + \mu(A \cap B).$$

*Moreover, if* $\mu(A \cap B) < \infty$, *then*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

(c) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ *if* $A, B \in \mathcal{A}$.

(d) $\mu$ *finitely sub-additive, that is*

$$\mu(A_1 \cup ... \cup A_n) \leq \mu(A_1) + ... + \mu(A_n)$$

*if* $A_1, ..., A_n$ *are members of* $\mathcal{A}$.

*If* $(X, \mathcal{M}, \mu)$ *is a positive measure space*

(e) $\mu(A_n) \to \mu(A)$ *if* $A = \cup_{n\in\mathbf{N}_+} A_n,\ A_n \in \mathcal{M},$ *and*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq ... \ .$$

(f) $\mu(A_n) \to \mu(A)$ *if* $A = \cap_{n\in\mathbf{N}_+} A_n,\ A_n \in \mathcal{M},$

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq ...$$

*and* $\mu(A_1) < \infty$.
(g) $\mu$ *is sub-additive, that is for any denumerable collection* $(A_n)_{n\in\mathbf{N}_+}$ *of members of* $\mathcal{M}$,

$$\mu(\cup_{n=1}^{\infty} A_n) \leq \Sigma_{n=1}^{\infty} \mu(A_n).$$

PROOF (a) If $A_1, ..., A_n$ are pairwise disjoint members of $\mathcal{A}$,

$$\mu(\cup_{k=1}^{n} A_k) = \mu(A_1 \cup (\cup_{k=2}^{n} A_k))$$

$$= \mu(A_1) + \mu(\cup_{k=2}^{n} A_k)$$

and, by induction, we conclude that $\mu$ is finitely additive.

(b) Recall that

$$A \setminus B = A \cap B^c.$$

Now $A = (A \setminus B) \cup (A \cap B)$ and we get

$$\mu(A) = \mu(A \setminus B) + \mu(A \cap B).$$

Moreover, since $A \cup B = (A \setminus B) \cup B$,

$$\mu(A \cup B) = \mu(A \setminus B) + \mu(B)$$

and, if $\mu(A \cap B) < \infty$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

(c) Part (b) yields $\mu(B) = \mu(B \setminus A) + \mu(A \cap B) = \mu(B \setminus A) + \mu(A)$, where the last member does not fall below $\mu(A)$.

(d) If $(A_i)_{i=1}^n$is a sequence of members of $\mathcal{A}$ define the so called disjunction $(B_k)_{k=1}^n$ of the sequence $(A_i)_{i=1}^n$ as

$$B_1 = A_1 \text{ and } B_k = A_k \setminus \cup_{i=1}^{k-1} A_i \text{ for } 2 \leq k \leq n.$$

Then $B_k \subseteq A_k$, $\cup_{i=1}^k A_i = \cup_{i=1}^k B_i$, $k = 1, .., n$, and $B_i \cap B_j = \phi$ if $i \neq j$. Hence, by Parts (a) and (c),

$$\mu(\cup_{k=1}^n A_k) = \Sigma_{k=1}^n \mu(B_k) \leq \Sigma_{k=1}^n \mu(A_k).$$

(e) Set $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then $A_n = B_1 \cup .... \cup B_n$, $B_i \cap B_j = \phi$ if $i \neq j$ and $A = \cup_{k=1}^\infty B_k$. Hence

$$\mu(A_n) = \Sigma_{k=1}^n \mu(B_k)$$

and

$$\mu(A) = \Sigma_{k=1}^\infty \mu(B_k).$$

Now e) follows, by the definition of the sum of an infinite series.

(f) Put $C_n = A_1 \setminus A_n$, $n \geq 1$. Then $C_1 \subseteq C_2 \subseteq C_3 \subseteq ...$,

$$A_1 \setminus A = \cup_{n=1}^\infty C_n$$

and $\mu(A) \leq \mu(A_n) \leq \mu(A_1) < \infty$. Thus

$$\mu(C_n) = \mu(A_1) - \mu(A_n)$$

and Part (e) shows that

$$\mu(A_1) - \mu(A) = \mu(A_1 \setminus A) = \lim_{n \to \infty} \mu(C_n) = \mu(A_1) - \lim_{n \to \infty} \mu(A_n).$$

This proves (f).

(g) The result follows from Parts d) and e).
This completes the proof of Theorem 1.1.2.

The hypothesis "$\mu(A_1) < \infty$ " in Theorem 1.1.2 ( f) is not superfluous. If $c_{\mathbf{N}_+}$ is the counting measure on $\mathbf{N}_+$ and $A_n = \{n, n+1, ...\}$ , then $c_{\mathbf{N}_+}(A_n) = \infty$ for all $n$ but $A_1 \supseteq A_2 \supseteq ....$ and $c_{\mathbf{N}_+}(\cap_{n=1}^{\infty} A_n) = 0$ since $\cap_{n=1}^{\infty} A_n = \phi$.

If $A, B \subseteq X$, the symmetric difference $A\Delta B$ is defined by the equation

$$A\Delta B =_{def} (A \setminus B) \cup (B \setminus A).$$

Note that

$$\chi_{A\Delta B} = \mid \chi_A - \chi_B \mid .$$

Moreover, we have

$$A\Delta B = A^c \Delta B^c$$

and

$$(\cup_{i=1}^{\infty} A_i)\Delta(\cup_{i=1}^{\infty} B_i) \subseteq \cup_{i=1}^{\infty}(A_i \Delta B_i).$$

**Example 1.1.1.** Let $\mu$ be a finite positive measure on $\mathcal{R}$. We claim that to each set $E \in \mathcal{R}$ and $\varepsilon > 0$, there exists a set $A$, which is finite union of intervals (that is, $A$ belongs to the Riemann algebra $\mathcal{R}_0$), such that

$$\mu(E\Delta A) < \varepsilon.$$

To see this let $\mathcal{S}$ be the class of all sets $E \in \mathcal{R}$ for which the conclusion is true. Clearly $\phi \in \mathcal{S}$ and, moreover, $\mathcal{R}_0 \subseteq \mathcal{S}$. If $A \in \mathcal{R}_0$, $A^c \in \mathcal{R}_0$ and therefore $E^c \in \mathcal{S}$ if $E \in \mathcal{S}$. Now suppose $E_i \in \mathcal{S}$, $i \in \mathbf{N}_+$. Then to each $\varepsilon > 0$ and $i$ there is a set $A_i \in \mathcal{R}_0$ such that $\mu(E_i \Delta A_i) < 2^{-i}\varepsilon$. If we set

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

then

$$\mu(E\Delta(\cup_{i=1}^{\infty} A_i)) \leq \Sigma_{i=1}^{\infty} \mu(E_i \Delta A_i) < \varepsilon.$$

Here

$$E\Delta(\cup_{i=1}^{\infty} A_i) = \{E \cap (\cap_{i=1}^{\infty} A_i^c)\} \cup \{E^c \cap (\cup_{i=1}^{\infty} A_i)\}$$

and Theorem 1.1.2 (f) gives that

$$\mu(\{E \cap (\cap_{i=1}^{n} A_i^c)\} \cup \{(E^c \cap (\cup_{i=1}^{\infty} A_i)\}) < \varepsilon$$

if $n$ is large enough (hint: $\cap_{i\in I}(D_i \cup F) = (\cap_{i\in I} D_i) \cup F$). But then

$$\mu(E\Delta \cup_{i=1}^{n} A_i) = \mu(\{E \cap (\cap_{i=1}^{n} A_i^c)\} \cup \{E^c \cap (\cup_{i=1}^{n} A_i)\}) < \varepsilon$$

if $n$ is large enough we conclude that the set $E \in \mathcal{S}$. Thus $\mathcal{S}$ is a $\sigma$-algebra and since $\mathcal{R}_0 \subseteq \mathcal{S} \subseteq \mathcal{R}$ it follows that $\mathcal{S} = \mathcal{R}$.

**Exercises**

1. Prove that the sets $\mathbf{N} \times \mathbf{N} = \{(i,j);\ i,j \in \mathbf{N}\}$ and $\mathbf{Q}$ are denumerable.

2. Suppose $\mathcal{A}$ is an algebra of subsets of $X$ and $\mu$ and $\nu$ two contents on $\mathcal{A}$ such that $\mu \leq \nu$ and $\mu(X) = \nu(X) < \infty$. Prove that $\mu = \nu$.

3. Suppose $\mathcal{A}$ is an algebra of subsets of $X$ and $\mu$ a content on $\mathcal{A}$ with $\mu(X) < \infty$. Show that

$$\mu(A \cup B \cup C) = \mu(A) + \mu(B) + \mu(C)$$

$$-\mu(A \cap B) - \mu(A \cap C) - \mu(B \cap C) + \mu(A \cap B \cap C).$$

4. (a) A collection $\mathcal{C}$ of subsets of $X$ is an algebra with the following property:
If $A_n \in \mathcal{C}$, $n \in \mathbf{N}_+$ and $A_k \cap A_n = \phi$ if $k \neq n$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{C}$.
Prove that $\mathcal{C}$ is a $\sigma$-algebra.

(b) A collection $\mathcal{C}$ of subsets of $X$ is an algebra with the following property:
If $E_n \in \mathcal{C}$ and $E_n \subseteq E_{n+1}$, $n \in \mathbf{N}_+$, then $\cup_1^{\infty} E_n \in \mathcal{C}$ .
Prove that $\mathcal{C}$ is a $\sigma$-algebra.

5. Let $(X, \mathcal{M})$ be a measurable space and $(\mu_k)_{k=1}^{\infty}$ a sequence of positive measures on $\mathcal{M}$ such that $\mu_1 \leq \mu_2 \leq \mu_3 \leq \ldots$ . Prove that the set function

$$\mu(A) = \lim_{k \to \infty} \mu_k(A),\ A \in \mathcal{M}$$

is a positive measure.

6. Let $(X, \mathcal{M}, \mu)$ be a positive measure space. Show that

$$\mu(\cap_{k=1}^{n} A_k) \leq \sqrt[n]{\Pi_{k=1}^{n} \mu(A_k)}$$

for all $A_1, ..., A_n \in \mathcal{M}$.

7. Let $(X, \mathcal{M}, \mu)$ be a $\sigma$-finite positive measure space with $\mu(X) = \infty$. Show that for any $r \in [0, \infty[$ there is some $A \in \mathcal{M}$ with $r < \mu(A) < \infty$.

8. Show that the symmetric difference of sets is associative:

$$A\Delta(B\Delta C) = (A\Delta B)\Delta C.$$

9. $(X, \mathcal{M}, \mu)$ is a finite positive measure space. Prove that

$$\mid \mu(A) - \mu(B) \mid \leq \mu(A\Delta B).$$

10. Let $E = 2\mathbf{N}$. Prove that

$$c_{\mathbf{N}}(E\Delta A) = \infty$$

if $A$ is a finite union of intervals.

11. Suppose $(X, \mathcal{P}(X), \mu)$ is a finite positive measure space such that $\mu(\{x\}) > 0$ for every $x \in X$. Set

$$d(A, B) = \mu(A\Delta B),\ \ A, B \in \mathcal{P}(X).$$

Prove that

$$d(A, B) = 0 \ \Leftrightarrow A = B,$$

$$d(A, B) = d(B, A)$$

and

$$d(A, B) \leq d(A, C) + d(C, B).$$

12. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space. Prove that

$$\mu(\cup_{i=1}^n A_i) \geq \Sigma_{i=1}^n \mu(A_i) - \Sigma_{1 \leq i < j \leq n} \mu(A_i \cap A_j)$$

for all $A_1, ..., A_n \in \mathcal{M}$ and integers $n \geq 2$.

13. Let $(X, \mathcal{M}, \mu)$ be a probability space and suppose the sets $A_1, ..., A_n \in \mathcal{M}$ satisfy the inequality $\sum_1^n \mu(A_i) > n - 1$. Show that $\mu(\cap_1^n A_i) > 0$.

## 1.2. Measure Determining Classes

Suppose $\mu$ and $\nu$ are probability measures defined on the same $\sigma$-algebra $\mathcal{M}$, which is generated by a class $\mathcal{E}$. If $\mu$ and $\nu$ agree on $\mathcal{E}$, is it then true that $\mu$ and $\nu$ agree on $\mathcal{M}$? The answer is in general no. To show this, let

$$X = \{1, 2, 3, 4\}$$

and

$$\mathcal{E} = \{\{1, 2\}, \{1, 3\}\}.$$

Then $\sigma(\mathcal{E}) = \mathcal{P}(X)$. If $\mu = \frac{1}{4} c_X$ and

$$\nu = \frac{1}{6}\delta_{X,1} + \frac{1}{3}\delta_{X,2} + \frac{1}{3}\delta_{X,3} + \frac{1}{6}\delta_{X,4}$$

then $\mu = \nu$ on $\mathcal{E}$ and $\mu \neq \nu$.

In this section we will prove a basic result on measure determining classes for $\sigma$-finite measures. In this context we will introduce so called $\pi$-systems and $\sigma$-additive classes, which will also be of great value later in connection with the construction of so called product measures in Chapter 3.

**Definition 1.2.1.** A class $\mathcal{G}$ of subsets of $X$ is a $\pi$-system if $A \cap B \in \mathcal{G}$ for all $A, B \in \mathcal{G}$.

The class of all open $n$-cells in $\mathbf{R}^n$ is a $\pi$-system.

**Definition 1.2.2.** A class $\mathcal{D}$ of subsets of $X$ is called a $\sigma$-additive class if the following properties hold:

(a) $X \in \mathcal{D}$.
(b) If $A, B \in \mathcal{D}$ and $A \subseteq B$, then $B \setminus A \in \mathcal{D}$.
(c) If $(A_n)_{n \in \mathbf{N}_+}$ is a disjoint denumerable collection of members of the class $\mathcal{D}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.

**Theorem 1.2.1.** *If a $\sigma$-additive class $\mathcal{M}$ is a $\pi$-system, then $\mathcal{M}$ is a $\sigma$-algebra.*

PROOF. If $A \in \mathcal{M}$, then $A^c = X \setminus A \in \mathcal{M}$ since $X \in \mathcal{M}$ and $\mathcal{M}$ is a $\sigma$-additive class. Moreover, if $(A_n)_{n \in \mathbf{N}_+}$ is a denumerable collection of members of $\mathcal{M}$,

$$A_1 \cup ... \cup A_n = (A_1^c \cap ... \cap A_n^c)^c \in \mathcal{M}$$

for each $n$, since $\mathcal{M}$ is a $\sigma$-additive class and a $\pi$-system. Let $(B_n)_{n=1}^{\infty}$ be the disjunction of $(A_n)_{n=1}^{\infty}$. Then $(B_n)_{n \in \mathbf{N}_+}$ is a disjoint denumerable collection of members of $\mathcal{M}$ and Definition 1.2.2(c) implies that $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n \in \mathcal{M}$.

**Theorem 1.2.2.** *Let $\mathcal{G}$ be a $\pi$-system and $\mathcal{D}$ a $\sigma$-additive class such that $\mathcal{G} \subseteq \mathcal{D}$. Then $\sigma(\mathcal{G}) \subseteq \mathcal{D}$.*

PROOF. Let $\mathcal{M}$ be the intersection of all $\sigma$-additive classes containing $\mathcal{G}$. The class $\mathcal{M}$ is a $\sigma$-additive class and $\mathcal{G} \subseteq \mathcal{M} \subseteq \mathcal{D}$. In view of Theorem 1.2.1 $\mathcal{M}$ is a $\sigma$-algebra, if $\mathcal{M}$ is a $\pi$-system and in that case $\sigma(\mathcal{G}) \subseteq \mathcal{M}$. Thus the theorem follows if we show that $\mathcal{M}$ is a $\pi$-system.

Given $C \subseteq X$, denote by $\mathcal{D}_C$ be the class of all $D \subseteq X$ such that $D \cap C \in \mathcal{M}$.

CLAIM 1. If $C \in \mathcal{M}$, then $\mathcal{D}_C$ is a $\sigma$-additive class.

PROOF OF CLAIM 1. First $X \in \mathcal{D}_C$ since $X \cap C = C \in \mathcal{M}$. Moreover, if $A, B \in \mathcal{D}_C$ and $A \subseteq B$, then $A \cap C, B \cap C \in \mathcal{M}$ and

$$(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in \mathcal{M}.$$

Accordingly from this, $B \setminus A \in \mathcal{D}_C$. Finally, if $(A_n)_{n \in \mathbf{N}_+}$ is a disjoint denumerable collection of members of $\mathcal{D}_C$, then $(A_n \cap C)_{n \in \mathbf{N}_+}$ is disjoint denumerable collection of members of $\mathcal{M}$ and

$$(\cup_{n \in \mathbf{N}_+} A_n) \cap C = \cup_{n \in \mathbf{N}_+} (A_n \cap C) \in \mathcal{M}.$$

Thus $\cup_{n \in \mathbf{N}_+} A_n \in \mathcal{D}_C$.

CLAIM 2. If $A \in \mathcal{G}$, then $\mathcal{M} \subseteq \mathcal{D}_A$.

PROOF OF CLAIM 2. If $B \in \mathcal{G}$, $A \cap B \in \mathcal{G} \subseteq \mathcal{M}$. Thus $B \in \mathcal{D}_A$. We have proved that $\mathcal{G} \subseteq \mathcal{D}_A$ and remembering that $\mathcal{M}$ is the intersection of all $\sigma$-additive classes containing $\mathcal{G}$ Claim 2 follows since $\mathcal{D}_A$ is a $\sigma$-additive class.

To complete the proof of Theorem 1.2.2, observe that $B \in \mathcal{D}_A$ if and only if $A \in \mathcal{D}_B$. By Claim 2, if $A \in \mathcal{G}$ and $B \in \mathcal{M}$, then $B \in \mathcal{D}_A$ that is $A \in \mathcal{D}_B$. Thus $\mathcal{G} \subseteq \mathcal{D}_B$ if $B \in \mathcal{M}$. Now the definition of $\mathcal{M}$ implies that $\mathcal{M} \subseteq \mathcal{D}_B$ if $B \in \mathcal{M}$. The proof is almost finished. In fact, if $A, B \in \mathcal{M}$ then $A \in \mathcal{D}_B$ that is $A \cap B \in \mathcal{M}$. Theorem 1.2.2 now follows from Theorem 1.2.1.

**Theorem 1.2.3.** *Let $\mu$ and $\nu$ be positive measures on $\mathcal{M} = \sigma(\mathcal{G})$, where $\mathcal{G}$ is a $\pi$-system, and suppose $\mu(A) = \nu(A)$ for every $A \in \mathcal{G}$.*

(a) *If $\mu$ and $\nu$ are probability measures, then $\mu = \nu$.*

(b) *Suppose there exist $E_n \in \mathcal{G}$, $n \in \mathbf{N}_+$, such that $X = \cup_{n=1}^{\infty} E_n$,*

*$E_1 \subseteq E_2 \subseteq ...$, and*

$$\mu(E_n) = \nu(E_n) < \infty, \text{ all } n \in \mathbf{N}_+.$$

*Then $\mu = \nu$.*

PROOF. (a) Let

$$\mathcal{D} = \{A \in \mathcal{M};\ \ \mu(A) = \ \nu(A)\}.$$

It is immediate that $\mathcal{D}$ is a $\sigma$-additive class and Theorem 1.2.2 implies that $\mathcal{M} = \sigma(\mathcal{G}) \subseteq \mathcal{D}$ since $\mathcal{G} \subseteq \mathcal{D}$ and $\mathcal{G}$ is a $\pi$-system.

(b) If $\mu(E_n) = \nu(E_n) = 0$ for all all $n \in \mathbf{N}_+$, then

$$\mu(X) = \lim_{n\to\infty} \mu(E_n) = 0$$

and, in a similar way, $\nu(X) = 0$. Thus $\mu = \nu$. If $\mu(E_n) = \nu(E_n) > 0$, set

$$\mu_n(A) = \frac{1}{\mu(E_n)}\mu(A \cap E_n) \text{ and } \nu_n(A) = \frac{1}{\nu(E_n)}\nu(A \cap E_n)$$

for each $A \in \mathcal{M}$. By Part (a) $\mu_n = \nu_n$ and we get

$$\mu(A \cap E_n) = \nu(A \cap E_n)$$

for each $A \in \mathcal{M}$. Theorem 1.1.2(e) now proves that $\mu = \nu$.

Theorem 1.2.3 implies that there is at most one positive measure defined on $\mathcal{R}_n$ such that the measure of any open $n$-cell in $\mathbf{R}^n$ equals its volume.

Next suppose $f : X \to Y$ and let $A \subseteq X$ and $B \subseteq Y$. The image of $A$ and the inverse image of $B$ are

$$f(A) = \{y;\ y = f(x) \text{ for some } x \in A\}$$

and

$$f^{-1}(B) = \{x;\ f(x) \in B\}$$

respectively. Note that

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

and

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

Moreover, if $(A_i)_{i\in I}$ is a collection of subsets of $X$ and $(B_i)_{i\in I}$ is a collection of subsets of $Y$

$$f(\cup_{i\in I} A_i) = \cup_{i\in I} f(A_i)$$

and

$$f^{-1}(\cup_{i\in I} B_i) = \cup_{i\in I} f^{-1}(B_i).$$

Given a class $\mathcal{E}$ of subsets of $Y$, set

$$f^{-1}(\mathcal{E}) = \left\{ f^{-1}(B);\ B \in \mathcal{E} \right\}.$$

If $(Y, \mathcal{N})$ is a measurable space, it follows that the class $f^{-1}(\mathcal{N})$ is a $\sigma$-algebra in $X$. If $(X, \mathcal{M})$ is a measurable space

$$\left\{ B \in \mathcal{P}(Y);\ f^{-1}(B) \in \mathcal{M} \right\}$$

is a $\sigma$-algebra in $Y$. Thus, given a class $\mathcal{E}$ of subsets of $Y$,

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

**Definition 1.2.3.** Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. The function $f : X \to Y$ is said to be $(\mathcal{M}, \mathcal{N})$-measurable if $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$. If we say that $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$ is measurable this means that $f : X \to Y$ is an $(\mathcal{M}, \mathcal{N})$-measurable function.

**Theorem 1.2.4.** *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and suppose $\mathcal{E}$ generates $\mathcal{N}$. The function $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable if*

$$f^{-1}(\mathcal{E}) \subseteq \mathcal{M}.$$

PROOF. The assumptions yield

$$\sigma(f^{-1}(\mathcal{E})) \subseteq \mathcal{M}.$$

Since

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

we are done.

**Corollary 1.2.1.** *A function* $f : X \to \mathbf{R}$ *is* $(\mathcal{M}, \mathcal{R})$*-measurable if and only if the set* $f^{-1}(]\alpha, \infty[) \in \mathcal{M}$ *for all* $\alpha \in \mathbf{R}$.

If $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and $\mu$ is a positive measure on $\mathcal{M}$, the equation

$$\nu(B) = \mu(f^{-1}(B)),\ B \in \mathcal{N}$$

defines a positive measure $\nu$ on $\mathcal{N}$. We will write $\nu = \mu f^{-1}$, $\nu = f(\mu)$ or $\nu = \mu_f$. The measure $\nu$ is called the image measure of $\mu$ under $f$ and $f$ is said to transport $\mu$ to $\nu$. Two $(\mathcal{M}, \mathcal{N})$-measurable functions $f : X \to Y$ and $g : X \to Y$ are said to be $\mu$-equimeasurable if $f(\mu) = g(\mu)$.

As an example, let $a \in \mathbf{R}^n$ and define $f(x) = x + a$ if $x \in \mathbf{R}^n$. If $B \subseteq \mathbf{R}^n$,

$$f^{-1}(B) = \{x;\ x + a \in B\} = B - a.$$

Thus $f^{-1}(B)$ is an open $n$-cell if $B$ is, and Theorem 1.2.4 proves that $f$ is $(\mathcal{R}_n, \mathcal{R}_n)$-measurable. Now, granted the existence of volume measure $v_n$, for every $B \in \mathcal{R}_n$ define

$$\mu(B) = f(v_n)(B) = v_n(B - a).$$

Then $\mu(B) = v_n(B)$ if $B$ is an open $n$-cell and Theorem 1.2.3 implies that $\mu = v_n$. We have thus proved the following

**Theorem 1.2.5.** *For any* $A \in \mathcal{R}_n$ *and* $x \in \mathbf{R}^n$

$$A + x \in \mathcal{R}_n$$

*and*

$$v_n(A + x) = v_n(A).$$

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. A measurable function $\xi$ defined on $\Omega$ is called a random variable and the image measure $P_\xi$ is called the probability law of $\xi$. We sometimes write

$$\mathcal{L}(\xi) = P_\xi.$$

Here are two simple examples.

If the range of a random variable $\xi$ consists of $n$ points $S = \{s_1, ..., s_n\}$ $(n \geq 1)$ and $P_\xi = \frac{1}{n} c_S$, $\xi$ is said to have a uniform distribution in $S$. Note that

$$P_\xi = \frac{1}{n} \Sigma_{k=1}^n \delta_{s_k}.$$

Suppose $\lambda > 0$ is a constant. If a random variable $\xi$ has its range in $\mathbf{N}$ and

$$P_\xi = \Sigma_{n=0}^\infty \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$$

then $\xi$ is said to have a Poisson distribution with parameter $\lambda$.

**Exercises**

1. Let $f : X \to Y$, $A \subseteq X$, and $B \subseteq Y$. Show that

$$f(f^{-1}(B)) \subseteq B \text{ and } f^{-1}(f(A)) \supseteq A.$$

2. Let $(X, \mathcal{M})$ be a measurable space and suppose $A \subseteq X$. Show that the function $\chi_A$ is $(\mathcal{M}, \mathcal{R})$-measurable if and only if $A \in \mathcal{M}$.

3. Suppose $(X, \mathcal{M})$ is a measurable space and $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}$, a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions such that

$$\lim_{n\to\infty} f_n(x) \text{ exists and } = f(x) \in \mathbf{R}$$

for each $x \in X$. Prove that $f$ is $(\mathcal{M}, \mathcal{R})$-measurable.

4. Suppose $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$ and $g : (Y, \mathcal{N}) \to (Z, \mathcal{S})$ are measurable. Prove that $g \circ f$ is $(\mathcal{M}, \mathcal{S})$-measurable.

5. Granted the existence of volume measure $v_n$, show that $v_n(rA) = r^n v_n(A)$ if $r \geq 0$ and $A \in \mathcal{R}_n$.

6. Let $\mu$ be the counting measure on $\mathbf{Z}^2$ and $f(x, y) = x$, $(x, y) \in \mathbf{Z}^2$. The positive measure $\mu$ is $\sigma$-finite. Prove that the image measure $f(\mu)$ is not a $\sigma$-finite positive measure.

7. Let $\mu, \nu : \mathcal{R} \to [0, \infty]$ be two positive measures such that $\mu(I) = \nu(I) < \infty$ for each open subinterval of $\mathbf{R}$. Prove that $\mu = \nu$.

8. Let $f : \mathbf{R}^n \to \mathbf{R}^k$ be continuous. Prove that $f$ is $(\mathcal{R}_n, \mathcal{R}_k)$-measurable.

9. Suppose $\xi$ has a Poisson distribution with parameter $\lambda$. Show that $P_\xi\,[2\mathbf{N}] = e^{-\lambda} \cosh \lambda$.

9. Find a $\sigma$-additive class which is not a $\sigma$-algebra.

## 1.3. Lebesgue Measure

Once the problem about the existence of volume measure is solved the existence of the so called Lebesgue measure is simple to establish as will be seen in this section. We start with some concepts of general interest.

If $(X, \mathcal{M}, \mu)$ is a positive measure space, the zero set $\mathcal{Z}_\mu$ of $\mu$ is, by definition, the set at all $A \in \mathcal{M}$ such that $\mu(A) = 0$. An element of $\mathcal{Z}_\mu$ is called a null set or $\mu$-null set. If

$$(A \in \mathcal{Z}_\mu \text{ and } B \subseteq A) \Rightarrow B \in \mathcal{M}$$

the measure space $(X, \mathcal{M}, \mu)$ is said to be complete. In this case the measure $\mu$ is also said to be complete. The positive measure space $(X, \{\phi, X\}, \mu)$, where $X = \{0, 1\}$ and $\mu = 0$, is not complete since $X \in \mathcal{Z}_\mu$ and $\{0\} \notin \{\phi, X\}$.

**Theorem 1.3.1** *If* $(E_n)_{n=1}^\infty$ *is a denumerable collection of members of* $\mathcal{Z}_\mu$ *then* $\cup_{n=1}^\infty E_n \in \mathcal{Z}_\mu$.

PROOF We have

$$0 \leq \mu(\cup_{n=1}^\infty E_n) \leq \Sigma_{n=1}^\infty \mu(E_n) = 0$$

which proves the result.

Granted the existence of linear measure $v_1$ it follows from Theorem 1.3.1 that $\mathbf{Q} \in \mathcal{Z}_{v_1}$ since $\mathbf{Q}$ is countable and $\{a\} \in \mathcal{Z}_{v_1}$ for each real number $a$.

Suppose $(X, \mathcal{M}, \mu)$ is an arbitrary positive measure space. It turns out that $\mu$ is the restriction to $\mathcal{M}$ of a complete measure. To see this suppose $\mathcal{M}^-$ is the class of all $E \subseteq X$ is such that there exist sets $A, B \in \mathcal{M}$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_\mu$. It is obvious that $X \in \mathcal{M}^-$ since $\mathcal{M} \subseteq \mathcal{M}^-$. If $E \in \mathcal{M}^-$, choose $A, B \in \mathcal{M}$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_\mu$. Then $B^c \subseteq E^c \subseteq A^c$ and $A^c \setminus B^c = B \setminus A \in \mathcal{Z}_\mu$ and we conclude that $E^c \in \mathcal{M}^-$. If $(E_i)_{i=1}^\infty$ is a denumerable collection of members of $\mathcal{M}^-$, for each $i$ there exist sets $A_i, B_i \in \mathcal{M}$ such that $A_i \subseteq E \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$. But then

$$\cup_{i=1}^\infty A_i \subseteq \cup_{i=1}^\infty E_i \subseteq \cup_{i=1}^\infty B_i$$

where $\cup_{i=1}^\infty A_i, \cup_{i=1}^\infty B_i \in \mathcal{M}$. Moreover, $(\cup_{i=1}^\infty B_i) \setminus (\cup_{i=1}^\infty A_i) \in \mathcal{Z}_\mu$ since

$$(\cup_{i=1}^\infty B_i) \setminus (\cup_{i=1}^\infty A_i) \subseteq \cup_{i=1}^\infty (B_i \setminus A_i).$$

Thus $\cup_{i=1}^\infty E_i \in \mathcal{M}^-$ and $\mathcal{M}^-$ is a $\sigma$-algebra.

If $E \in \mathcal{M}$, suppose $A_i, B_i \in \mathcal{M}$ are such that $A_i \subseteq E \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$ for $i = 1, 2$. Then for each $i$, $(B_1 \cap B_2) \setminus A_i \in \mathcal{Z}_\mu$ and

$$\mu(B_1 \cap B_2) = \mu((B_1 \cap B_2) \setminus A_i) + \mu(A_i) = \mu(A_i).$$

Thus the real numbers $\mu(A_1)$ and $\mu(A_2)$ are the same and we define $\bar{\mu}(E)$ to be equal to this common number. Note also that $\mu(B_1) = \bar{\mu}(E)$. It is plain that $\bar{\mu}(\phi) = 0$. If $(E_i)_{i=1}^{\infty}$ is a disjoint denumerable collection of members of $\mathcal{M}$, for each $i$ there exist sets $A_i, B_i \in \mathcal{M}$ such that $A_i \subseteq E_i \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$. From the above it follows that

$$\bar{\mu}(\cup_{i=1}^{\infty} E_i) = \mu(\cup_{i=1}^{\infty} A_i) = \Sigma_{n=1}^{\infty} \mu(A_i) = \Sigma_{n=1}^{\infty} \bar{\mu}(E_i).$$

We have proved that $\bar{\mu}$ is a positive measure on $\mathcal{M}^-$. If $E \in \mathcal{Z}_{\bar{\mu}}$ the definition of $\bar{\mu}$ shows that any set $A \subseteq E$ belongs to the $\sigma$-algebra $\mathcal{M}^-$. It follows that the measure $\bar{\mu}$ is complete and its restriction to $\mathcal{M}$ equals $\mu$.

The measure $\bar{\mu}$ is called the completion of $\mu$ and $\mathcal{M}^-$ is called the completion of $\mathcal{M}$ with respect to $\mu$.

**Definition 1.3.1** The completion of volume measure $v_n$ on $\mathbf{R}^n$ is called Lebesgue measure on $\mathbf{R}^n$ and is denoted by $m_n$. The completion of $\mathcal{R}_n$ with respect to $v_n$ is called the Lebesgue $\sigma$-algebra in $\mathbf{R}^n$ and is denoted by $\mathcal{R}_n^-$. A member of the class $\mathcal{R}_n^-$ is called a Lebesgue measurable set in $\mathbf{R}^n$ or a Lebesgue set in $\mathbf{R}^n$. A function $f : \mathbf{R}^n \to \mathbf{R}$ is said to be Lebesgue measurable if it is $(\mathcal{R}_n^-, \mathcal{R})$-measurable. Below, $m_1$ is written $m$ if this notation will not lead to misunderstanding. Furthermore, $\mathcal{R}_1^-$ is written $\mathcal{R}^-$.

**Theorem 1.3.2.** *Suppose $E \in \mathcal{R}_n^-$ and $x \in \mathbf{R}^n$. Then $E + x \in \mathcal{R}_n^-$ and $m_n(E + x) = m_n(E)$.*

PROOF. Choose $A, B \in \mathcal{R}_n$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_{v_n}$. Then, by Theorem 1.2.5, $A + x, B + x \in \mathcal{R}_n$, $v_n(A + x) = v_n(A) = m_n(E)$, and $(B + x) \setminus (A + x) = (B \setminus A) + x \in \mathcal{Z}_{v_n}$. Since $A + x \subseteq E + x \subseteq B + x$ the theorem is proved.

The Lebesgue $\sigma$-algebra in $\mathbf{R}^n$ is very large and contains each set of interest in analysis and probability. In fact, in most cases, the $\sigma$-algebra $\mathcal{R}_n$ is sufficiently large but there are some exceptions. For example, if $f : \mathbf{R}^n \to \mathbf{R}^n$ is continuous and $A \in \mathcal{R}_n$, the image set $f(A)$ need not belong to the class $\mathcal{R}_n$ (see e.g. the Dudley book [D]). To prove the existence of a subset of the real line, which is not Lebesgue measurable we will use the so called Axiom of Choice.

**Axiom of Choice.** If $(A_i)_{i\in I}$ is a non-empty collection of non-empty sets, there exists a function $f : I \to \cup_{i\in I} A_i$ such that $f(i) \in A_i$ for every $i \in I$.

Let $X$ and $Y$ be sets. The set of all ordered pairs $(x, y)$, where $x \in X$ and $y \in Y$ is denoted by $X \times Y$. An arbitrary subset $R$ of $X \times Y$ is called a relation. If $(x, y) \in R$ , we write $x \sim y$. A relation is said to be an equivalence relation on $X$ if $X = Y$ and

(i) $x \sim x$ (reflexivity)
(ii) $x \sim y \Rightarrow y \sim x$ (symmetry)
(iii) $(x \sim y$ and $y \sim z) \Rightarrow x \sim z$ (transitivity)

The equivalence class $R(x) =_{def} \{y;\ y \sim x\}$ . The definition of the equivalence relation $\sim$ implies the following:

(a) $x \in R(x)$
(b) $R(x) \cap R(y) \neq \phi \Rightarrow R(x) = R(y)$
(c) $\cup_{x\in X} R(x) = X$.

An equivalence relation leads to a partition of $X$ into a disjoint collection of subsets of $X$.

Let $X = \left[-\frac{1}{2}, \frac{1}{2}\right]$ and define an equivalence relation for numbers $x, y$ in $X$ by stating that $x \sim y$ if $x - y$ is a rational number. By the Axiom of Choice it is possible to pick exactly one element from each equivalence class. Thus there exists a subset $NL$ of $X$ which contains exactly one element from each equivalence class.

If we assume that $NL \in \mathcal{R}^-$ we get a contradiction as follows. Let $(r_i)_{i=1}^{\infty}$ be an enumeration of the rational numbers in $[-1, 1]$. Then

$$X \subseteq \cup_{i=1}^{\infty}(r_i + NL)$$

and it follows from Theorem 1.3.1 that $r_i + NL \notin \mathcal{Z}_m$ for some $i$. Thus, by Theorem 1.3.2, $NL \notin \mathcal{Z}_m$.

Now assume $(r_i + NL) \cap (r_j + NL) \neq \phi$. Then there exist $a', a'' \in NL$ such that $r_i + a' = r_j + a''$ or $a' - a'' = r_j - r_i$. Hence $a' \sim a''$ and it follows that $a'$ and $a''$ belong to the same equivalence class. But then $a' = a''$. Thus $r_i = r_j$ and we conclude that $(r_i + NL)_{i \in \mathbf{N}_+}$ is a disjoint enumeration of Lebesgue sets. Now, since

$$\cup_{i=1}^{\infty}(r_i + NL) \subseteq \left[-\frac{3}{2}, \frac{3}{2}\right]$$

it follows that

$$3 \geq m(\cup_{i=1}^{\infty}(r_i + NL)) = \Sigma_{n=1}^{\infty} m(NL).$$

But then $NL \in \mathcal{Z}_m$, which is a contradiction. Thus $NL \notin \mathcal{R}^-$.

In the early 1970' Solovay $[S]$ proved that it is consistent with the usual axioms of Set Theory, excluding the Axiom of Choice, that every subset of $\mathbf{R}$ is Lebesgue measurable.

From the above we conclude that the Axiom of Choice implies the existence of a subset of the set of real numbers which does not belong to the class $\mathcal{R}$. Interestingly enough, such an example can be given without any use of the Axiom of Choice and follows naturally from the theory of analytic sets. The interested reader may consult the Dudley book $[D]$.

**Exercises**

1. $(X, \mathcal{M}, \mu)$ is a positive measure space. Prove or disprove: If $A \subseteq E \subseteq B$ and $\mu(A) = \mu(B)$ then $E$ belongs to the domain of the completion $\bar{\mu}$.

2. Prove or disprove: If $A$ and $B$ are not Lebesgue measurable subsets of $\mathbf{R}$, then $A \cup B$ is not Lebesgue measurable.

3. Let $(X, \mathcal{M}, \mu)$ be a complete positive measure space and suppose $A, B \in \mathcal{M}$, where $B \setminus A$ is a $\mu$-null set. Prove that $E \in \mathcal{M}$ if $A \subseteq E \subseteq B$ (stated otherwise $\mathcal{M}^- = \mathcal{M}$).

4. Suppose $E \subseteq \mathbf{R}$ and $E \notin \mathcal{R}^-$. Show there is an $\varepsilon > 0$ such that

$$m(B \setminus A) \geq \varepsilon$$

for all $A, B \in \mathcal{R}^-$ such that $A \subseteq E \subseteq B$.

5. Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space and $(Y, \mathcal{N})$ a measurable space. Furthermore, suppose $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and let $\nu = \mu f^{-1}$, that is $\nu(B) = \mu(f^{-1}(B))$, $B \in \mathcal{N}$. Show that $f$ is $(\mathcal{M}^-, \mathcal{N}^-)$-measurable, where $\mathcal{M}^-$ denotes the completion of $\mathcal{M}$ with respect to $\mu$ and $\mathcal{N}^-$ the completion of $\mathcal{N}$ with respect to $\nu$.

### 1.4. Carathéodory's Theorem

In these notes we exhibit two famous approaches to Lebesgue measure. One is based on the Carathéodory Theorem, which we present in this section, and the other one, due to F. Riesz, is a representation theorem of positive linear functionals on spaces of continuous functions in terms of positive measures. The latter approach, is presented in Chapter 3. Both methods depend on topological concepts such as compactness.

**Definition 1.4.1.** A function $\theta : \mathcal{P}(X) \to [0, \infty]$ is said to be an outer measure if the following properties are satisfied:

(i) $\theta(\phi) = 0$.
(ii) $\theta(A) \leq \theta(B)$ if $A \subseteq B$.
(iii) for any denumerable collection $(A_n)_{n=1}^{\infty}$ of subsets of $X$

$$\theta(\cup_{n=1}^{\infty} A_n) \leq \Sigma_{n=1}^{\infty} \theta(A_n).$$

Since
$$E = (E \cap A) \cup (E \cap A^c)$$
an outer measure $\theta$ satisfies the inequality
$$\theta(E) \leq \theta(E \cap A) + \theta(E \cap A^c).$$

If $\theta$ is an outer measure on $X$ we define $\mathcal{M}(\theta)$ as the set of all $A \subseteq X$ such that
$$\theta(E) = \theta(E \cap A) + \theta(E \cap A^c) \text{ for all } E \subseteq X$$
or, what amounts to the same thing,
$$\theta(E) \geq \theta(E \cap A) + \theta(E \cap A^c) \text{ for all } E \subseteq X.$$

The next theorem is one of the most important in measure theory.

**Theorem 1.4.1. (Carathéodory's Theorem)** *Suppose $\theta$ is an outer measure. The class $\mathcal{M}(\theta)$ is a $\sigma$-algebra and the restriction of $\theta$ to $\mathcal{M}(\theta)$ is a complete measure.*

PROOF. Clearly, $\phi \in \mathcal{M}(\theta)$ and $A^c \in \mathcal{M}(\theta)$ if $A \in \mathcal{M}(\theta)$. Moreover, if $A, B \in \mathcal{M}(\theta)$ and $E \subseteq X$,
$$\theta(E) = \theta(E \cap A) + \theta(E \cap A^c)$$
$$= \theta(E \cap A \cap B) + \theta(E \cap A \cap B^c)$$
$$+\theta(E \cap A^c \cap B) + \theta(E \cap A^c \cap B^c).$$
But
$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$$
and
$$A^c \cap B^c = (A \cup B)^c$$
and we get
$$\theta(E) \geq \theta(E \cap (A \cup B)) + \theta(E \cap (A \cup B)^c).$$
It follows that $A \cup B \in \mathcal{M}(\theta)$ and we have proved that the class $\mathcal{M}(\theta)$ is an algebra. Now if $A, B \in \mathcal{M}(\theta)$ are disjoint
$$\theta(A \cup B) = \theta((A \cup B) \cap A) + \theta((A \cup B) \cap A^c) = \theta(A) + \theta(B)$$

and therefore the restriction of $\theta$ to $\mathcal{M}(\theta)$ is a content.

Next we prove that $\mathcal{M}(\theta)$ is a $\sigma$-algebra. Let $(A_i)_{i=1}^{\infty}$ be a disjoint denumerable collection of members of $\mathcal{M}(\theta)$ and set for each $n \in \mathbf{N}$

$$B_n = \cup_{1\leq i\leq n} A_i \text{ and } B = \cup_{i=1}^{\infty} A_i$$

(here $B_0 = \phi$). Then for any $E \subseteq X$,

$$\theta(E \cap B_n) = \theta(E \cap B_n \cap A_n) + \theta(E \cap B_n \cap A_n^c)$$

$$= \theta(E \cap A_n) + \theta(E \cap B_{n-1})$$

and, by induction,

$$\theta(E \cap B_n) = \Sigma_{i=1}^{n} \theta(E \cap A_i).$$

But then

$$\theta(E) = \theta(E \cap B_n) + \theta(E \cap B_n^c)$$
$$\geq \Sigma_{i=1}^{n} \theta(E \cap A_i) + \theta(E \cap B^c)$$

and letting $n \to \infty$,

$$\theta(E) \geq \Sigma_{i=1}^{\infty} \theta(E \cap A_i) + \theta(E \cap B^c)$$

$$\geq \theta(\cup_{i=1}^{\infty} (E \cap A_i)) + \theta(E \cap B^c)$$
$$= \theta(E \cap B) + \theta(E \cap B^c) \geq \theta(E).$$

All the inequalities in the last calculation must be equalities and we conclude that $B \in \mathcal{M}(\theta)$ and, choosing $E = B$, results in

$$\theta(B) = \Sigma_{i=1}^{\infty} \theta(A_i).$$

Thus $\mathcal{M}(\theta)$ is a $\sigma$-algebra and the restriction of $\theta$ to $\mathcal{M}(\theta)$ is a positive measure.

Finally we prove that the the restriction of $\theta$ to $\mathcal{M}(\theta)$ is a complete measure. Suppose $B \subseteq A \in \mathcal{M}(\theta)$ and $\theta(A) = 0$. If $E \subseteq X$,

$$\theta(E) \leq \theta(E \cap B) + \theta(E \cap B^c) \leq \theta(E \cap B^c) \leq \theta(E)$$

and so $B \in \mathcal{M}(\theta)$. The theorem is proved.

**Exercises**

1. Suppose $\theta_i : \mathcal{P}(X) \to [0, \infty]$ , $i = 1, 2$, are outer measures. Prove that $\theta = \max(\theta_1, \theta_2)$ is an outer measure.

2. Suppose $a, b \in \mathbf{R}$ and $a \neq b$. Set $\theta = \max(\delta_a, \delta_b)$. Prove that

$$\{a\}, \{b\} \notin \mathcal{M}(\theta).$$

## 1.5. Existence of Linear Measure

The purpose of this section is to show the existence of linear measure on $\mathbf{R}$ using the Carathéodory Theorem and a minimum of topology.

First let us recall the definition of infimum and supremum of a non-empty subset of the extended real line. Suppose $A$ is a non-empty subset of $[-\infty, \infty] = \mathbf{R} \cup \{-\infty, \infty\}$. We define $-\infty \leq x$ and $x \leq \infty$ for all $x \in [-\infty, \infty]$. An element $b \in [-\infty, \infty]$ is called a majorant of $A$ if $x \leq b$ for all $x \in A$ and a minorant if $x \geq b$ for all $x \in A$. The Supremum Axiom states that $A$ possesses a least majorant, which is denoted by $\sup A$. From this follows that if $A$ is non-empty, then $A$ possesses a greatest minorant, which is denoted by $\inf A$. (Actually, the Supremum Axiom is a theorem in courses where time is spent on the definition of real numbers.)

**Theorem 1.5.1. (The Heine-Borel Theorem; weak form)** *Let $[a, b]$ be a closed bounded interval and $(U_i)_{i \in I}$ a collection of open sets such that*

$$\cup_{i \in I} U_i \supseteq [a, b].$$

*Then*

$$\cup_{i \in J} U_i \supseteq [a, b]$$

*for some finite subset $J$ of $I$.*

PROOF. Let $A$ be the set of all $x \in [a, b]$ such that

$$\cup_{i \in J} U_i \supseteq [a, x]$$

for some finite subset $J$ of $I$. Clearly, $a \in A$ since $a \in U_i$ for some $i$. Let $c = \sup A$. There exists an $i_0$ such that $c \in U_{i_0}$. Let $c \in ]a_0, b_0[ \subseteq U_{i_0}$, where $a_0 < b_0$. Furthermore, by the very definition of least upper bound, there exists a finite set $J$ such that

$$\cup_{i \in J} U_i \supseteq [a, (a_0 + c)/2] .$$

Hence

$$\cup_{i \in J \cup \{i_0\}} U_k \supseteq [a, (c + b_0)/2]$$

and it follows that $c \in A$ and $c = b$. The lemma is proved.

A subset $K$ of $\mathbf{R}$ is called compact if for every family of open subsets $U_i$, $i \in I$, with $\cup_{i \in I} U_i \supseteq K$ we have $\cup_{i \in J} U_i \supseteq K$ for some finite subset $J$ of $I$. The Heine-Borel Theorem shows that a closed bounded interval is compact.

If $x, y \in \mathbf{R}$ and $E$, $F \subseteq \mathbf{R}$, let

$$d(x, y) = \mid x - y \mid$$

be the distance between $x$ and $y$, let

$$d(x, E) = \inf_{u \in E} d(x, u)$$

be the distance from $x$ to $E$, and let

$$d(E, F) = \inf_{u \in E, v \in F} d(u, v)$$

be the distance between $E$ and $F$ (here the infimum of the emty set equals $\infty$). Note that for any $u \in E$,

$$d(x, u) \leq d(x, y) + d(y, u)$$

and, hence

$$d(x, E) \leq d(x, y) + d(y, u)$$

and

$$d(x, E) \leq d(x, y) + d(y, E).$$

By interchanging the roles of $x$ and $y$ and assuming that $E \neq \phi$, we get

$$| \, d(x, E) - d(y, E) \, | \leq d(x, y).$$

Note that if $F \subseteq \mathbf{R}$ is closed and $x \notin F$, then $d(x, F) > 0$.

An outer measure $\theta : \mathcal{P}(\mathbf{R}) \to [0, \infty]$ is called a metric outer measure if

$$\theta(A \cup B) = \theta(A) + \theta(B)$$

for all $A, B \in \mathcal{P}(\mathbf{R})$ such that $d(A, B) > 0$.

**Theorem 1.5.2.** *If* $\theta : \mathcal{P}(\mathbf{R}) \to [0, \infty]$ *is a metric outer measure, then* $\mathcal{R} \subseteq \mathcal{M}(\theta)$.

PROOF. Let $F \in \mathcal{P}(\mathbf{R})$ be closed. It is enough to show that $F \in \mathcal{M}(\theta)$. To this end we choose $E \subseteq X$ with $\theta(E) < \infty$ and prove that

$$\theta(E) \geq \theta(E \cap F) + \theta(E \cap F^c).$$

Let $n \geq 1$ be an integer and define

$$A_n = \left\{ x \in E \cap F^c; \; d(x, F) \geq \frac{1}{n} \right\}.$$

Note that $A_n \subseteq A_{n+1}$ and

$$E \cap F^c = \cup_{n=1}^{\infty} A_n.$$

Moreover, since $\theta$ is a metric outer measure

$$\theta(E) \geq \theta((E \cap F) \cup A_n) = \theta(E \cap F) + \theta(A_n)$$

and, hence, proving

$$\theta(E \cap F^c) = \lim_{n \to \infty} \theta(A_n)$$

we are done.

Let $B_n = A_{n+1} \cap A_n^c$. It is readily seen that

$$d(B_{n+1}, A_n) \geq \frac{1}{n(n+1)}$$

since if $x \in B_{n+1}$ and

$$d(x, y) < \frac{1}{n(n+1)}$$

then

$$d(y, F) \leq d(y, x) + d(x, F) < \frac{1}{n(n+1)} + \frac{1}{n+1} = \frac{1}{n}.$$

Now

$$\begin{aligned}\theta(A_{2k+1}) &\geq \theta(B_{2k} \cup A_{2k-1}) = \theta(B_{2k}) + \theta(A_{2k-1}) \\ &\geq ... \geq \Sigma_{i=1}^{k} \theta(B_{2i})\end{aligned}$$

and in a similar way

$$\theta(A_{2k}) \geq \Sigma_{i=1}^{k} \theta(B_{2i-1}).$$

But $\theta(A_n) \leq \theta(E) < \infty$ and we conclude that

$$\Sigma_{i=1}^{\infty} \theta(B_i) < \infty.$$

We now use that

$$E \cap F^c = A_n \cup (\cup_{i=n}^{\infty} B_i)$$

to obtain

$$\theta(E \cap F^c) \leq \theta(A_n) + \Sigma_{i=n}^{\infty} \theta(B_i).$$

Now, since $\theta(E \cap F^c) \geq \theta(A_n)$,

$$\theta(E \cap F^c) = \lim_{n \to \infty} \theta(A_n)$$

and the theorem is proved.

PROOF OF THEOREM 1.1.1 IN ONE DIMENSION. Suppose $\delta > 0$. If $A \subseteq \mathbf{R}$, define

$$\theta_\delta(A) = \inf \Sigma_{k=1}^{\infty} l(I_k)$$

the infimum being taken over all open intervals $I_k$ with $l(I_k) < \delta$ such that

$$A \subseteq \cup_{k=1}^{\infty} I_k.$$

Obviously, $\theta_\delta(\phi) = 0$ and $\theta_\delta(A) \le \theta_\delta(B)$ if $A \subseteq B$. Suppose $(A_n)_{n=1}^\infty$ is a denumerable collection of subsets of $\mathbf{R}$ and let $\varepsilon > 0$. For each $n$ there exist open intervals $I_{kn}, k \in \mathbf{N}_+$, such that $l(I_{kn}) < \delta$,

$$A_n \subseteq \cup_{k=1}^\infty I_{kn}$$

and

$$\Sigma_{k=1}^\infty l(I_{kn}) \le \theta_\delta(A_n) + \varepsilon 2^{-n}.$$

Then

$$A =_{def} \cup_{n=1}^\infty A_n \subseteq \cup_{k,n=1}^\infty I_{kn}$$

and

$$\Sigma_{k,n=1}^\infty l(I_{kn}) \le \Sigma_{n=1}^\infty \theta_\delta(A_n) + \varepsilon.$$

Thus

$$\theta_\delta(A) \le \Sigma_{n=1}^\infty \theta_\delta(A_n) + \varepsilon$$

and, since $\varepsilon > 0$ is arbitrary,

$$\theta_\delta(A) \le \Sigma_{n=1}^\infty \theta_\delta(A_n).$$

It follows that $\theta_\delta$ is an outer measure.

If $I$ is an open interval it is simple to see that

$$\theta_\delta(I) \le l(I).$$

To prove the reverse inequality, choose a closed bounded interval $J \subseteq I$. Now, if

$$I \subseteq \cup_{k=1}^\infty I_k$$

where each $I_k$ is an open interval of $l(I_k) < \delta$, it follows from the Heine-Borel Theorem that

$$J \subseteq \cup_{k=1}^n I_k$$

for some $n$. Hence

$$l(J) \le \Sigma_{k=1}^n l(I_k) \le \Sigma_{k=1}^\infty l(I_k)$$

and it follows that

$$l(J) \le \theta_\delta(I)$$

and, accordingly from this,

$$l(I) \le \theta_\delta(I).$$

Thus, if $I$ is an open interval, then

$$\theta_\delta(I) = l(I).$$

Note that $\theta_{\delta_1} \geq \theta_{\delta_2}$ if $0 < \delta_1 \leq \delta_2$. We define

$$\theta_0(A) = \lim_{\delta \to 0} \theta_\delta(A) \text{ if } A \subseteq \mathbf{R}.$$

It obvious that $\theta_0$ is an outer measure such that $\theta_0(I) = l(I)$, if $I$ is an open interval.

To complete the proof we show that $\theta_0$ is a metric outer measure. To this end let $A, B \subseteq \mathbf{R}$ and $d(A, B) > 0$. Suppose $0 < \delta < d(A, B)$ and

$$A \cup B \subseteq \cup_{k=1}^{\infty} I_k$$

where each $I_k$ is an open interval with $l(I_k) < \delta$. Let

$$\alpha = \{k;\ I_k \cap A \neq \phi\}$$

and

$$\beta = \{k;\ I_k \cap B \neq \phi\}.$$

Then $\alpha \cap \beta = \phi$,

$$A \subseteq \cup_{k \in \alpha} I_k$$

and

$$B \subseteq \cup_{k \in \beta} I_k$$

and it follows that

$$\Sigma_{k=1}^{\infty} l(I_k) \geq \Sigma_{k \in \alpha} l(I_k) + \Sigma_{k \in \beta} l(I_k)$$

$$\geq \theta_\delta(A) + \theta_\delta(B).$$

Thus

$$\theta_\delta(A \cup B) \geq \theta_\delta(A) + \theta_\delta(B)$$

and by letting $\delta \to 0$ we have

$$\theta_0(A \cup B) \geq \theta_0(A) + \theta_0(B)$$

and

$$\theta_0(A \cup B) = \theta_0(A) + \theta_0(B).$$

Finally by applying the Carathéodory Theorem and Theorem 1.5.2 it follows that the restriction of $\theta_0$ to $\mathcal{R}$ equals $v_1$.

We end this section with some additional results of great interest.

**Theorem 1.5.3.** *For any* $\delta > 0$, $\theta_\delta = \theta_0$. *Moreover, if* $A \subseteq \mathbf{R}$

$$\theta_0(A) = \inf \Sigma_{k=1}^{\infty} l(I_k)$$

*the infimum being taken over all open intervals* $I_k$, $k \in \mathbf{N}_+$, *such that* $\cup_{k=1}^{\infty} I_k \supseteq A$.

PROOF. It follows from the definition of $\theta_0$ that $\theta_\delta \leq \theta_0$. To prove the reverse inequality let $A \subseteq \mathbf{R}$ and choose open intervals $I_k$, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$. Then

$$\theta_0(A) \leq \theta_0(\cup_{k=1}^{\infty} I_k) \leq \Sigma_{k=1}^{\infty} \theta_0(I_k)$$

$$= \Sigma_{k=1}^{\infty} l(I_k).$$

Hence

$$\theta_0(A) \leq \inf \Sigma_{k=1}^{\infty} l(I_k)$$

the infimum being taken over all open intervals $I_k$, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$. Thus $\theta_0(A) \leq \theta_\delta(A)$, which completes the proof of Theorem 1.5.3.

**Theorem 1.5.4.** *If* $A \subseteq \mathbf{R}$,

$$\theta_0(A) = \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

*Moreover, if* $A \in \mathcal{M}(\theta_0)$,

$$\theta_0(A) = \sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K).$$

PROOF. If $A \subseteq U$, $\theta_0(A) \leq \theta_0(U)$. Hence

$$\theta_0(A) \leq \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

Next let $\varepsilon > 0$ be fixed and choose open intervals $I_k$, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$ and

$$\Sigma_{k=1}^{\infty} l(I_k) \leq \theta_0(A) + \varepsilon$$

(here observe that it may happen that $\theta_0(A) = \infty$). Then the set $U =_{def} \cup_{k=1}^{\infty} I_k$ is open and

$$\theta_0(U) \leq \Sigma_{k=1}^{\infty} \theta_0(I_k) = \Sigma_{k=1}^{\infty} l(I_k) \leq \theta_0(A) + \varepsilon.$$

Thus

$$\inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U) \leq \theta_0(A)$$

and we have proved that

$$\theta_0(A) = \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

If $K \subseteq A$, $\theta_0(K) \leq \theta_0(A)$ and, accordingly from this,

$$\sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K) \leq \theta_0(A).$$

To prove the reverse inequality we first assume that $A \in \mathcal{M}(\theta_0)$ is bounded. Let $\varepsilon > 0$ be fixed and suppose $J$ is a closed bounded interval containing $A$. Then we know from the first part of Theorem 1.5.4 already proved that there exists an open set $U \supseteq J \setminus A$ such that

$$\theta_0(U) < \theta_0(J \setminus A) + \varepsilon.$$

But then

$$\theta_0(J) \leq \theta_0(J \setminus U) + \theta_0(U) < \theta_0(J \setminus U) + \theta_0(J \setminus A) + \varepsilon$$

and it follows that

$$\theta_0(A) - \varepsilon < \theta_0(J \setminus U).$$

Since $J \setminus U$ is a closed bounded set contained in $A$ we conclude that

$$\theta_0(A) \leq \sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K).$$

If $A \in \mathcal{M}(\theta_0)$ let $A_n = A \cap [-n, n]$ , $n \in \mathbf{N}_+$. Then given $\varepsilon > 0$ and $n \in \mathbf{N}_+$, let $K_n$ be a closed bounded subset of $A_n$ such that $\theta_0(K_n) > \theta_0(A_n) - \varepsilon$. Clearly, there is no loss of generality to assume that $K_1 \subseteq K_2 \subseteq K_3 \subseteq ...$ and by letting $n$ tend to plus infinity we get

$$\lim_{n \to \infty} \theta_0(K_n) \geq \theta_0(A) - \varepsilon.$$

Hence

$$\theta_0(A) = \sup_{\substack{K \subseteq A \\ K \text{ compact}}} \theta_0(K).$$

and Theorem 1.5.4 is completely proved.

**Theorem 1.5.5.** *Lebesgue measure $m_1$ equals the restriction of $\theta_0$ to $\mathcal{M}(\theta_0)$.*

PROOF. Recall that linear measure $v_1$ equals the restriction of $\theta_0$ to $\mathcal{R}$ and $m_1 = \bar{v}_1$. First suppose $E \in \mathcal{R}^-$ and choose $A, B \in \mathcal{R}$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_{v_1}$. But then $\theta_0(E \setminus A) = 0$ and $E = A \cup (E \setminus A) \in \mathcal{M}(\theta_0)$ since the Carathéodory Theorem gives us a complete measure. Hence $m_1(E) = v_1(A) = \theta_0(E)$.

Conversely suppose $E \in \mathcal{M}(\theta_0)$. We will prove that $E \in \mathcal{R}^-$ and $m_1(E) = \theta_0(E)$. First assume that $E$ is bounded. Then for each positive integer $n$ there exist open $U_n \supseteq E$ and closed bounded $K_n \subseteq E$ such that

$$\theta_0(U_n) < \theta_0(E) + 2^{-n}$$

and

$$\theta_0(K_n) > \theta_0(E) - 2^{-n}.$$

The definitions yield $A = \cup_1^\infty K_n$, $B = \cap_1^\infty U_n \in \mathcal{R}$ and

$$\theta_0(E) = \theta_0(A) = \theta_0(B) = v_1(A) = v_1(B) = m_1(E).$$

It follows that $E \in \mathcal{R}^-$ and $\theta_0(E) = m_1(E)$.

In the general case set $E_n = E \cap [-n, n]$ , $n \in \mathbf{N}_+$. Then from the above $E_n \in \mathcal{R}^-$ and $\theta_0(E_n) = m_1(E_n)$ for each $n$ and Theorem 1.5.5 follows by letting $n$ go to infinity.

The Carathéodory Theorem can be used to show the existence of volume measure on $\mathbf{R}^n$ but we do not go into this here since its existence follows by several other means below. By passing, let us note that the Carathéodory Theorem is very efficient to prove the existence of so called Haussdorff measures (see e.g. $[F]$), which are of great interest in Geometric Measure Theory.

**Exercises**

1. Prove that a subset $K$ of $\mathbf{R}$ is compact if and only if $K$ is closed and bounded.

2. Suppose $A \in \mathcal{R}^-$ and $m(A) < \infty$. Set $f(x) = m(A \cap \,]-\infty, x])$, $x \in \mathbf{R}$. Prove that $f$ is continuous.

3. Suppose $A \in \mathcal{Z}_m$ and $B = \{x^3; x \in A\}$ . Prove that $B \in \mathcal{Z}_m$.

4. Let $A$ be the set of all real numbers $x$ such that

$$\mid x - \frac{p}{q} \mid \leq \frac{1}{q^3}$$

for infinitely many pairs of positive integers $p$ and $q$. Prove that $A \in \mathcal{Z}_m$.

5. Let $I_1, ..., I_n$ be open subintervals of $\mathbf{R}$ such that

$$\mathbf{Q} \cap [0, 1] \subseteq \cup_{k=1}^n I_k.$$

Prove that $\Sigma_{k=1}^n m(I_k) \geq 1$.

6. If $E \in \mathcal{R}^-$ and $m(E) > 0$, for every $\alpha \in ]0,1[$ there is an interval $I$ such that $m(E \cap I) > \alpha m(I)$. (Hint: $m(E) = \inf \Sigma_{k=1}^{\infty} m(I_k)$, where the infimum is taken over all intervals such that $\cup_{k=1}^{\infty} I_k \supseteq E$.)

7. If $E \in \mathcal{R}^-$ and $m(E) > 0$, then the set $E - E = \{x - y; x, y \in E\}$ contains an open non-empty interval centred at 0.(Hint: Take an interval $I$ with $m(E \cap I) \geq \frac{3}{4} m(I)$. Set $\varepsilon = \frac{1}{2} m(I)$. If $\mid x \mid \leq \varepsilon$, then $(E \cap I) \cap (x + (E \cap I)) \neq \phi$.)

8. Let $\mu$ be the restriction of the positive measure $\Sigma_{k=1}^{\infty} \delta_{\mathbf{R}, \frac{1}{k}}$ to $\mathcal{R}$. Prove that

$$\inf_{\substack{U \supseteq A \\ U \text{ open}}} \mu(U) > \mu(A)$$

if $A = \{0\}$.

## 1.6. Positive Measures Induced by Increasing Right Continuous Functions

Suppose $F : \mathbf{R} \to [0, \infty[$ is a right continuous increasing function such that

$$\lim_{x \to -\infty} F(x) = 0.$$

Set

$$L = \lim_{x \to \infty} F(x).$$

We will prove that there exists a unique positive measure $\mu : \mathcal{R} \to [0, L]$ such that

$$\mu(]-\infty, x]) = F(x), \ x \in \mathbf{R}.$$

This measure will often be denoted by $\mu_F$.

The special case $L = 0$ is trivial so let us assume $L > 0$ and introduce

$$H(y) = \inf \{x \in \mathbf{R}; F(x) \geq y\}, \ 0 < y < L.$$

The definition implies that the function $H$ increases.

Suppose $a$ is a fixed real number. We claim that

$$\{y \in ]0, L[\,;\ H(y) \leq a\} = ]0, F(a)] \cap ]0, L[\,.$$

To prove this first suppose that $y \in ]0, L[$ and $H(y) \leq a$. Then to each positive integer $n$, there is an $x_n \in [H(y), H(y) + 2^{-n}[$ such that $F(x_n) \geq y$. Then $x_n \to H(y)$ as $n \to \infty$ and we obtain that $F(H(y)) \geq y$ since $F$ is right continuous. Thus, remembering that $F$ increases, $F(a) \geq y$. On the other hand, if $0 < y < L$ and $0 < y \leq F(a)$, then, by the very definition of $H(y)$, $H(y) \leq a$.

We now define

$$\mu = H(v_{1_{|]0,L[}})$$

and get

$$\mu(]-\infty, x]) = F(x),\ \ x \in \mathbf{R}.$$

The uniqueness follows at once from Theorem 1.2.3. Note that the measure $\mu$ is a probability measure if $L = 1$.

**Example 1.1.1.** If

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

then $\mu_F$ is the Dirac measure at the point 0 restricted to $\mathcal{R}$.

**Example 1.1.2.** If

$$F(x) = \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} \text{ (a Riemann integral)}$$

then $\mu_F$ is called the standard Gaussian measure on $\mathbf{R}$.

**Exercises**

1. Suppose $F : \mathbf{R} \to \mathbf{R}$ is a right continuous increasing function. Prove that there is a unique positive measure $\mu$ on $\mathcal{R}$ such that

$$\mu(]a, x]) = F(x) - F(a),\ \text{ if } a, x \in \mathbf{R} \text{ and } a < x.$$

2. Suppose $F : \mathbf{R} \to \mathbf{R}$ is an increasing function. Prove that the set of all discontinuity points of $F$ is at most denumerable. (Hint: Assume first that $F$ is bounded and prove that the set of all points $x \in \mathbf{R}$ such that $F(x+) - F(x-) > \varepsilon$ is finite for every $\varepsilon > 0$.)

3. Suppose $\mu$ is a $\sigma$-finite positive measure on $\mathcal{R}$. Prove that the set of all $x \in \mathbf{R}$ such that $\mu(\{x\}) > 0$ is at most denumerable.

4. Suppose $\mu$ is a $\sigma$-finite positive measure on $\mathcal{R}_n$. Prove that there is an at most denumerable set of hyperplanes of the type

$$x_k = c \qquad (k = 1, ..., n,\ c \in \mathbf{R})$$

with positive $\mu$-measure.

5. Construct an increasing function $f : \mathbf{R} \to \mathbf{R}$ such that the set of discontinuity points of $f$ equals $\mathbf{Q}$.

# CHAPTER 2

# INTEGRATION

## Introduction

In this chapter Lebesgue integration in abstract positive measure spaces is introduced. A series of famous theorems and lemmas will be proved.

## 2.1. Integration of Functions with Values in $[0,\infty]$

Recall that $[0,\infty] = [0,\infty[ \cup \{\infty\}$. A subinterval of $[0,\infty]$ is defined in the natural way. We denote by $\mathcal{R}_{0,\infty}$ the $\sigma$-algebra generated by all subintervals of $[0,\infty]$. The class of all intervals of the type $]\alpha,\infty]$, $0 \leq \alpha < \infty$, (or of the type $[\alpha,\infty]$, $0 \leq \alpha < \infty$) generates the $\sigma$-algebra $\mathcal{R}_{0,\infty}$ and we get the following

**Theorem 2.1.1.** *Let $(X,\mathcal{M})$ be a measurable space and suppose $f : X \to [0,\infty]$.*

(a) *The function $f$ is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if $f^{-1}(]\alpha,\infty]) \in \mathcal{M}$ for every $0 \leq \alpha < \infty$.*

(b) *The function $f$ is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if $f^{-1}([\alpha,\infty]) \in \mathcal{M}$ for every $0 \leq \alpha < \infty$.*

Note that the set $\{f > \alpha\} \in \mathcal{M}$ for all real $\alpha$ if $f$ is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable.

If $f,g : X \to [0,\infty]$ are $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable, then $\min(f,g)$, $\max(f,g)$, and $f+g$ are $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable, since, for each $\alpha \in [0,\infty[$,

$$\min(f,g) \geq \alpha \Leftrightarrow (f \geq \alpha \text{ and } g \geq \alpha)$$

$$\max(f,g) \geq \alpha \Leftrightarrow (f \geq \alpha \text{ or } g \geq \alpha)$$

and

$$\{f+g>\alpha\}=\bigcup_{q\in\mathbf{Q}}(\{f>\alpha-q\}\cap\{g>q\}).$$

Given functions $f_n : X \to [0,\infty]$ , $n = 1, 2, ...,$ $f = \sup_{n\geq 1} f_n$ is defined by the equation

$$f(x)=\sup\{f_n(x);\ n=1,2,...\}.$$

Note that

$$f^{-1}(]\alpha,\infty])=\cup_{n=1}^{\infty}f_n^{-1}(]\alpha,\infty])$$

for every real $\alpha \geq 0$ and, accordingly from this, the function $\sup_{n\geq 1} f_n$ is $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable if each $f_n$ is $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable. Moreover, $f = \inf_{n\geq 1} f_n$ is given by

$$f(x)=\inf\{f_n(x);\ n=1,2,...\}.$$

Since

$$f^{-1}([0,\alpha[)=\cup_{n=1}^{\infty}f_n^{-1}([0,\alpha[)$$

for every real $\alpha \geq 0$ we conclude that the function $f = \inf_{n\geq 1} f_n$ is $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable if each $f_n$ is $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable.

Below we write

$$f_n \uparrow f$$

if $f_n$, $n = 1, 2, ...,$ and $f$ are functions from $X$ into $[0,\infty]$ such that $f_n \leq f_{n+1}$ for each $n$ and $f_n(x) \to f(x)$ for each $x \in X$ as $n \to \infty$.

An $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function $\varphi : X \to [0,\infty]$ is called a simple measurable function if $\varphi(X)$ is a finite subset of $[0,\infty[$ . If it is neccessary to be more precise, we say that $\varphi$ is a simple $\mathcal{M}$-measurable function.

**Theorem 2.1.2.** *Let $f : X \to [0,\infty]$ be $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable. There exist simple measurable functions $\varphi_n$, $n \in \mathbf{N}_+$, on $X$ such that $\varphi_n \uparrow f$ .*

PROOF. Given $n \in \mathbf{N}_+$, set

$$E_{in}=f^{-1}(\left[\frac{i-1}{2^n},\frac{i}{2^n}\right[),\ i\in\mathbf{N}_+$$

and

$$\rho_n = \sum_{i=1}^{\infty} \frac{i-1}{2^n} \chi_{E_{in}} + \infty \chi_{f^{-1}(\{\infty\})}.$$

It is obvious that $\rho_n \leq f$ and that $\rho_n \leq \rho_{n+1}$. Now set $\varphi_n = \min(n, \rho_n)$ and we are done.

Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $\varphi : X \to [0, \infty[$ a simple measurable function. If $\alpha_1, ..., \alpha_n$ are the distinct values of the simple function $\varphi$, and if $E_i = \varphi^{-1}(\{\alpha_i\})$, $i = 1, ..., n$, then

$$\varphi = \Sigma_{i=1}^n \alpha_i \chi_{E_i}.$$

Furthermore, if $A \in \mathcal{M}$ we define

$$\nu(A) = \int_A \varphi d\mu = \Sigma_{i=1}^n \alpha_i \mu(E_i \cap A) = \Sigma_{k=1}^n \alpha_i \mu^{E_i}(A).$$

Note that this formula still holds if $(E_i)_1^n$ is a measurable partition of $X$ and $\varphi = \alpha_i$ on $E_i$ for each $i = 1, ..., n$. Clearly, $\nu$ is a positive measure since each term in the right side is a positive measure as a function of $A$. Note that

$$\int_A \alpha\varphi d\mu = \alpha \int_A \varphi d\mu \text{ if } 0 \leq \alpha < \infty$$

and

$$\int_A \varphi d\mu = a\mu(A)$$

if $a \in [0, \infty[$ and $\varphi$ is a simple measurable function such that $\varphi = a$ on $A$.

If $\psi$ is another simple measurable function and $\varphi \leq \psi$,

$$\int_A \varphi d\mu \leq \int_A \psi d\mu.$$

To see this, let $\beta_1, ..., \beta_p$ be the distinct values of $\psi$ and $F_j = \psi^{-1}(\{\beta_j\})$, $j = 1, ..., p$. Now, putting $B_{ij} = E_i \cap F_j$,

$$\int_A \varphi d\mu = \nu(\cup_{ij}(A \cap B_{ij}))$$

$$= \Sigma_{ij} \nu(A \cap B_{ij}) = \Sigma_{ij} \int_{A \cap B_{ij}} \varphi d\mu = \Sigma_{ij} \int_{A \cap B_{ij}} \alpha_i d\mu$$

$$\leq \Sigma_{ij} \int_{A\cap B_{ij}} \beta_j d\mu = \int_A \psi d\mu.$$

In a similar way one proves that

$$\int_A (\varphi + \psi) d\mu = \int_A \varphi d\mu + \int_A \psi d\mu.$$

From the above it follows that

$$\int_A \varphi \chi_A d\mu = \int_A \Sigma_{i=1}^n \alpha_i \chi_{E_i \cap A} d\mu$$

$$= \Sigma_{i=1}^n \alpha_i \int_A \chi_{E_i \cap A} d\mu = \Sigma_{i=1}^n \alpha_i \mu(E_i \cap A)$$

and

$$\int_A \varphi \chi_A d\mu = \int_A \varphi d\mu.$$

If $f : X \to [0, \infty]$ is an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function and $A \in \mathcal{M}$, we define

$$\int_A f d\mu = \sup \left\{ \int_A \varphi d\mu;\ 0 \leq \varphi \leq f,\ \varphi \text{ simple measurable} \right\}$$

$$= \sup \left\{ \int_A \varphi d\mu;\ 0 \leq \varphi \leq f,\ \varphi \text{ simple measurable and } \varphi = 0 \text{ on } A^c \right\}.$$

The left member in this equation is called the Lebesgue integral of $f$ over $A$ with respect to the measure $\mu$. Sometimes we also speek of the $\mu$-integral of $f$ over $A$. The two definitions of the $\mu$-integral of a simple measurable function $\varphi : X \to [0, \infty[$ over $A$ agree.

From now on in this section, an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function $f : X \to [0, \infty]$ is simply called measurable.

The following properties are immediate consequences of the definitions. The functions and sets occurring in the equations are assumed to be measurable.

(a) If $f$, $g \geq 0$ and $f \leq g$ on $A$, then $\int_A f d\mu \leq \int_A g d\mu$.

(b) $\int_A f d\mu = \int_X \chi_A f d\mu$.

(c) If $f \geq 0$ and $\alpha \in [0, \infty[$, then $\int_A \alpha f d\mu = \alpha \int_A f d\mu$.

(d) $\int_A f d\mu = 0$ if $f = 0$ and $\mu(A) = \infty$.

(e) $\int_A f d\mu = 0$ if $f = \infty$ and $\mu(A) = 0$.

If $f : X \to [0, \infty]$ is measurable and $0 < \alpha < \infty$, then $f \geq \alpha \chi_{f^{-1}([\alpha,\infty])} = \alpha \chi_{\{f \geq \alpha\}}$and

$$\int_X f d\mu \geq \int_X \alpha \chi_{\{f \geq \alpha\}} d\mu = \alpha \int_X \chi_{\{f \geq \alpha\}} d\mu.$$

This proves the so called Markov Inequality

$$\mu(f \geq \alpha) \leq \frac{1}{\alpha} \int_X f d\mu$$

where we write $\mu(f \geq \alpha)$ instead of the more precise expression $\mu(\{f \geq \alpha\})$.

**Example 2.1.1.** Suppose $f : X \to [0, \infty]$ is measurable and

$$\int_X f d\mu < \infty.$$

We claim that

$$\{f = \infty\} = f^{-1}(\{\infty\}) \in \mathcal{Z}_\mu.$$

To prove this we use the Markov Inequality and have

$$\mu(f = \infty) \leq \mu(f \geq \alpha) \leq \frac{1}{\alpha} \int_X f d\mu$$

for each $\alpha \in \,]0,\infty[\,$. Thus $\mu(f=\infty)=0$.

**Example 2.1.2.** Suppose $f : X \to [0,\infty]$ is measurable and

$$\int_X f d\mu = 0.$$

We claim that

$$\{f > 0\} = f^{-1}(]0,\infty]) \in \mathcal{Z}_\mu.$$

To see this, note that

$$f^{-1}(]0,\infty]) = \cup_{n=1}^{\infty} f^{-1}(\left]\frac{1}{n},\infty\right])$$

Furthermore, for every fixed $n \in \mathbf{N}_+$, the Markov Inequality yields

$$\mu(f > \frac{1}{n}) \leq n \int_X f d\mu = 0$$

and we get $\{f > 0\} \in \mathcal{Z}_\mu$ since a countable union of null sets is a null set.

We now come to one of the most important results in the theory.

**Theorem 2.1.3. (Monotone Convergence Theorem)** *Let* $f_n : X \to [0,\infty]$ , $n = 1,2,3,....$, *be a sequence of measurable functions and suppose that* $f_n \uparrow f$, that is $0 \leq f_1 \leq f_2 \leq ...$ *and*

$$f_n(x) \to f(x) \text{ as } n \to \infty, \text{ for every } x \in X.$$

*Then* $f$ *is measurable and*

$$\int_X f_n d\mu \to \int_X f d\mu \text{ as } n \to \infty.$$

PROOF. The function $f$ is measurable since $f = \sup_{n \geq 1} f_n$.

The inequalities $f_n \leq f_{n+1} \leq f$ yield $\int_X f_n d\mu \leq \int_X f_{n+1} d\mu \leq \int_X f d\mu$ and we conclude that there exists an $\alpha \in [0, \infty]$ such that

$$\int_X f_n d\mu \to \alpha \text{ as } n \to \infty$$

and

$$\alpha \leq \int_X f d\mu.$$

To prove the reverse inequality, let $\varphi$ be any simple measurable function such that $0 \leq \varphi \leq f$, let $0 < \theta < 1$ be a constant, and define, for fixed $n \in \mathbf{N}_+$,

$$A_n = \{x \in X;\ f_n(x) \geq \theta\varphi(x)\}.$$

If $\alpha_1, ..., \alpha_p$ are the distinct values of $\varphi$,

$$A_n = \cup_{k=1}^{p} (\{x \in X;\ f_n(x) \geq \theta\alpha_k\} \cap \{\varphi = \alpha_k\})$$

and it follows that $A_n$ is measurable. Clearly, $A_1 \subseteq A_2 \subseteq ...$ . Moreover, if $f(x) = 0$, then $x \in A_1$ and if $f(x) > 0$, then $\theta\varphi(x) < f(x)$ and $x \in A_n$ for all sufficiently large $n$. Thus $\cup_{n=1}^{\infty} A_n = X$. Now

$$\alpha \geq \int_{A_n} f_n d\mu \geq \theta \int_{A_n} \varphi d\mu$$

and we get

$$\alpha \geq \theta \int_X \varphi d\mu$$

since the map $A \to \int_A \varphi d\mu$ is a positive measure on $\mathcal{M}$. By letting $\theta \uparrow 1$,

$$\alpha \geq \int_X \varphi d\mu$$

and, hence

$$\alpha \geq \int_X f d\mu.$$

The theorem follows.

**Theorem 2.1.4.** (a) *Let $f, g : X \to [0, \infty]$ be measurable functions. Then*

$$\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

(b) **(Beppo Levi's Theorem)** *If $f_k : X \to [0, \infty]$ , $k = 1, 2, ...$ are measurable,*

$$\int_X \Sigma_{k=1}^{\infty} f_k d\mu = \Sigma_{k=1}^{\infty} \int_X f_k d\mu$$

PROOF. (a) Let $(\varphi_n)_{n=1}^{\infty}$ and $(\psi_n)_{n=1}^{\infty}$ be sequences of simple and measurable functions such that $0 \leq \varphi_n \uparrow f$ and $0 \leq \psi_n \uparrow g$. We proved above that

$$\int_X (\varphi_n + \psi_n) d\mu = \int_X \varphi_n d\mu + \int_X \psi_n d\mu$$

and, by letting $n \to \infty$, Part (a) follows from the Monotone Convergence Theorem.

(b) Part (a) and induction imply that

$$\int_X \Sigma_{k=1}^{n} f_k d\mu = \Sigma_{k=1}^{n} \int_X f_k d\mu$$

and the result follows from monotone convergence.

**Theorem 2.1.5.** *Suppose $w : X \to [0, \infty]$ is a measurable function and define*

$$\nu(A) = \int_A w d\mu, \ A \in \mathcal{M}.$$

*Then $\nu$ is a positive measure and*

$$\int_A f d\nu = \int_A f w d\mu, \ A \in \mathcal{M}$$

*for every measurable function $f : X \to [0, \infty]$.*

PROOF. Clearly, $\nu(\phi) = 0$. Suppose $(E_k)_{k=1}^{\infty}$ is a disjoint denumerable collection of members of $\mathcal{M}$ and set $E = \cup_{k=1}^{\infty} E_k$. Then

$$\nu(\cup_{k=1}^{\infty} E_k) = \int_E w d\mu = \int_X \chi_E w d\mu = \int_X \Sigma_{k=1}^{\infty} \chi_{E_k} w d\mu$$

where, by the Beppo Levi Theorem, the right member equals

$$\Sigma_{k=1}^{\infty} \int_X \chi_{E_k} w d\mu = \Sigma_{k=1}^{\infty} \int_{E_k} w d\mu = \Sigma_{k=1}^{\infty} \nu(E_k).$$

This proves that $\nu$ is a positive measure.

Let $A \in \mathcal{M}$. To prove the last part in Theorem 2.1.5 we introduce the class $\mathcal{C}$ of all measurable functions $f : X \to [0, \infty]$ such that

$$\int_A f d\nu = \int_A f w d\mu.$$

The indicator function of a measurable set belongs to $\mathcal{C}$ and from this we conclude that every simple measurable function belongs to $\mathcal{C}$. Furthermore, if $f_n \in \mathcal{C}$, $n \in \mathbf{N}$, and $f_n \uparrow f$ , the Monotone Convergence Theorem proves that $f \in \mathcal{C}$. Thus in view of Theorem 2.1.2 the class $\mathcal{C}$ contains every measurable function $f : X \to [0, \infty]$ . This completes the proof of Theorem 2.1.5.

The measure $\nu$ in Theorem 2.1.5 is written

$$\nu = w\mu$$

or

$$d\nu = w d\mu.$$

Let $(\alpha_n)_{n=1}^{\infty}$ be a sequence in $[-\infty, \infty]$ . First put $\beta_k = \inf \{\alpha_k, \alpha_{k+1}, \alpha_{k+2}, ...\}$ and $\gamma = \sup \{\beta_1, \beta_2, \beta_3, ..\} = \lim_{n\to\infty} \beta_n$. We call $\gamma$ the lower limit of $(\alpha_n)_{n=1}^{\infty}$ and write

$$\gamma = \liminf_{n\to\infty} \alpha_n.$$

Note that

$$\gamma = \lim_{n\to\infty} \alpha_n$$

if the limit exists. Now put $\beta_k = \sup \{\alpha_k, \alpha_{k+1}, \alpha_{k+2}, ...\}$ and $\gamma = \inf \{\beta_1, \beta_2, \beta_3, ..\} = \lim_{n\to\infty} \beta_n$. We call $\gamma$ the upper limit of $(\alpha_n)_{n=1}^{\infty}$ and write

$$\gamma = \limsup_{n\to\infty} \alpha_n.$$

Note that

$$\gamma = \lim_{n\to\infty} \alpha_n$$

if the limit exists.

Given measurable functions $f_n : X \to [0,\infty]$, $n = 1,2,...$, the function $\liminf_{n\to\infty} f_n$ is measurable. In particular, if

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for every $x \in X$, then $f$ is measurable.

**Theorem 2.1.6. (Fatou's Lemma)** *If $f_n : X \to [0,\infty]$, $n = 1,2,...,$ are measurable*

$$\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu.$$

PROOF. Introduce

$$g_k = \inf_{n\geq k} f_n.$$

The definition gives that $g_k \uparrow \liminf_{n\to\infty} f_n$ and, moreover,

$$\int_X g_k d\mu \leq \int_X f_n d\mu,\ n \geq k$$

and

$$\int_X g_k d\mu \leq \inf_{n\geq k} \int_X f_n d\mu.$$

The Fatou Lemma now follows by monotone convergence.

Below we often write

$$\int_E f(x) d\mu(x)$$

instead of

$$\int_E f d\mu.$$

**Example 2.1.3.** Suppose $a \in \mathbf{R}$ and $f : (\mathbf{R},\mathcal{R}^-) \to ([0,\infty], \mathcal{R}_{0,\infty})$ is measurable. We claim that

$$\int_{\mathbf{R}} f(x+a) dm(x) = \int_{\mathbf{R}} f(x) dm(x).$$

First if $f = \chi_A$, where $A \in \mathcal{R}^-$,

$$\int_{\mathbf{R}} f(x+a)dm(x) = \int_{\mathbf{R}} \chi_{A-a}(x)dm(x) = m(A-a) =$$

$$m(A) = \int_{\mathbf{R}} f(x)dm(x).$$

Next it is clear that the relation we want to prove is true for simple measurable functions and finally, we use the Monotone Convergence Theorem to deduce the general case.

**Example 2.1.3,** Suppose $\Sigma_1^\infty a_n$ is a positive convergent series and let $E$ be the set of all $x \in [0,1]$ such that

$$\min_{p \in \{0,\dots,n\}} \mid x - \frac{p}{n} \mid < \frac{a_n}{n}$$

for infinitely many $n \in \mathbf{N}_+$. We claim that $E$ is a Lebesgue null set.

To prove this claim for fixed $n \in \mathbf{N}_+$, let $E_n$ be the set of all $x \in [0,1]$ such that

$$\min_{p \in \mathbf{N}_+} \mid x - \frac{p}{n} \mid < \frac{a_n}{n}.$$

Then if $B(x,r) = \,]x-r, x+r[\,$, $x \in [0,1]$, $r > 0$, we have

$$E_n \subseteq \bigcup_{p=0}^{n} B(\frac{p}{n}, \frac{a_n}{n})$$

and

$$m(E_n) \leq (n+1)\frac{2a_n}{n} \leq 4a_n.$$

Hence

$$\sum_1^\infty m(E_n) < \infty$$

and by the Beppo Levi theorem

$$\int_0^1 \sum_1^\infty \chi_{E_n} dm < \infty.$$

Accordingly from this the set

$$F = \left\{ x \in [0,1]\,;\ \sum_{1}^{\infty} \chi_{E_n}(x) < \infty \right\}$$

is of Lebesgue measure 1. Since $E \subseteq [0,1] \setminus F$ we have $m(E) = 0$.

**Exercises**

1. Suppose $f_n : X \to [0,\infty]$, $n = 1, 2, ...,$ are measurable and

$$\Sigma_{n=1}^{\infty} \mu(f_n > 1) < \infty.$$

Prove that

$$\left\{ \limsup_{n\to\infty} f_n > 1 \right\} \in \mathcal{Z}_\mu \ .$$

2. Set $f_n = n^2 \chi_{[0,\frac{1}{n}]}$, $n \in \mathbf{N}_+$. Prove that

$$\int_{\mathbf{R}} \liminf_{n\to\infty} f_n dm = 0 < \infty = \liminf_{n\to\infty} \int_{\mathbf{R}} f_n dm$$

(the inequality in the Fatou Lemma may be strict).

3. Suppose $f : (\mathbf{R},\mathcal{R}^-) \to ([0,\infty], \mathcal{R}_{0,\infty})$ is measurable and set

$$g(x) = \Sigma_{k=1}^{\infty} f(x+k),\ x \in \mathbf{R}.$$

Show that

$$\int_{\mathbf{R}} g dm < \infty \text{ if and only if } \{f > 0\} \in \mathcal{Z}_m.$$

4. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $f : X \to [0,\infty]$ an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function such that

$$f(X) \subseteq \mathbf{N}$$

and

$$\int_X f d\mu < \infty.$$

For every $t \geq 0$, set

$$F(t) = \mu(f > t) \text{ and } G(t) = \mu(f \geq t).$$

Prove that

$$\int_X f d\mu = \Sigma_{n=0}^{\infty} F(n) = \Sigma_{n=1}^{\infty} G(n).$$

## 2.2. Integration of Functions with Arbitrary Sign

As usual suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. In this section when we speak of a measurable function $f : X \to \mathbf{R}$ it is understood that $f$ is an $(\mathcal{M}, \mathcal{R})$-measurable function, if not otherwise stated. If $f, g : X \to \mathbf{R}$ are measurable, the sum $f + g$ is measurable since

$$\{f + g > \alpha\} = \bigcup_{q \in \mathbf{Q}} (\{f > \alpha - q\} \cap \{g > q\})$$

for each real $\alpha$. Besides the function $-f$ and the difference $f - g$ are measurable. It follows that a function $f : X \to \mathbf{R}$ is measurable if and only if the functions $f^+ = \max(0, f)$ and $f^- = \max(0, -f)$ are measurable since $f = f^+ - f^-$.

We write $f \in \mathcal{L}^1(\mu)$ if $f : X \to \mathbf{R}$ is measurable and

$$\int_X \mid f \mid d\mu < \infty$$

and in this case we define

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

Note that

$$\mid \int_X f d\mu \mid \leq \int_X \mid f \mid d\mu$$

since $| f |= f^+ + f^-$. Moreover, if $E \in \mathcal{M}$ we define

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

and it follows that

$$\int_E f d\mu = \int_X \chi_E f d\mu.$$

Note that

$$\int_E f d\mu = 0 \text{ if } \mu(E) = 0.$$

Sometimes we write

$$\int_E f(x) d\mu(x)$$

instead of

$$\int_E f d\mu.$$

If $f, g \in \mathcal{L}^1(\mu)$, setting $h = f + g$,

$$\int_X | h | d\mu \leq \int_X | f | d\mu + \int_X | g | d\mu < \infty$$

and it follows that $h + g \in \mathcal{L}^1(\mu)$. Moreover,

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

and the equation

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

gives

$$\int_X h^+ d\mu + \int_X f^- d\mu + \int_X g^- d\mu = \int_X f^+ d\mu + \int_X g^+ d\mu + \int_X h^- d\mu.$$

Thus

$$\int_X h d\mu = \int_X f d\mu + \int_X g d\mu.$$

Moreover,

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu$$

for each real $\alpha$. The case $\alpha \geq 0$ follows from (c) in Section 2.1. The case $\alpha = -1$ is also simple since $(-f)^+ = f^-$ and $(-f)^- = f^+$.

**Theorem 2.2.1. (Lebesgue's Dominated Convergence Theorem)** *Suppose $f_n : X \to \mathbf{R}$, $n = 1, 2, ...,$ are measurable and*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

*exists for every $x \in X$. Moreover, suppose there exists a function $g \in \mathcal{L}^1(\mu)$ such that*

$$\mid f_n(x) \mid \leq g(x), \text{ all } x \in X \text{ and } n \in \mathbf{N}_+.$$

*Then $f \in \mathcal{L}^1(\mu)$,*

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0$$

*and*

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$$

**Proof.** Since $\mid f \mid \leq g$, the function $f$ is real-valued and measurable since $f^+$ and $f^-$ are measurable. Note here that

$$f^{\pm}(x) = \lim_{n\to\infty} f_n^{\pm}(x), \text{all } x \in X.$$

We now apply the Fatous Lemma to the functions $2g- \mid f_n - f \mid, n = 1, 2, ...,$ and have

$$\int_X 2g d\mu \leq \liminf_{n\to\infty} \int_X (2g- \mid f_n - f \mid) d\mu$$

$$= \int_X 2g d\mu - \limsup_{n\to\infty} \int_X \mid f_n - f \mid d\mu.$$

But $\int_X 2g d\mu$ is finite and we get

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0.$$

Since

$$\mid \int_X f_n d\mu - \int_X f d\mu \mid = \mid \int_X (f - f_n) d\mu \mid \leq \int_X \mid f - f_n \mid d\mu$$

the last part in Theorem 2.2.1 follows from the first part. The theorem is proved.

**Example 2.2.1.** Suppose $f : ]a, b[ \times X \to \mathbf{R}$ is a function such that $f(t, \cdot) \in \mathcal{L}^1(\mu)$ for each $t \in ]a, b[$ and, moreover, assume $\frac{\partial f}{\partial t}$ exists and

$$\mid \frac{\partial f}{\partial t}(t, x) \mid \leq g(x) \text{ for all } (t, x) \in ]a, b[ \times X$$

where $g \in \mathcal{L}^1(\mu)$. Set

$$F(t) = \int_X f(t, x) d\mu(x) \text{ if } t \in ]a, b[\,.$$

We claim that $F$ is differentiable and

$$F'(t) = \int_X \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

To see this let $t_* \in ]a, b[$ be fixed and choose a sequence $(t_n)_{n=1}^{\infty}$ in $]a, b[ \setminus \{t_*\}$ which converges to $t_*$. Define

$$h_n(x) = \frac{f(t_n, x) - f(t_*, x)}{t_n - t_*} \text{ if } x \in X.$$

Here each $h_n$ is measurable and

$$\lim_{n \to \infty} h_n(x) = \frac{\partial f}{\partial t}(t_*, x) \text{ for all } x \in X.$$

Furthermore, for each fixed $n$ and $x$ there is a $\tau_{n,x} \in ]t_n, t_*[$ such that $h_n(x) = \frac{\partial f}{\partial t}(\tau_{n,x}, x)$ and we conclude that $\mid h_n(x) \mid \leq g(x)$ for every $x \in X$. Since

$$\frac{F(t_n) - F(t_*)}{t_n - t_*} = \int_X h_n(x) d\mu(x)$$

the claim above now follows from the Lebesgue Dominated Convergence Theorem.

Suppose $S(x)$ is a statement, which depends on $x \in X$. We will say that $S(x)$ holds almost (or $\mu$-almost) everywhere if there exists an $N \in \mathcal{Z}_\mu$ such that $S(x)$ holds at every point of $X \setminus N$. In this case we write "$S$ holds a.e. " or "$S$ holds a.e. $[\mu]$". Sometimes we prefer to write "$S(x)$ holds a.e." or "$S(x)$ holds a.e. $[\mu]$". If the underlying measure space is a probability space, we often say "almost surely" instead of almost everywhere. The term "almost surely" is abbreviated a.s.

Suppose $f : X \to \mathbf{R}$, is an $(\mathcal{M}, \mathcal{R})$-measurable functions and $g : X \to \mathbf{R}$. If $f = g$ a.e. $[\mu]$ there exists an $N \in \mathcal{Z}_\mu$ such that $f(x) = g(x)$ for every $x \in X \setminus N$. We claim that $g$ is $(\mathcal{M}^-, \mathcal{R})$-measurable. To see this let $\alpha \in \mathbf{R}$ and use that

$$\{g > \alpha\} = [\{f > \alpha\} \cap (X \setminus N)] \cup [\{g > \alpha\} \cap N] .$$

Now if we define

$$A = \{f > \alpha\} \cap (X \setminus N)$$

the set $A \in \mathcal{M}$ and

$$A \subseteq \{g > \alpha\} \subseteq A \cup N.$$

Accordingly from this $\{g > \alpha\} \in \mathcal{M}^-$ and $g$ is $(\mathcal{M}^-, \mathcal{R})$-measurable since $\alpha$ is an arbitrary real number.

Next suppose $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}_+$, is a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions and $f : X \to \mathbf{R}$ a function. Recall if

$$\lim_{n\to\infty} f_n(x) = f(x), \text{ all } x \in X$$

then $f$ is $(\mathcal{M}, \mathcal{R})$-measurable since

$$\{f > \alpha\} = \cup_{k,l\in\mathbf{N}_+} \cap_{n\geq k} \left\{f_n > \alpha + l^{-1}\right\}, \text{ all } \alpha \in \mathbf{R}.$$

If we only assume that

$$\lim_{n\to\infty} f_n(x) = f(x), \text{ a.e. } [\mu]$$

then $f$ need not be $(\mathcal{M}, \mathcal{R})$-measurable but $f$ is $(\mathcal{M}^-, \mathcal{R})$-measurable. To see this suppose $N \in \mathcal{Z}_\mu$ and

$$\lim_{n\to\infty} f_n(x) = f(x), \text{ all } x \in X \setminus N.$$

Then

$$\lim_{n\to\infty} \chi_{X\setminus N}(x) f_n(x) = \chi_{X\setminus N}(x) f(x)$$

and it follows that the function $\chi_{X\setminus N}f$ is $(\mathcal{M},\mathcal{R})$-measurable. Since $f = \chi_{X\setminus N}f$ a.e. $[\mu]$ it follows that $f$ is $(\mathcal{M}^-,\mathcal{R})$-measurable. The next example shows that $f$ need not be $(\mathcal{M},\mathcal{R})$-measurable.

**Example 2.2.2.** Let $X = \{0,1,2\}$, $\mathcal{M} = \{\phi, \{0\}, \{1,2\}, X\}$, and $\mu(A) = \chi_A(0)$, $A \in \mathcal{M}$. Set $f_n = \chi_{\{1,2\}}$, $n \in \mathbf{N}_+$, and $f(x) = x$, $x \in X$. Then each $f_n$ is $(\mathcal{M},\mathcal{R})$-measurable and

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ a.e. } [\mu]$$

since

$$\left\{x \in X;\ \lim_{n\to\infty} f_n(x) = f(x)\right\} = \{0,1\}$$

and $N = \{1,2\}$ is a $\mu$-null set. The function $f$ is not $(\mathcal{M},\mathcal{R})$-measurable.

Suppose $f,g \in \mathcal{L}^1(\mu)$. The functions $f$ and $g$ are equal almost everywhere with respect to $\mu$ if and only if $\{f \neq g\} \in \mathcal{Z}_\mu$. This is easily seen to be an equivalence relation and the set of all equivalence classes is denoted by $L^1(\mu)$. Moreover, if $f = g$ a.e. $[\mu]$, then

$$\int_X f d\mu = \int_X g d\mu$$

since

$$\int_X f d\mu = \int_{\{f=g\}} f d\mu + \int_{\{f\neq g\}} f d\mu = \int_{\{f=g\}} f d\mu = \int_{\{f=g\}} g d\mu$$

and, in a similar way,

$$\int_X g d\mu = \int_{\{f=g\}} g d\mu.$$

Below we consider the elements of $L^1(\mu)$ members of $\mathcal{L}^1(\mu)$ and two members of $L^1(\mu)$ are identified if they are equal a.e. $[\mu]$. From this convention it is straight-forward to define $f+g$ and $\alpha f$ for all $f,g \in L^1(\mu)$ and $\alpha \in \mathbf{R}$. Moreover, we get

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu \text{ if } f,g \in L^1(\mu)$$

and

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu \text{ if } f \in L^1(\mu) \text{ and } \alpha \in \mathbf{R}.$$

Next we give two theorems where exceptional null sets enter. The first one is a mild variant of Theorem 2.2.1 and needs no proof.

**Theorem 2.2.2.** *Suppose $(X, \mathcal{M}, \mu)$ is a positive complete measure space and let $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}_+$, be measurable functions such that*

$$\sup_{n \in \mathbf{N}_+} \mid f_n(x) \mid \leq g(x) \text{ a.e. } [\mu]$$

*where $g \in L^1(\mu)$. Moreover, suppose $f : X \to \mathbf{R}$ is a function and*

$$f(x) = \lim_{n \to \infty} f_n(x) \text{ a.e. } [\mu].$$

*Then, $f \in L^1(\mu)$,*

$$\lim_{n \to \infty} \int_X \mid f_n - f \mid d\mu = 0$$

*and*

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

**Theorem 2.2.3.** *Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space.*

(a) *If $f : (X, \mathcal{M}^-) \to ([0, \infty], \mathcal{R}_{0,\infty})$ is measurable there exists a measurable function $g : (X, \mathcal{M}) \to ([0, \infty], \mathcal{R}_{0,\infty})$ such that $f = g$ a.e. $[\mu]$.*

(b) *If $f : (X, \mathcal{M}^-) \to (\mathbf{R}, \mathcal{R})$ is measurable there exists a measurable function $g : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ such that $f = g$ a.e. $[\mu]$.*

PROOF. Since $f = f^+ - f^-$ it is enough to prove Part (a). There exist simple $\mathcal{M}^-$-measurable functions $\varphi_n$ , $n \in \mathbf{N}_+$, such that $0 \leq \varphi_n \uparrow f$. For each fixed $n$ suppose $\alpha_{1n}, ..., \alpha_{k_n n}$ are the distinct values of $\varphi_n$ and choose for each fixed $i = 1, ..., k_n$ a set $A_{in} \subseteq \varphi_n^{-1}(\{\alpha_{in}\})$ such that $A_{in} \in \mathcal{M}$ and $\varphi_n^{-1}(\alpha_{in}) \setminus A_{in} \in \mathcal{Z}_{\bar{\mu}}$. Set

$$\psi_n = \Sigma_{i=1}^{k_n} \alpha_{in} \chi_{A_{in}}.$$

Clearly $\psi_n(x) \uparrow f(x)$ if $x \in E =_{def} \cap_{n=1}^{\infty}(\cup_{i=1}^{k_n} A_{in})$ and $\mu(X \setminus E) = 0$. We now define $g(x) = f(x),\;$ if $x \in E$, and $g(x) = 0$ if $x \in X \setminus E$. The theorem is proved.

**Exercises**

1. Suppose $f$ and $g$ are real-valued measurable functions. Prove that $f^2$ and $fg$ are measurable functions.

2. Suppose $f \in L^1(\mu)$. Prove that

$$\lim_{\alpha\to\infty} \int_{|f|\geq\alpha} \mid f \mid d\mu = 0.$$

(Here $\int_{|f|\geq\alpha}$ means $\int_{\{|f|\geq\alpha\}}$.)

3. Suppose $f \in L^1(\mu)$. Prove that to each $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\int_E \mid f \mid d\mu < \varepsilon$$

whenever $\mu(E) < \delta$.

4. Let $(f_n)_{n=1}^{\infty}$ be a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions. Prove that the set of all $x \in \mathbf{R}$ such that the sequence $(f_n(x))_{n=1}^{\infty}$ converges to a real limit belongs to $\mathcal{M}$.

5. Let $(X, \mathcal{M}, \mathcal{R})$ be a positive measure space such that $\mu(A) = 0$ or $\infty$ for every $A \in \mathcal{M}$. Show that $f \in L^1(\mu)$ if and only if $f(x) = 0$ a.e. $[\mu]$.

6. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose $f$ and $g$ are non-negative measurable functions such that

$$\int_A f d\mu = \int_A g d\mu, \text{ all } A \in \mathcal{M}.$$

(a) Prove that $f = g$ a.e. $[\mu]$ if $\mu$ is $\sigma$-finite.
(b) Prove that the conclusion in Part (a) may fail if $\mu$ is not $\sigma$-finite.

7. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose the functions $f_n : X \to \mathbf{R}$, $n = 1, 2, ...,$ are measurable. Show that there is a sequence $(\alpha_n)_{n=1}^{\infty}$ of positive real numbers such that

$$\lim_{n\to\infty} \alpha_n f_n = 0 \text{ a.e. } [\mu]\,.$$

8. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and let $f_n : X \to \mathbf{R}$, $n = 1, 2, ...,$ be a sequence in $L^1(\mu)$ which converges to $f$ a.e. $[\mu]$ as $n \to \infty$. Suppose $f \in L^1(\mu)$ and

$$\lim_{n\to\infty} \int_X \mid f_n \mid d\mu = \int_X \mid f \mid d\mu.$$

Show that

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0.$$

9. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose $f \in L^1(\mu)$ is a bounded function such that

$$\int_X f^2 d\mu = \int_X f^3 d\mu = \int_X f^4 d\mu.$$

Prove that $f = \chi_A$ for an appropriate $A \in \mathcal{M}$.

10. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and $f : X \to \mathbf{R}$ a measurable function. Prove that $f \in L^1(\mu)$ if and only if

$$\Sigma_{k=1}^{\infty} \mu(\mid f \mid \geq k) < \infty.$$

11. Suppose $f \in L^1(m)$. Prove that the series $\Sigma_{k=-\infty}^{\infty} f(x+k)$ converges for $m$-almost all $x$.

12. a) Suppose $f : \mathbf{R} \to [0, \infty[$ is Lebesgue measurable and $\int_{\mathbf{R}} f dm < \infty$. Prove that

$$\lim_{\alpha \to \infty} \alpha m(f \geq \alpha) = 0.$$

b) Find a Lebesgue measurable function $f : \mathbf{R} \to [0, \infty[$ such that $f \notin L^1(m)$, $m(f > 0) < \infty$, and

$$\lim_{\alpha \to \infty} \alpha m(f \geq \alpha) = 0.$$

13. (a) Suppose $\mathcal{M}$ is an $\sigma$-algebra of subsets of $X$ and $\mu$ a positive measure on $\mathcal{M}$ with $\mu(X) < \infty$. Let $A_1, ..., A_n \in \mathcal{M}$. Show that

$$\chi_{A_1 \cup A_2 \cup ... \cup A_n} = 1 - (1 - \chi_{A_1}) \cdot ... \cdot (1 - \chi_{A_n})$$

and conclude that

$$\mu(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{1 \leq i \leq n} \mu(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mu(A_{i_1} \cap A_{i_2})$$

$$+ \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mu(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - ... + (-1)^{n+1} \mu(A_1 \cap ... \cap A_n).$$

(b) Let $X$ be the set of all permutations (bijections) $x : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$ and let $\mu = \frac{1}{n!} c_X$. A random variable $\xi : \Omega \to X$ has the uniform distribution in $X$ or, stated otherwise, the image measure $P_\xi$ equals $\mu$. Find the probability that $\xi$ has a fixed point, that is find

$$P\left[\xi(i) = i \text{ for some } i \in \{1, 2, ..., n\}\right].$$

(Hint: Set $A_i = \{x \in X;\ x(i) = i\}$, $i = 1, ..., n$, and note that the probability in question equals $\mu(A_1 \cup A_2 \cup ... \cup A_n)$.)

14. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $f : X \to \mathbf{R}$ an $(\mathcal{M}, \mathcal{R})$-measurable function. Moreover, for each $t > 1$, let

$$a(t) = \sum_{n=-\infty}^{\infty} t^n \mu(t^n \leq\mid f \mid< t^{n+1}).$$

Show that

$$\lim_{t \to 1^+} a(t) = \int_X \mid f \mid d\mu.$$

15. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $f_n{:}X \to \mathbf{R}$, $n \in \mathbf{N}_+$, a sequence of measurable functions such that

$$\limsup_{n \to \infty} n^2 \mu(\mid f_n \mid \geq n^{-2}) < \infty.$$

Prove that the series $\sum_{n=1}^{\infty} f_n(x)$ converges for $\mu$-almost all $x \in X$.

16. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $f{:}X \to \mathbf{R}$ a measurable function. Furthermore, suppose there are strictly positive constants $B$ and $C$ such that

$$\int_X e^{af} d\mu \leq Be^{\frac{a^2 C}{2}} \text{ if } a \in \mathbf{R}.$$

Prove that

$$\mu(\mid f \mid \geq t) \leq 2Be^{-\frac{t^2}{2C}} \text{ if } t > 0.$$

## 2.3 Comparison of Riemann and Lebesgue Integrals

In this section we will show that the Lebesgue integral is a natural generalization of the Riemann integral. For short, the discussion is restricted to a closed and bounded interval.

Let $[a, b]$ be a closed and bounded interval and suppose $f : [a, b] \to \mathbf{R}$ is a bounded function. For any partition

$$\Delta : a = x_0 < x_1 < \ldots < x_n = b$$

of $[a, b]$ define

$$S_\Delta f = \Sigma_{i=1}^n (\sup_{]x_{i-1}, x_i]} f)(x_i - x_{i-1})$$

and

$$s_\Delta f = \Sigma_{i=1}^n (\inf_{]x_{i-1}, x_i]} f)(x_i - x_{i-1}).$$

The function $f$ is Riemann integrable if

$$\inf_{\Delta} S_{\Delta} f = \sup_{\Delta} s_{\Delta} f$$

and the Riemann integral $\int_a^b f(x)dx$ is, by definition, equal to this common value.

Below an $((\mathcal{R}^-)_{[a,b]}, \mathcal{R})$-measurable function is simply called Lebesgue measurable. Furthermore, we write $m$ instead of $m_{|[a,b]}$.

**Theorem 2.3.1.** *A bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable if and only if the set of discontinuity points of $f$ is a Lebesgue null set. Moreover, if the set of discontinuity points of $f$ is a Lebesgue null set, then $f$* is *Lebesgue measurable and*

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

PROOF. A partition $\Delta' : a = x'_0 < x'_1 < ... < x'_{n'} = b$ is a refinement of a partition $\Delta : a = x_0 < x_1 < ... < x_n = b$ if each $x_k$ is equal to some $x'_l$ and in this case we write $\Delta \prec \Delta'$. The definitions give $S_{\Delta} f \geq S_{\Delta'} f$ and $s_{\Delta} f \leq s_{\Delta'} f$ if $\Delta \prec \Delta'$. We define, $\text{mesh}(\Delta) = \max_{1 \leq i \leq n}(x_i - x_{i-1})$.

First suppose $f$ is Riemann integrable. For each partition $\Delta$ let

$$G_{\Delta} = f(a)\chi_{\{a\}} + \Sigma_{i=1}^n (\sup_{]x_{i-1}, x_i]} f)\chi_{]x_{i-1}, x_i]}$$

and

$$g_{\Delta} = f(a)\chi_{\{a\}} + \Sigma_{i=1}^n (\inf_{]x_{i-1}, x_i]} f)\chi_{]x_{i-1}, x_i]}$$

and note that

$$\int_{[a,b]} G_{\Delta} dm = S_{\Delta} f$$

and

$$\int_{[a,b]} g_{\Delta} dm = s_{\Delta} f.$$

Suppose $\Delta_k$, $k = 1, 2, ...$, is a sequence of partitions such that $\Delta_k \prec \Delta_{k+1}$,

$$S_{\Delta_k} f \downarrow \int_a^b f(x)dx$$

and

$$s_{\Delta_k} f \uparrow \int_a^b f(x)dx$$

as $k \to \infty$. Let $G = \lim_{k\to\infty} G_{\Delta_k}$ and $g = \lim_{k\to\infty} g_{\Delta_k}$. Then $G$ and $g$ are $(\mathcal{R}_{[a,b]}, \mathcal{R})$-measurable, $g \leq f \leq G$, and by dominated convergence

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

But then

$$\int_{[a,b]} (G - g)dm = 0$$

so that $G = g$ a.e. $[m]$ and therefore $G = f$ a.e. $[m]$. In particular, $f$ is Lebesgue measurable and

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

Set

$$N = \{x;\ g(x) < f(x) \text{ or } f(x) < G(x)\}.$$

We proved above that $m(N) = 0$. Let $M$ be the union of all those points which belong to some partition $\Delta_k$. Clearly, $m(M) = 0$ since $M$ is denumerable. We claim that $f$ is continuous off $N \cup M$. If $f$ is not continuous at a point $c \notin N \cup M$, there is an $\varepsilon > 0$ and a sequence $(c_n)_{n=1}^{\infty}$ converging to $c$ such that

$$\mid f(c_n) - f(c) \mid \geq \varepsilon \text{ all } n.$$

Since $c \notin M$, $c$ is an interior point to exactly one interval of each partition $\Delta_k$ and we get

$$G_{\Delta_k}(c) - g_{\Delta_k}(c) \geq \varepsilon$$

and in the limit

$$G(c) - g(c) \geq \varepsilon.$$

But then $c \in N$ which is a contradiction.

Conversely, suppose the set of discontinuity points of $f$ is a Lebesgue null set and let $(\Delta_k)_{k=1}^{\infty}$ is an arbitrary sequence of partitions of $[a, b]$ such that $\Delta_k \prec \Delta_{k+1}$ and $\mathrm{mesh}(\Delta_k) \to 0$ as $k \to \infty$. By assumption,

$$\lim_{k\to\infty} G_{\Delta_k}(x) = \lim_{k\to\infty} g_{\Delta_k}(x) = f(x)$$

at each point $x$ of continuity of $f$. Therefore $f$ is Lebesgue measurable and dominated convergence yields

$$\lim_{k\to\infty} \int_{[a,b]} G_{\Delta_k} dm = \int_{[a,b]} f dm$$

and

$$\lim_{k\to\infty} \int_{[a,b]} g_{\Delta_k} dm = \int_{[a,b]} f dm.$$

Thus $f$ is Riemann integrable and

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

In the following we sometimes write

$$\int_A f(x)dx \qquad (A \in \mathcal{R}^-)$$

instead of

$$\int_A f dm \qquad (A \in \mathcal{R}^-).$$

In a similar way we often prefer to write

$$\int_A f(x)dx \qquad (A \in \mathcal{R}_n^-)$$

instead of

$$\int_A f dm_n \qquad (A \in \mathcal{R}_n^-).$$

Furthermore, $\int_a^b f dm$ means $\int_{[a,b]} f dm$. Here, however, a warning is motivated. It is simple to find a real-valued function $f$ on $[0, \infty[$, which is bounded on each bounded subinterval of $[0, \infty[$, such that the generalized Riemann integral

$$\int_0^\infty f(x)dx$$

is convergent, that is

$$\lim_{b\to\infty} \int_0^b f(x)dx$$

exists and the limit is a real number, while the Riemann integral

$$\int_0^\infty \mid f(x) \mid dx$$

is divergent (take e.g. $f(x) = \frac{\sin x}{x}$). In this case the function $f$ does not belong to $\mathcal{L}^1$ with respect to Lebesgue measure on $[0, \infty[$ since

$$\int_{[0,\infty[} \mid f \mid dm = \lim_{b\to\infty} \int_0^b \mid f(x) \mid dx = \infty.$$

**Example 2.3.1.** To compute

$$\lim_{n\to\infty} \int_0^n \frac{(1-\frac{x}{n})^n}{\sqrt{x}} dx$$

suppose $n \in \mathbf{N}_+$ and use the inequality $1 + t \leq e^t$, $t \in \mathbf{R}$, to get

$$\chi_{[0,n]}(x)(1 - \frac{x}{n})^n \leq e^{-x} \text{ if } x \geq 0.$$

From this

$$f_n(x) =_{def} \chi_{[0,n]}(x)\frac{(1-\frac{x}{n})^n}{\sqrt{x}} \leq \frac{e^{-x}}{\sqrt{x}},\ x \geq 0$$

and, in addition,

$$\lim_{n\to\infty} f_n(x) = \frac{e^{-x}}{\sqrt{x}}.$$

Here $\frac{e^{-x}}{\sqrt{x}} \in L^1(m_1$ on $[0, \infty[)$ since $\frac{e^{-x}}{\sqrt{x}} \geq 0$ and

$$\int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = 2\int_0^\infty e^{-x^2} dx = \sqrt{\pi}.$$

Moreover $f_n \geq 0$ for every $n \in \mathbf{N}_+$ and by using dominated convergence we get

$$\lim_{n\to\infty}\int_0^n \frac{(1-\frac{x}{n})^n}{\sqrt{x}}dx = \lim_{n\to\infty}\int_0^\infty f_n(x)dx =$$

$$\int_0^\infty \lim_{n\to\infty} f_n(x)dx = \int_0^\infty \frac{e^{-x}}{\sqrt{x}}dx = \sqrt{\pi}.$$

**Exercises**

1. Let $f_n : [0,1] \to [0,1]$, $n \in \mathbf{N}$, be a sequence of Riemann integrable functions such that

$$\lim_{n\to\infty} f_n(x) \text{ exists } = f(x) \text{ all } x \in [0,1]\,.$$

Show by giving an example that $f$ need not be Riemann integrable.

2. Suppose $f_n(x) = n^2 \mid x \mid e^{-n|x|}$, $x \in \mathbf{R}$, $n \in \mathbf{N}_+$. Compute $\lim_{n\to\infty} f_n$ and $\lim_{n\to\infty}\int_{\mathbf{R}} f_n dm$.

3. Compute the following limits and justify the calculations:

(a)

$$\lim_{n\to\infty}\int_0^\infty \frac{\sin(e^x)}{1+nx^2}dx.$$

(b)

$$\lim_{n\to\infty}\int_0^n (1+\frac{x}{n})^{-n}\cos x dx.$$

(c)

$$\lim_{n\to\infty}\int_0^n (1+\frac{x}{n})^n e^{-2x}dx.$$

(d)

$$\lim_{n\to\infty}\int_0^\infty (1+\frac{x}{n})^n \exp(-(1+\frac{x}{n})^n)dx.$$

(e)

$$\lim_{n\to\infty} n\int_0^\infty \frac{\sin(\frac{x}{n})}{x(1+x^2)}dx.$$

(f)

$$\lim_{n\to\infty}\int_0^n (1-\frac{x}{n})^n \frac{1+nx}{n+x}\cos x dx$$

(g)

$$\lim_{n\to\infty}\int_0^\infty (1+\frac{x}{n})^{n^2} e^{-nx} dx.$$

(h)

$$\lim_{n\to\infty}\int_0^1 \frac{1+nx^2}{(1+x^2)^n} dx.$$

(i)

$$\lim_{n\to\infty}\sqrt{n}\int_{-1}^1 (1-t^2)^n(1+\sqrt{n}\mid \sin t\mid)dt.$$

4. Let $(r_n)_{n=1}^\infty$ be an enumeration of $\mathbf{Q}$ and define

$$f(x) = \Sigma_{n=1}^\infty 2^{-n}\varphi(x-r_n)$$

where $\varphi(x) = x^{-\frac{1}{2}}$ if $0 < x < 1$ and $\varphi(x) = 0$ if $x \leq 0$ or $x \geq 1$. Show that

a)

$$\int_{-\infty}^\infty f(x)dx = 2.$$

b)

$$\int_a^b f^2(x)dx = \infty \text{ if } a < b.$$

c)

$$f < \infty \text{ a.s. } [m].$$

d)

$$\sup_{a<x<b} f(x) = +\infty \text{ if } a < b.$$

5. Let $n \in \mathbf{N}_+$ and define $f_n(x) = e^x(1-\frac{x^2}{2n})^n$, $x \in \mathbf{R}$. Compute

$$\lim_{n\to\infty}\int_{-\sqrt{2n}}^{\sqrt{2n}} f_n(x)dx.$$

6. Suppose $p \in \mathbf{N}_+$ and define $f_n(x) = n^p x^{p-1}(1-x)^n$, $0 \le x \le 1$, for every $n \in \mathbf{N}_+$. Show that

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = (p-1)!.$$

7. Suppose $f{:}[0,1] \to \mathbf{R}$ is a continuous function. Find

$$\lim_{n\to\infty} n \int_0^1 f(x) e^{-(n \min(x,1-x))^2} dx.$$

## 2.4. Expectation

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\xi : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ a random variable. Recall that the probability law $\mu$ of $\xi$ is given by the image measure $P_\xi$. By definition,

$$\int_S \chi_B d\mu = \int_\Omega \chi_B(\xi) dP$$

for every $B \in \mathcal{S}$, and, hence

$$\int_S \varphi d\mu = \int_\Omega \varphi(\xi) dP$$

for each simple $\mathcal{S}$-measurable function $\varphi$ on $S$ (we sometimes write $f \circ g = f(g)$). By monotone convergence, we get

$$\int_S f d\mu = \int_\Omega f(\xi) dP$$

for every measurable $f : S \to [0, \infty]$. Thus if $f : S \to \mathbf{R}$ is measurable, $f \in L^1(\mu)$ if and only if $f(\xi) \in L^1(P)$ and in this case

$$\int_S f d\mu = \int_\Omega f(\xi) dP.$$

In the special case when $\xi$ is real-valued and $\xi \in L^1(P)$,

$$\int_{\mathbf{R}} x d\mu(x) = \int_\Omega \xi dP.$$

The integral in the right-hand side is called the expectation of $\xi$ and is denoted by $E[\xi]$.

# CHAPTER 3

# Further Construction Methods of Measures

## Introduction

In the first section of this chapter we collect some basic results on metric spaces, which every mathematician must know about. Section 3.2 gives a version of the Riesz Representation Theorem, which leads to another and perhaps simpler approach to Lebesgue measure than the Carathéodory Theorem. A reader can skip Section 3.2 without losing the continuity in this paper. The chapter also treats so called product measures and Stieltjes integrals.

## 3.1. Metric Spaces

The construction of our most important measures requires topological concepts. For our purpose it will be enough to restrict ourselves to so called metric spaces.

A metric $d$ on a set $X$ is a mapping $d : X \times X \to [0, \infty[$ such that

(a) $d(x, y) = 0$ if and only if $x = y$
(b) $d(x, y) = d(y, x)$ (symmetry)
(c) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Here recall, if $A_1, ..., A_n$ are sets,

$$A_1 \times ... \times A_n = \{(x_1, ..., x_n);\ x_i \in A_i \text{ for all } i = 1, ..., n\}$$

A set $X$ equipped with a metric $d$ is called a metric space. Sometimes we write $X = (X, d)$ to emphasize the metric $d$. If $E$ is a subset of the metric space $(X, d)$, the function $d_{|E \times E}(x, y) = d(x, y)$, if $x, y \in E$, is a metric on $E$. Thus $(E, d_{|E \times E})$ is a metric space.

The function $\varphi(t) = \min(1, t)$, $t \geq 0$, satisfies the inequality

$$\varphi(s + t) \leq \varphi(s) + \varphi(t).$$

Therefore, if $d$ is a metric on $X$, $\min(1, d)$ is a metric on $X$. The metric $\min(1, d)$ is a bounded metric.

The set $\mathbf{R}$ equipped with the metric $d_1(x, y) =\mid x - y \mid$ is a metric space. More generally, $\mathbf{R}^n$ equipped with the metric

$$d_n(x, y) = d_n((x_1, ..., x_n), (y_1, ..., y_n)) = \max_{1 \leq k \leq n} \mid x_k - y_k \mid$$

is a metric space. If not otherwise stated, it will always be assumed that $\mathbf{R}^n$ is equipped with this metric.

Let $C\,[0, T]$ denote the vector space of all real-valued continuous functions on the interval $[0, T]$, where $T > 0$. Then

$$d_\infty(x, y) = \max_{0 \leq t \leq T} \mid x(t) - y(t) \mid$$

is a metric on $C\,[0, T]$.

If $(X_k, e_k)$, $k = 1, ..., n$, are metric spaces,

$$d(x, y) = \max_{1 \leq k \leq n} e_k(x_k, y_k),\ x = (x_1, ..., x_n)\ , y = (y_1, ..., y_n)$$

is a metric on $X_1 \times ... \times X_n$. The metric $d$ is called the product metric on $X_1 \times ... \times X_n$.

If $X = (X, d)$ is a metric space and $x \in X$ and $r > 0$, the open ball with centre at $x$ and radius $r$ is the set $B(x, r) = \{y \in X; d(y, x) < r\}$. If $E \subseteq X$ and $E$ is contained in an appropriate open ball in $X$ it is said to be bounded. The diameter of $E$ is, by definition,

$$\text{diam } E = \sup_{x, y \in E} d(x, y)$$

and it follows that $E$ is bounded if and only if $\text{diam } E < \infty$. A subset of $X$ which is a union of open balls in $X$ is called open. In particular, an open ball is an open set. The empty set is open since the union of an empty family of sets is empty. An arbitrary union of open sets is open. The class of all open subsets of $X$ is called the topology of $X$. The metrics $d$ and $\min(1, d)$ determine the same topology. A subset $E$ of $X$ is said to be closed if its complement $E^c$ relative to $X$ is open. An intersection of closed subsets of $X$ is closed. If $E \subseteq X$, $E^\circ$ denotes the largest open set contained in $E$ and $E^-$ (or $\bar{E}$) the smallest closed set containing $E$. $E^\circ$ is the interior of $E$ and $E^-$ its closure. The $\sigma$-algebra generated by the open sets in $X$ is called the Borel $\sigma$-algebra in $X$ and is denoted by $\mathcal{B}(X)$. A positive measure on $\mathcal{B}(X)$ is called a positive Borel measure.

A sequence $(x_n)_{n=1}^\infty$ in $X$ converges to $x \in X$ if

$$\lim_{n\to\infty} d(x_n, x) = 0.$$

If, in addition, the sequence $(x_n)_{n=1}^\infty$ converges to $y \in X$, the inequalities

$$0 \leq d(x, y) \leq d(x_n, x) + d(x_n, y)$$

imply that $y = x$ and the limit point $x$ is unique.

If $E \subseteq X$ and $x \in X$, the following properties are equivalent:

(i) $x \in E^-$.
(ii) $B(x, r) \cap E \neq \phi$, all $r > 0$.
(iii) There is a sequence $(x_n)_{n=1}^\infty$ in $E$ which converges to $x$.

If $B(x, r) \cap E = \phi$, then $B(x, r)^c$ is a closed set containing $E$ but not $x$. Thus $x \notin E^-$. This proves that (i)$\Rightarrow$(ii). Conversely, if $x \notin E^-$, since $\bar{E}^c$ is open there exists an open ball $B(y, s)$ such that $x \in B(y, s) \subseteq \bar{E}^c \subseteq E^c$. Now choose $r = s - d(x, y) > 0$ so that $B(x, r) \subseteq B(y, s)$. Then $B(x, r) \cap E = \phi$. This proves (ii)$\Rightarrow$(i).

If (ii) holds choose for each $n \in \mathbf{N}_+$ a point $x_n \in E$ with $d(x_n, x) < \frac{1}{n}$ and (iii) follows. If there exists an $r > 0$ such that $B(x, r) \cap E = \phi$, then (iii) cannot hold. Thus (iii)$\Rightarrow$(ii).

If $E \subseteq X$, the set $E^- \setminus E^\circ$ is called the boundary of $E$ and is denoted by $\partial E$.

A set $A \subseteq X$ is said to be dense in $X$ if $A^- = X$. The metric space $X$ is called separable if there is an at most denumerable dense subset of $X$. For example, $\mathbf{Q}^n$ is a dense subset of $\mathbf{R}^n$. The space $\mathbf{R}^n$ is separable.

**Theorem 3.1.1.** $\mathcal{B}(\mathbf{R}^n) = \mathcal{R}_n$.

PROOF. The $\sigma$-algebra $\mathcal{R}_n$ is generated by the open $n$-cells in $\mathbf{R}^n$ and an open $n$-cell is an open subset of $\mathbf{R}^n$. Hence $\mathcal{R}_n \subseteq \mathcal{B}(\mathbf{R}^n)$. Let $U$ be an open subset in $\mathbf{R}^n$ and note that an open ball in $\mathbf{R}^n = (\mathbf{R}^n, d_n)$ is an open $n$-cell. If $x \in U$ there exist an $a \in \mathbf{Q}^n \cap U$ and a rational number $r > 0$ such that $x \in B(a, r) \subseteq U$. Thus $U$ is an at most denumerable union of open $n$-cells and it follows that $U \in \mathcal{R}_n$. Thus $\mathcal{B}(\mathbf{R}^n) \subseteq \mathcal{R}_n$ and the theorem is proved.

Let $X = (X, d)$ and $Y = (Y, e)$ be two metric spaces. A mapping $f : X \to Y$ (or $f : (X, d) \to (Y, e)$ to emphasize the underlying metrics) is said to be continuous at the point $a \in X$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B(a, \delta) \Rightarrow f(x) \in B(f(a), \varepsilon).$$

Equivalently this means that for any sequence $(x_n)_{n=1}^{\infty}$ in $X$ which converges to $a$ in $X$, the sequence $(f(x_n))_{n=1}^{\infty}$ converges to $f(a)$ in $Y$. If $f$ is continuous at each point of $X$, the mapping $f$ is called continuous. Stated otherwise this means that

$$f^{-1}(V) \text{ is open if } V \text{ is open}$$

or

$$f^{-1}(F) \text{ is closed if } F \text{ is closed.}$$

The mapping $f$ is said to be Borel measurable if

$$f^{-1}(B) \in \mathcal{B}(X) \text{ if } B \in \mathcal{B}(Y)$$

or, what amounts to the same thing,

$$f^{-1}(V) \in \mathcal{B}(X) \text{ if } V \text{ is open.}$$

A Borel measurable function is sometimes called a Borel function. A continuous function is a Borel function.

**Example 3.1.1.** Let $f : (\mathbf{R}, d_1) \to (\mathbf{R}, d_1)$ be a continuous strictly increasing function and set $\rho(x, y) = \mid f(x) - f(y) \mid$, $x, y \in \mathbf{R}$. Then $\rho$ is a metric on $\mathbf{R}$.

Define $j(x) = x$, $x \in \mathbf{R}$. The mapping $j : (\mathbf{R},d_1) \rightarrow (\mathbf{R},\rho)$ is continuous. We claim that the map $j : (\mathbf{R},\rho) \rightarrow (\mathbf{R},d_1)$ is continuous. To see this, let $a \in \mathbf{R}$ and suppose the sequence $(x_n)_{n=1}^{\infty}$ converges to $a$ in the metric space $(\mathbf{R},\rho)$, that is $\mid f(x_n) - f(a) \mid \rightarrow 0$ as $n \rightarrow \infty$. Let $\varepsilon > 0$. Then

$$f(x_n) - f(a) \geq f(a+\varepsilon) - f(a) > 0 \text{ if } x_n \geq a + \varepsilon$$

and

$$f(a) - f(x_n) \geq f(a) - f(a-\varepsilon) > 0 \text{ if } x_n \leq a - \varepsilon.$$

Thus $x_n \in \, ]a-\varepsilon, a+\varepsilon[$ if $n$ is sufficiently large. This proves that he map $j : (\mathbf{R},\rho) \rightarrow (\mathbf{R},d_1)$ is continuous.

The metrics $d_1$ and $\rho$ determine the same topology and Borel subsets of $\mathbf{R}$.

A mapping $f : (X,d) \rightarrow (Y,e)$ is said to be uniformly continuous if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $e(f(x),f(y)) < \varepsilon$ as soon as $d(x,y) < \delta$.

If $x \in X$ and $E$, $F \subseteq X$, let

$$d(x,E) = \inf_{u \in E} d(x,u)$$

be the distance from $x$ to $E$ and let

$$d(E,F) = \inf_{u \in E, v \in F} d(u,v)$$

be the distance between $E$ and $F$. Note that $d(x,E) = 0$ if and only if $x \in \bar{E}$.

If $x,y \in X$ and $u \in E$,

$$d(x,u) \leq d(x,y) + d(y,u)$$

and, hence

$$d(x,E) \leq d(x,y) + d(y,u)$$

and

$$d(x,E) \leq d(x,y) + d(y,E).$$

Next suppose $E \neq \phi$. Then by interchanging the roles of $x$ and $y$, we get

$$\mid d(x,E) - d(y,E) \mid \leq d(x,y)$$

and conclude that the distance function $d(x, E)$, $x \in X$, is continuous. In fact, it is uniformly continuous. If $x \in X$ and $r > 0$, the so called closed ball $\bar{B}(x, r) = \{y \in X;\ d(y, x) \leq r\}$ is a closed set since the map $y \to d(y, x)$, $y \in X$, is continuous.

If $F \subseteq X$ is closed and $\varepsilon > 0$, the continuous function

$$\Pi_{F,\varepsilon}^X = \max(0, 1 - \frac{1}{\varepsilon} d(\cdot, F))$$

fulfils $0 \leq \Pi_{F,\varepsilon}^X \leq 1$ and $\Pi_{F,\varepsilon}^X = 1$ on $F$. Furthermore, $\Pi_{F,\varepsilon}^X(a) > 0$ if and only if $a \in F_\varepsilon =_{def} \{x \in X;\ d(x, F) < \varepsilon\}$. Thus

$$\chi_F \leq \Pi_{F,\varepsilon}^X \leq \chi_{F_\varepsilon}.$$

Let $X = (X, d)$ be a metric space. A sequence $(x_n)_{n=1}^\infty$ in $X$ is called a Cauchy sequence if to each $\varepsilon > 0$ there exists a positive integer $p$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq p$. If a Cauchy sequence $(x_n)_{n=1}^\infty$ contains a convergent subsequence $(x_{n_k})_{k=1}^\infty$ it must be convergent. To prove this claim, suppose the subsequence $(x_{n_k})_{k=1}^\infty$ converges to a point $x \in X$. Then

$$d(x_m, x) \leq d(x_m, x_{n_k}) + d(x_{n_k}, x)$$

can be made arbitrarily small for all sufficiently large $m$ by choosing $k$ sufficiently large. Thus $(x_n)_{n=1}^\infty$ converges to $x$.

A subset $E$ of $X$ is said to be complete if every Cauchy sequence in $E$ converges to a point in $E$. If $E \subseteq X$ is closed and $X$ is complete it is clear that $E$ is complete. Conversely, if $X$ is a metric space and a subset $E$ of $X$ is complete, then $E$ is closed.

It is important to know that $\mathbf{R}$ is complete equipped with its standard metric. To see this let $(x_n)_{n=1}^\infty$ be a Cauchy sequence. There exists a positive integer such that $\mid x_n - x_m \mid < 1$ if $n, m \geq p$. Therefore

$$\mid x_n \mid \leq \mid x_n - x_p \mid + \mid x_p \mid \leq 1 + \mid x_p \mid$$

for all $n \geq p$. We have proved that the sequence $(x_n)_{n=1}^\infty$ is bounded (the reader can check that every Cauchy sequence in a metric space has this property). Now define

$$a = \sup \{x \in \mathbf{R};\ \text{there are only finitely many } n \text{ with } x_n \leq x\}.$$

The definition implies that there exists a subsequence $(x_{n_k})_{k=1}^\infty$, which converges to $a$ (since for any $r > 0$, $x_n \in B(a, r)$ for infinitely many $n$). The original sequence is therefore convergent and we conclude that $\mathbf{R}$ is complete (equipped with its standard metric $d_1$). It is simple to prove that the product of $n$ complete spaces is complete and we conclude that $\mathbf{R}^n$ is complete.

Let $E \subseteq X$. A family $(V_i)_{i\in I}$ of subsets of $X$ is said to be a cover of $E$ if $\cup_{i\in I} V_i \supseteq E$ and $E$ is said to be covered by the $V_i's$. The cover $(V_i)_{i\in I}$ is said to be an open cover if each member $V_i$ is open. The set $E$ is said to be totally bounded if, for every $\varepsilon > 0$, $E$ can be covered by finitely many open balls of radius $\varepsilon$. A subset of a totally bounded set is totally bounded.

The following definition is especially important.

**Definition 3.1.1.** A subset $E$ of a metric space $X$ is said to be compact if to every open cover $(V_i)_{i\in I}$ of $E$, there is a finite subcover of $E$, which means there is a finite subset $J$ of $I$ such that $(V_i)_{i\in J}$ is a cover of $E$.

If $K$ is closed, $K \subseteq E$, and $E$ is compact, then $K$ is compact. To see this, let $(V_i)_{i\in I}$ be an open cover of $K$. This cover, augmented by the set $X \setminus K$ is an open cover of $E$ and has a finite subcover since $E$ is compact. Noting that $K \cap (X \setminus K) = \phi$, the assertion follows.

**Theorem 3.1.2.** *The following conditions are equivalent:*

(a) *$E$ is complete and totally bounded.*

(b) *Every sequence in $E$ contains a subsequence which converges to a point of $E$.*

(c) *$E$ is compact.*

PROOF. (a)$\Rightarrow$(b). Suppose $(x_n)_{n=1}^{\infty}$ is a sequence in $E$. The set $E$ can be covered by finitely many open balls of radius $2^{-1}$ and at least one of them must contain $x_n$ for infinitely many $n \in \mathbf{N}_+$. Suppose $x_n \in B(a_1, 2^{-1})$ if $n \in N_1 \subseteq N_0 =_{def} \mathbf{N}_+$, where $N_1$ is infinite. Next $E \cap B(a_1, 2^{-1})$ can be covered by finitely many balls of radius $2^{-2}$ and at least one of them must contain $x_n$ for infinitely many $n \in N_1$. Suppose $x_n \in B(a_2, 2^{-1})$ if $n \in N_2$, where $N_2 \subseteq N_1$ is infinite. By induction, we get open balls $B(a_j, 2^{-j})$ and infinite sets $N_j \subseteq N_{j-1}$ such that $x_n \in B(a_j, 2^{-j})$ for all $n \in N_j$ and $j \geq 1$.

Let $n_1 < n_2 < ...$, where $n_k \in N_k$, $k = 1, 2, ...$ . The sequence $(x_{n_k})_{k=1}^{\infty}$ is a Cauchy sequence, and since $E$ is complete it converges to a point of $E$ .

(b)⇒(a). If $E$ is not complete there is a Cauchy sequence in $E$ with no limit in $E$. Therefore no subsequence can converge in $E$, which contradicts (b). On the other hand if $E$ is not totally bounded, there is an $\varepsilon > 0$ such that $E$ cannot be covered by finitely many balls of radius $\varepsilon$. Let $x_1 \in E$ be arbitrary. Having chosen $x_1, ..., x_{n-1}$, pick $x_n \in E \setminus \cup_{i=1}^{n-1} B(x_i, \varepsilon)$, and so on. The sequence $(x_n)_{n=1}^{\infty}$ cannot contain any convergent subsequence as $d(x_n, x_m) \geq \varepsilon$ if $n \neq m$, which contradicts (b).

{(a) and (b)} ⇒(c). Let $(V_i)_{i \in I}$ be an open cover of $E$. Since $E$ is totally bounded it is enough to show that there is an $\varepsilon > 0$ such that any open ball of radius $\varepsilon$ which intersects $E$ is contained in some $V_i$. Suppose on the contrary that for every $n \in \mathbf{N}_+$ there is an open ball $B_n$ of radius $\leq 2^{-n}$ which intersects $E$ and is contained in no $V_i$. Choose $x_n \in B_n \cap E$ and assume without loss of generality that $(x_n)_{n=1}^{\infty}$ converges to some point $x$ in $E$ by eventually going to a subsequence. Suppose $x \in V_{i_0}$ and choose $r > 0$ such that $B(x, r) \subseteq V_{i_0}$. But then $B_n \subseteq B(x, r) \subseteq V_{i_0}$ for large $n$, which contradicts the assumption on $B_n$.

(c)⇒(b). If $(x_n)_{n=1}^{\infty}$ is a sequence in $E$ with no convergent subsequence in $E$, then for every $x \in E$ there is an open ball $B(x, r_x)$ which contains $x_n$ for only finitely many $n$. Then $(B(x, r_x))_{x \in E}$ is an open cover of $E$ without a finite subcover.

**Corollary 3.1.1.** *A subset of* $\mathbf{R}^n$ *is compact if and only if it is closed and bounded.*

PROOF. Suppose $K$ is compact. If $x_n \in K$ and $x_n \notin B(0, n)$ for every $n \in \mathbf{N}_+$, the sequence $(x_n)_{n=1}^{\infty}$ cannot contain a convergent subsequence. Thus $K$ is bounded. Since $K$ is complete it is closed.

Conversely, suppose $K$ is closed and bounded. Since $\mathbf{R}^n$ is complete and $K$ is closed, $K$ is complete. We next prove that a bounded set is totally bounded. It is enough to prove that any $n$-cell in $\mathbf{R}^n$ is a union of finitely many $n$-cells $I_1 \times ... \times I_n$ where each interval $I_1, ..., I_n$ has a prescribed positive length. This is clear and the theorem is proved.

**Corollary 3.1.2.** *Suppose $f : X \to \mathbf{R}$ is continuous and $X$ compact.*
(a) *There exists an $a \in X$ such that $\max_X f = f(a)$ and a $b \in X$ such that $\min_X f = f(b)$.*
(b) *The function $f$ is uniformly continuous.*

PROOF. (a) For each $a \in X$, let $V_a = \{x \in X : f(x) < 1 + f(a)\}$. The open cover $(V_a)_{a\in K}$ of $X$ has a finite subcover and it follows that $f$ is bounded. Let $(x_n)_{n=1}^{\infty}$ be a sequence in $X$ such that $f(x_n) \to \sup_K f$ as $n \to \infty$. Since $X$ is compact there is a subsequence $(x_{n_k})_{k=1}^{\infty}$ which converges to a point $a \in X$. Thus, by the continuity of $f$, $f(x_{n_k}) \to f(a)$ as $k \to \infty$.

The existence of a minimum is proved in a similar way.

(b) If $f$ is not uniformly continuous there exist $\varepsilon > 0$ and sequences $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ such that $| f(x_n) - f(y_n) | \geq \varepsilon$ and $| x_n - y_n | < 2^{-n}$ for every $n \geq 1$. Since $X$ is compact there exists a subsequence $(x_{n_k})_{k=1}^{\infty}$ of $(x_n)_{n=1}^{\infty}$ which converges to a point $a \in X$. Clearly the sequence $(y_{n_k})_{k=1}^{\infty}$ converges to $a$ and therefore

$$| f(x_{n_k}) - f(y_{n_k}) | \leq | f(x_{n_k}) - f(a) | + | f(a) - f(y_{n_k}) | \to 0$$

as $k \to \infty$ since $f$ is continuous. But $| f(x_{n_k}) - f(y_{n_k}) | \geq \varepsilon$ and we have got a contradiction. The corollary is proved.

**Example 3.1.2.** Suppose $X = ]0, 1]$ and define $\rho_1(x, y) = d_1(x, y)$ and $\rho_2(x, y) = | \frac{1}{x} - \frac{1}{y} |$, $x, y \in X$. As in Example 3.1.1 we conclude that the metrics $\rho_1$ and $\rho_2$ determine the same topology of subsets of $X$. The space $(X, \rho_1)$

totally bounded but not complete. However, the space $(X, \rho_2)$ is not totally bounded but it is complete. To see this, let $(x_n)_{n=1}^{\infty}$be a Cauchy sequence in $(X, \rho_2)$. As a Cauchy sequence it must be bounded and therefore there exists an $\varepsilon \in ]0, 1]$ such that $x_n \in [\varepsilon, 1]$ for all $n$. But then, by Corollary 3.1.1, $(x_n)_{n=1}^{\infty}$ contains a convergent subsequence in $(X, \rho_1)$ and, accordingly from this, the same property holds in $(X, \rho_2)$. The space $(X, \rho_2)$ is not compact, since $(X, \rho_1)$ is not compact, and we conclude from Theorem 3.1.2 that the space $(X, \rho_2)$ cannot be totally bounded.

**Example 3.1.3.** Set $\hat{\mathbf{R}}$=$\mathbf{R}\cup \{-\infty, \infty\}$ and

$$\hat{d}(x, y) =\mid \arctan x - \arctan y \mid$$

if $x, y \in \hat{\mathbf{R}}$. Here

$$\arctan\infty = \frac{\pi}{2} \text{ and } \arctan -\infty = -\frac{\pi}{2}.$$

Example 3.1.1 shows that the standard metric $d_1$ and the metric $\hat{d}_{|\mathbf{R}\times\mathbf{R}}$ determine the same topology.

We next prove that the metric space $\hat{\mathbf{R}}$ is compact. To this end, consider a sequence $(x_n)_{n=1}^{\infty}$ in $\hat{\mathbf{R}}$**.** If there exists a real number $M$ such that $\mid x_n \mid\leq M$ for infinitely many $n$, the sequence $(x_n)_{n=1}^{\infty}$ contains a convergent subsequence since the interval $[-M, M]$ is compact. In the opposite case, for each positive real number $M$, either $x_n \geq M$ for infinitely many $n$ or $x_n \leq -M$ for infinitely many $n$. Suppose $x_n \geq M$ for infinitely many $n$ for every $M \in \mathbf{N}_+$. Then $\hat{d}(x_{n_k}, \infty) =\mid \arctan x_{n_k} - \frac{\pi}{2} \mid\to 0$ as $k \to \infty$ for an appropriate subsequence $(x_{n_k})_{k=1}^{\infty}$.

The space $\hat{\mathbf{R}}$= $(\hat{\mathbf{R}}, \hat{d})$ is called a two-point compactification of $\mathbf{R}$.

It is an immediate consequence of Theorem 3.1.2 that the product of finitely many compact metric spaces is compact. Thus $\hat{\mathbf{R}}^n$ equipped with the product metric is compact.

We will finish this section with several useful approximation theorems.

**Theorem 3.1.3.** *Suppose $X$ is a metric space and $\mu$ positive Borel measure in $X$. Moreover, suppose there is a sequence $(U_n)_{n=1}^{\infty}$of open subsets of $X$ such that*

$$X = \cup_{n=1}^{\infty} U_n$$

*and*

$$\mu(U_n) < \infty, \text{ all } n \in \mathbf{N}_+.$$

*Then for each $A \in \mathcal{B}(X)$ and $\varepsilon > 0$, there are a closed set $F \subseteq A$ and an open set $V \supseteq A$ such that*

$$\mu(V \setminus F) < \varepsilon.$$

*In particular, for every $A \in \mathcal{B}(X)$,*

$$\mu(A) = \inf_{\substack{V \supseteq A \\ V \text{ open}}} \mu(V)$$

*and*

$$\mu(A) = \sup_{\substack{F \subseteq A \\ F \text{ closed}}} \mu(F)$$

If $X = \mathbf{R}$ and $\mu(A) = \Sigma_{n=1}^{\infty} \delta_{\frac{1}{n}}(A)$ , $A \in \mathcal{R}$, then $\mu(\{0\}) = 0$ and $\mu(V) = \infty$ for every open set containing $\{0\}$. The hypothesis that the sets $U_n$, $n \in \mathbf{N}_+$, are open (and not merely Borel sets) is very important in Theorem 3.1.3.

PROOF. First suppose that $\mu$ is a finite positive measure.

Let $\mathcal{A}$ be the class of all Borel sets $A$ in $X$ such that for every $\varepsilon > 0$ there exist a closed $F \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. If $F$ is a closed subset of $X$ and $V_n = \left\{x; d(x, F) < \frac{1}{n}\right\}$, then $V_n$ is open and, by Theorem 1.1.2 (f), $\mu(V_n) \downarrow \mu(F)$ as $n \to \infty$. Thus $F \in \mathcal{A}$ and we conclude that $\mathcal{A}$ contains all closed subsets of $X$.

Now suppose $A \in \mathcal{A}$. We will prove that $A^c \in \mathcal{A}$. To this end, we choose $\varepsilon > 0$ and a closed set $F \subseteq A$ and an open set $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. Then $V^c \subseteq A^c \subseteq F^c$ and, moreover, $\mu(F^c \setminus V^c) < \varepsilon$ since

$$V \setminus F = F^c \setminus V^c.$$

If we note that $V^c$ is closed and $F^c$ open it follows that $A^c \in \mathcal{A}$.

Next let $(A_i)_{i=1}^{\infty}$ be a denumerable collection of members of $\mathcal{A}$. Choose $\varepsilon > 0$. By definition, for each $i \in \mathbf{N}_+$ there exist a closed $F_i \subseteq A_i$ and an open $V_i \supseteq A_i$ such that $\mu(V_i \setminus F_i) < 2^{-i}\varepsilon$. Set

$$V = \cup_{i=1}^{\infty} V_i.$$

Then

$$\mu(V \setminus (\cup_{i=1}^{\infty} F_i)) \leq \mu(\cup_{i=1}^{\infty}(V_i \setminus F_i))$$
$$\leq \Sigma_{i=1}^{\infty} \mu(V_i \setminus F_i) < \varepsilon.$$

But

$$V \setminus (\cup_{i=1}^{\infty} F_i) = \cap_{n=1}^{\infty} \{V \setminus (\cup_{i=1}^{n} F_i)\}$$

and since $\mu$ is a finite positive measure

$$\mu(V \setminus (\cup_{i=1}^{\infty} F_i)) = \lim_{n\to\infty} \mu(V \setminus (\cup_{i=1}^{n} F_i)).$$

Accordingly, from these equations

$$\mu(V \setminus (\cup_{i=1}^{n} F_i)) < \varepsilon$$

if $n$ is large enough. Since a union of open sets is open and a finite union of closed sets is closed, we conclude that $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$. This proves that $\mathcal{A}$ is a $\sigma$-algebra. Since $\mathcal{A}$ contains each closed subset of $X$, $\mathcal{A} = \mathcal{B}(X)$.

We now prove the general case. Suppose $A \in \mathcal{B}(X)$. Since $\mu^{U_n}$ is a finite positive measure the previous theorem gives us an open set $V_n \supseteq A \cap U_n$ such that $\mu^{U_n}(V_n \setminus (A \cap U_n)) < \varepsilon 2^{-n}$. By eventually replacing $V_n$ by $V_n \cap U_n$ we can assume that $V_n \subseteq U_n$. But then $\mu(V_n \setminus (A \cap U_n)) = \mu^{U_n}(V_n \setminus (A \cap U_n)) < \varepsilon 2^{-n}$.

Set $V = \cup_{n=1}^{\infty} V_n$ and note that $V$ is open. Moreover,

$$V \setminus A \subseteq \cup_{n=1}^{\infty}(V_n \setminus (A \cap U_n))$$

and we get

$$\mu(V \setminus A) \leq \Sigma_{n=1}^{\infty} \mu(V_n \setminus (A \cap U_n)) < \varepsilon.$$

By applying the result already proved to the complement $A^c$ we conclude there exists an open set $W \supseteq A^c$ such that

$$\mu(A \setminus W^c) = \mu(W \setminus A^c) < \varepsilon.$$

Thus if $F =_{def} W^c$ it follows that $F \subseteq A \subseteq V$ and $\mu(V \setminus F) < 2\varepsilon$. The theorem is proved.

If $X$ is a metric space $C(X)$ denotes the vector space of all real-valued continuous functions $f : X \to \mathbf{R}$. If $f \in C(X)$, the closure of the set of all $x$ where $f(x) \neq 0$ is called the support of $f$ and is denoted by supp$f$. The vector space of all all real-valued continuous functions $f : X \to \mathbf{R}$ with compact support is denoted by $C_c(X)$.

**Corollary 3.1.3**. *Suppose $\mu$ and $\nu$ are positive Borel measures in $\mathbf{R}^n$ such that*

$$\mu(K) < \infty \text{ and } \nu(K) < \infty$$

*for every compact subset $K$ of $\mathbf{R}^n$. If*

$$\int_{\mathbf{R}^n} f(x)d\mu(x) = \int_{\mathbf{R}^n} f(x)d\nu(x), \text{ all } f \in C_c(\mathbf{R}^n)$$

*then $\mu = \nu$.*

PROOF. Let $F$ be closed. Clearly $\mu(B(0,i)) < \infty$ and $\nu(B(0,i)) < \infty$ for every positive integer $i$. Hence, by Theorem 3.1.3 it is enough to show that $\mu(F) = \nu(F)$. Now fix a positive integer $i$ and set $K = \bar{B}(0,i) \cap F$. It is enough to show that $\mu(K) = \nu(K)$. But

$$\int_{\mathbf{R}^n} \Pi^{\mathbf{R}^n}_{K,2^{-j}}(x)d\mu(x) = \int_{\mathbf{R}^n} \Pi^{\mathbf{R}^n}_{K,2^{-j}}(x)d\nu(x)$$

for each positive integer $j$ and letting $j \to \infty$ we are done.

A metric space $X$ is called a standard space if it is separable and complete. Standard spaces have a series of very nice properties related to measure theory; an example is furnished by the following

**Theorem 3.1.4. (Ulam's Theorem)** *Let $X$ be a standard space and suppose $\mu$ is a finite positive Borel measure on $X$. Then to each $A \in \mathcal{B}(X)$ and $\varepsilon > 0$ there exist a compact $K \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus K) < \varepsilon$.*

PROOF. Let $\varepsilon > 0$. We first prove that there is a compact subset $K$ of $X$ such that $\mu(K) > \mu(X) - \varepsilon$. To this end, let $A$ be a dense denumerable subset of $X$ and let $(a_i)_{i=1}^{\infty}$ be an enumeration of $A$. Now for each positive integer $j$, $\cup_{i=1}^{\infty} B(a_i, 2^{-j}\varepsilon) = X$, and therefore there is a positive integer $n_j$ such that

$$\mu(\cup_{i=1}^{n_j} B(a_i, 2^{-j}\varepsilon)) > \mu(X) - 2^{-j}\varepsilon.$$

Set

$$F_j = \cup_{i=1}^{n_j} \bar{B}(a_i, 2^{-j}\varepsilon)$$

and

$$L = \cap_{j=1}^{\infty} F_j.$$

The set $L$ is totally bounded. Since $X$ is complete and $L$ closed, $L$ is complete. Therefore, the set $L$ is compact and, moreover

$$\mu(K) = \mu(X) - \mu(L^c) = \mu(X) - \mu(\cup_{j=1}^{\infty} F_j^c)$$

$$\geq \mu(X) - \Sigma_{j=1}^{\infty} \mu(F_j^c) = \mu(X) - \Sigma_{j=1}^{\infty}(\mu(X) - \mu(F_j))$$

$$\geq \mu(X) - \Sigma_{j=1}^{\infty} 2^{-j}\varepsilon = \mu(X) - \varepsilon.$$

Depending on Theorem 3.1.3 to each $A \in \mathcal{B}(X)$ there exists a closed $F \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. But

$$V \setminus (F \cap L) = (V \setminus F) \cup (F \setminus L)$$

and we get

$$\mu(V \setminus (F \cap L)) \leq \mu(V \setminus F) + \mu(X \setminus K) < 2\varepsilon.$$

Since the set $F \cap L$ is compact Theorem 3.1.4 is proved.

Two Borel sets in $\mathbf{R}^n$ are said to be almost disjoint if their intersection has volume measure zero.

**Theorem 3.1.5.** *Every open set $U$ in $\mathbf{R}^n$ is the union of an at most denumerable collection of mutually almost disjoint cubes.*

Before the proof observe that a cube in $\mathbf{R}^n$ is the same as a closed ball in $\mathbf{R}^n$ equipped with the metric $d_n$.

PROOF. For each, $k \in \mathbf{N}_+$, let $\mathcal{Q}_k$be the class of all cubes of side length $2^{-k}$ whose vertices have coordinates of the form $i2^{-k}$, $i \in \mathbf{Z}$. Let $F_1$ be the union of those cubes in $\mathcal{Q}_1$ which are contained in $U$. Inductively, for $k \geq 1$, let $F_k$ be the union of those cubes in $\mathcal{Q}_k$ which are contained in $U$ and whose interiors are disjoint from $\cup_{j=1}^{k-1} F_j$. Since $d(x,\mathbf{R}^n \setminus U) > 0$ for every $x \in U$ it follows that $U = \cup_{j=1}^{\infty} F_j$.

**Exercises**

1. Suppose $f : (X, \mathcal{M}) \to (\mathbf{R}^d, \mathcal{R}_d)$ and $g : (X, \mathcal{M}) \to (\mathbf{R}^n, \mathcal{R}_n)$ are measurable. Set $h(x) = (f(x), g(x)) \in \mathbf{R}^{d+n}$ if $x \in X$. Prove that $h : (X, \mathcal{M}) \to (\mathbf{R}^{d+n}, \mathcal{R}_{d+n})$ is measurable.

2. Suppose $f : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ and $g : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ are measurable. Prove that $fg$ is $(\mathcal{M}, \mathcal{R})$-measurable.

3. The function $f : \mathbf{R} \to \mathbf{R}$ is a Borel function. Set $g(x, y) = f(x)$, $(x, y) \in \mathbf{R}^2$. Prove that $g : \mathbf{R}^2 \to \mathbf{R}$ is a Borel function.

4. Suppose $f : [0, 1] \to \mathbf{R}$ is a continuous function and $g : [0, 1] \to [0, 1]$ a Borel function. Compute the limit

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

5. Suppose $X$ and $Y$ are metric spaces and $f : X \to Y$ a continuous mapping. Show that $f(E)$ is compact if $E$ is a compact subset of $X$.

6. Suppose $X$ and $Y$ are metric spaces and $f : X \to Y$ a continuous bijection. Show that the inverse mapping $f^{-1}$ is continuous if $X$ is compact.

7. Construct an open bounded subset $V$ of $\mathbf{R}$ such that $m(\partial V) > 0$.

8. The function $f : [0,1] \to \mathbf{R}$ has a continuous derivative. Prove that the set $f(K) \in \mathcal{Z}_m$ if $K = (f')^{-1}(\{0\})$.

9. Let $P$ denote the class of all Borel probability measures on $[0,1]$ and $L$ the class of all functions $f : [0,1] \to [-1,1]$ such that

$$| f(x) - f(y) | \leq | x - y |, \ x, y \in [0,1] .$$

For any $\mu, \nu \in P$, define

$$\rho(\mu, \nu) = \sup_{f \in L} \left| \int_{[0,1]} f d\mu - \int_{[0,1]} f d\nu \right| .$$

(a) Show that $(P, \rho)$ is a metric space. (b) Compute $\rho(\mu, \nu)$ if $\mu$ is linear measure on $[0,1]$ and $\nu = \frac{1}{n}\Sigma_{k=0}^{n-1}\delta_{\frac{k}{n}}$, where $n \in \mathbf{N}_+$ (linear measure on $[0,1]$ is Lebesgue measure on $[0,1]$ restricted to the Borel sets in $[0,1]$).

10. Suppose $\mu$ is a finite positive Borel measure on $\mathbf{R}^n$. (a) Let $(V_i)_{i\in I}$ be a family of open subsets of $\mathbf{R}^n$ and $V = \cup_{i\in I} V_i$. Prove that

$$\mu(V) = \sup_{\substack{i_1,...,i_k \in I \\ k \in \mathbf{N}_+}} \mu(V_{i_1} \cup ... \cup V_{i_k}).$$

(b) Let $(F_i)_{i\in I}$ be a family of closed subsets of $\mathbf{R}^n$ and $F = \cap_{i\in I} F_i$. Prove that

$$\mu(F) = \inf_{\substack{i_1,...,i_k \in I \\ k \in \mathbf{N}_+}} \mu(F_{i_1} \cap ... \cap F_{i_k}).$$

$\downarrow\downarrow\downarrow$

## 3.2. Linear Functionals and Measures

Let $X$ be a metric space. A mapping $T : C_c(X) \to \mathbf{R}$ is said to be a linear functional on $C_c(X)$ if

$$T(f+g) = Tf + Tg, \text{ all } f, g \in C_c(X)$$

and

$$T(\alpha f) = \alpha Tf, \text{ all } \alpha \in \mathbf{R},\ f \in C_c(X).$$

If in addition $Tf \geq 0$ for all $f \geq 0$, $T$ is called a positive linear functional on $C_c(X)$. In this case $Tf \leq Tg$ if $f \leq g$ since $g - f \geq 0$ and $Tg - Tf = T(g-f) \geq 0$. Note that $C_c(X) = C(X)$ if $X$ is compact.

The main result in this section is the following

**Theorem 3.2.1. (The Riesz Representation Theorem)** *Suppose $X$ is a compact metric space and let $T$ be a positive linear functional on $C(X)$. Then there exists a unique finite positive Borel measure $\mu$ in $X$ with the following properties:*

(a)

$$Tf = \int_X f d\mu,\ f \in C(X).$$

(b) *For every $E \in \mathcal{B}(X)$*

$$\mu(E) = \sup_{\substack{K \subseteq E \\ K \text{ compact}}} \mu(K).$$

(c) *For every $E \in \mathcal{B}(X)$*

$$\mu(E) = \inf_{\substack{V \supseteq E \\ V \text{ open}}} \mu(V).$$

The property (c) is a consequence of (b), since for each $E \in \mathcal{B}(X)$ and $\varepsilon > 0$ there is a compact $K \subseteq X \setminus E$ such that

$$\mu(X \setminus E) < \mu(K) + \varepsilon.$$

But then

$$\mu(X \setminus K) < \mu(E) + \varepsilon$$

and $X \setminus K$ is open and contains $E$. In a similar way, (b) follows from (c) since $X$ is compact.

The proof of the Riesz Representation Theorem depends on properties of continuous functions of independent interest. Suppose $K \subseteq X$ is compact and $V \subseteq X$ is open. If $f : X \to [0,1]$ is a continuous function such that

$$f \leq \chi_V \text{ and } \mathrm{supp} f \subseteq V$$

we write

$$f \prec V$$

and if

$$\chi_K \leq f \leq \chi_V \text{ and } \mathrm{supp} f \subseteq V$$

we write

$$K \prec f \prec V.$$

**Theorem 3.2.2.** *Let $K$ be compact subset $X$.*

(a) *Suppose $K \subseteq V$ where $V$ is open. There exists a function $f$ on $X$ such that*

$$K \prec f \prec V.$$

(b) *Suppose $X$ is compact and $K \subseteq V_1 \cup ... \cup V_n$, where $K$ is compact and $V_1, ..., V_n$ are open. There exist functions $h_1, ..., h_n$ on $X$ such that*

$$h_i \prec V_i, \ i = 1, ..., n$$

*and*

$$h_1 + ... + h_n = 1 \text{ on } K.$$

PROOF. (a) Suppose $\varepsilon = \frac{1}{2}\min_K d(\cdot, V^c)$. By Corollary 3.1.2, $\varepsilon > 0$. The continuous function $f = \Pi^X_{K,\varepsilon}$ satisfies $\chi_K \leq f \leq \chi_{K_\varepsilon}$, that is $K \prec f \prec K_\varepsilon$. Part (a) follows if we note that the closure $(K_\varepsilon)^-$ of $K_\varepsilon$ is contained in $V$.

(b) For each $x \in K$ there exists an $r_x > 0$ such that $B(x, r_x) \subseteq V_i$ for some $i$. Let $U_x = B(x, \frac{1}{2}r_x)$. It is important to note that $(U_x)^- \subseteq V_i$ and $(U_x)^-$ is compact since $X$ is compact. There exist points $x_1, ..., x_m \in K$ such that $\cup_{j=1}^m U_{x_i} \supseteq K$. If $1 \leq i \leq n$, let $F_i$ denote the union of those $(U_{x_j})^-$ which are contained in $V_i$. By Part (a), there exist continuous functions $f_i$ such that $F_i \prec f_i \prec V_i$, $i = 1, ..., n$. Define

$$\begin{aligned} h_1 &= f_1 \\ h_2 &= (1 - f_1)f_2 \\ &\quad .... \\ h_n &= (1 - f_1)...(1 - f_{n-1})f_n. \end{aligned}$$

Clearly, $h_i \prec V_i$, $i = 1, ..., n$. Moreover, by induction, we get

$$h_1 + ... + h_n = 1 - (1 - f_1)...(1 - f_{n-1})(1 - f_n).$$

Since $\cup_{i=1}^n F_i \supseteq K$ we are done.

The uniqueness in Theorem 3.2.1 is simple to prove. Suppose $\mu_1$ and $\mu_2$ are two measures for which the theorem holds. Fix $\varepsilon > 0$ and compact $K \subseteq X$ and choose an open set $V$ so that $\mu_2(V) \leq \mu_2(K) + \varepsilon$. If $K \prec f \prec V$,

$$\mu_1(K) = \int_X \chi_K d\mu_1 \leq \int_X f d\mu_1 = Tf$$

$$= \int_X f d\mu_2 \leq \int_X \chi_V d\mu_2 = \mu_2(V) \leq \mu_2(K) + \varepsilon.$$

Thus $\mu_1(K) \leq \mu_2(K)$. If we interchange the roles of the two measures, the opposite inequality is obtained, and the uniqueness of $\mu$ follows.

To prove the existence of the measure $\mu$ in Theorem 3.2.1, define for every open $V$ in $X$,

$$\mu(V) = \sup_{f \prec V} Tf.$$

Here $\mu(\phi) = 0$ since the supremum over the empty set, by convention, equals 0. Note also that $\mu(X) = T1$. Moreover, $\mu(V_1) \leq \mu(V_2)$ if $V_1$ and $V_2$ are open and $V_1 \subseteq V_2$. Now set

$$\mu(E) = \inf_{\substack{V \supseteq E \\ V \text{ open}}} \mu(V) \text{ if } E \in \mathcal{B}(X).$$

Clearly, $\mu(E_1) \leq \mu(E_2)$, if $E_1 \subseteq E_2$ and $E_1, E_2 \in \mathcal{B}(X)$. We therefore say that $\mu$ is increasing.

**Lemma 3.2.1.** (a) *If* $V_1, ..., V_n$ *are open,*

$$\mu(\cup_{i=1}^n V_i) \leq \Sigma_{i=1}^n \mu(V_i).$$

(b) *If* $E_1, E_2, ... \in \mathcal{B}(X)$,

$$\mu(\cup_{i=1}^\infty E_i) \leq \Sigma_{i=1}^\infty \mu(E_i).$$

(c) *If* $K_1, ..., K_n$ *are compact and pairwise disjoint,*

$$\mu(\cup_{i=1}^n K_i) = \Sigma_{i=1}^n \mu(K_i).$$

PROOF. (a) It is enough to prove (a) for $n = 2$. To this end first choose $g \prec V_1 \cup V_2$ and then $h_i \prec V_i$, $i = 1, 2$, such that $h_1 + h_2 = 1$ on supp $g$. Then

$$g = h_1 g + h_2 g$$

and it follows that

$$Tg = T(h_1 g) + T(h_2 g) \leq \mu(V_1) + \mu(V_2).$$

Thus

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2).$$

(b) Choose $\varepsilon > 0$ and for each $i \in \mathbf{N}_+$, choose an open $V_i \supseteq E_i$ such $\mu(V_i) < \mu(E_i) + 2^{-i}\varepsilon$. Set $V = \cup_{i=1}^{\infty} V_i$ and choose $f \prec V$. Since supp$f$ is compact, $f \prec V_1 \cup ... \cup V_n$ for some $n$. Thus, by Part (a),

$$Tf \leq \mu(V_1 \cup ... \cup V_n) \leq \Sigma_{i=1}^{n} \mu(V_i) \leq \Sigma_{i=1}^{\infty} \mu(E_i) + \varepsilon$$

and we get

$$\mu(V) \leq \Sigma_{i=1}^{\infty} \mu(E_i)$$

since $\varepsilon > 0$ is arbitrary. But $\cup_{i=1}^{\infty} E_i \subseteq V$ and it follows that

$$\mu(\cup_{i=1}^{\infty} E_i) \leq \Sigma_{i=1}^{\infty} \mu(E_i).$$

(c) It is enough to treat the special case $n = 2$. Choose $\varepsilon > 0$. Set $\rho = d(K_1, K_2)$ and $V_1 = (K_1)_{\rho/2}$ and $V_2 = (K_2)_{\rho/2}$. There is an open set $U \supseteq K_1 \cup K_2$ such that $\mu(U) < \mu(K_1 \cup K_2) + \varepsilon$ and there are functions $f_i \prec U \cap V_i$ such that $Tf_i > \mu(U \cap V_i) - \varepsilon$ for $i = 1, 2$. Now, using that $\mu$ increases

$$\mu(K_1) + \mu(K_2) \leq \mu(U \cap V_1) + \mu(U \cap V_2)$$

$$\leq Tf_1 + Tf_2 + 2\varepsilon = T(f_1 + f_2) + 2\varepsilon.$$

Since $f_1 + f_2 \prec U$,

$$\mu(K_1) + \mu(K_2) \leq \mu(U) + 2\varepsilon \leq \mu(K_1 \cup K_2) + 3\varepsilon$$

and, by letting $\varepsilon \to 0$,

$$\mu(K_1) + \mu(K_2) \leq \mu(K_1 \cup K_2).$$

The reverse inequality follows from Part (b). The lemma is proved.

Next we introduce the class

$$\mathcal{M} = \left\{ E \in \mathcal{B}(X);\ \mu(E) = \sup_{\substack{K \subseteq E \\ K \text{ compact}}} \mu(K) \right\}$$

Since $\mu$ is increasing $\mathcal{M}$ contains every compact set. Recall that a closed set in $X$ is compact, since $X$ is compact. Especially, note that $\phi$ and $X \in \mathcal{M}$.

COMPLETION OF THE PROOF OF THEOREM 3.2.1:

CLAIM 1. $\mathcal{M}$ contains every open set.

PROOF OF CLAIM 1. Let $V$ be open and suppose $\alpha < \mu(V)$. There exists an $f \prec V$ such that $\alpha < Tf$. If $U$ is open and $U \supseteq K =_{def} \text{supp} f$, then $f \prec U$, and hence $Tf \leq \mu(U)$. But then $Tf \leq \mu(K)$. Thus $\alpha < \mu(K)$ and Claim 1 follows since $K$ is compact and $K \subseteq V$.

CLAIM 2. Let $(E_i)_{i=1}^{\infty}$ be a disjoint denumerable collection of members of $\mathcal{M}$ and put $E = \cup_{i=1}^{\infty} E_i$. Then

$$\mu(E) = \Sigma_{i=1}^{\infty} \mu(E_i)$$

and $E \in \mathcal{M}$.

PROOF OF CLAIM 2. Choose $\varepsilon > 0$ and for each $i \in \mathbf{N}_+$, choose a compact $K_i \subseteq E_i$ such that $\mu(K_i) > \mu(E_i) - 2^{-i}\varepsilon$. Set $H_n = K_1 \cup ... \cup K_n$. Then, by Lemma 3.2.1 (c),

$$\mu(E) \geq \mu(H_n) = \Sigma_{i=1}^{n} \mu(K_i) > \Sigma_{i=1}^{n} \mu(E_i) - \varepsilon$$

and we get

$$\mu(E) \geq \Sigma_{i=1}^{\infty} \mu(E_i).$$

Thus, by Lemma 3.2.1 (b), $\mu(E) = \Sigma_{i=1}^{\infty} \mu(E_i)$. To prove that $E \in \mathcal{M}$, let $\varepsilon$ be as in the very first part of the proof and choose $n$ such that

$$\mu(E) \leq \Sigma_{i=1}^{n} \mu(E_i) + \varepsilon.$$

Then

$$\mu(E) < \mu(H_n) + 2\varepsilon$$

and this shows that $E \in \mathcal{M}$.

CLAIM 3. Suppose $E \in \mathcal{M}$ and $\varepsilon > 0$. Then there exist a compact $K$ and an open $V$ such that $K \subseteq E \subseteq V$ and $\mu(V \setminus K) < \varepsilon$.

PROOF OF CLAIM 3. The definitions show that there exist a compact $K$ and an open $V$ such that

$$\mu(V) - \frac{\varepsilon}{2} < \mu(E) < \mu(K) + \frac{\varepsilon}{2}.$$

The set $V \setminus K$ is open and $V \setminus K \in \mathcal{M}$ by Claim 1. Thus Claim 2 implies that

$$\mu(K) + \mu(V \setminus K) = \mu(V) < \mu(K) + \varepsilon$$

and we get $\mu(V \setminus K) < \varepsilon$.

CLAIM 4. If $A \in \mathcal{M}$, then $X \setminus A \in \mathcal{M}$.

PROOF OF CLAIM 4. Choose $\varepsilon > 0$. Furthermore, choose compact $K \subseteq A$ and open $V \supseteq A$ such that $\mu(V \setminus K) < \varepsilon$. Then

$$X \setminus A \subseteq (V \setminus K) \cup (X \setminus V).$$

Now, by Lemma 3.2.1 (b),

$$\mu(X \setminus A) \leq \varepsilon + \mu(X \setminus V).$$

Since $X \setminus V$ is a compact subset of $X \setminus A$, we conclude that $X \setminus A \in \mathcal{M}$.

Claims 1, 2 and 4 prove that $\mathcal{M}$ is a $\sigma$-algebra which contains all Borel sets. Thus $\mathcal{M} = \mathcal{B}(X)$.

We finally prove (a). It is enough to show that

$$Tf \leq \int_X f d\mu$$

for each $f \in C(X)$. For once this is known

$$-Tf = T(-f) \leq \int_X -f d\mu \leq -\int_X f d\mu$$

and (a) follows.

Choose $\varepsilon > 0$. Set $f(X) = [a, b]$ and choose $y_0 < y_1 < ... < y_n$ such that $y_1 = a$, $y_{n-1} = b$, and $y_i - y_{i-1} < \varepsilon$. The sets

$$E_i = f^{-1}([y_{i-1}, y_i[), \ i = 1, ..., n$$

constitute a disjoint collection of Borel sets with the union $X$. Now, for each $i$, pick an open set $V_i \supseteq E_i$ such that $\mu(V_i) \leq \mu(E_i) + \frac{\varepsilon}{n}$ and $V_i \subseteq f^{-1}(]-\infty, y_i[)$. By Theorem 3.2.2 there are functions $h_i \prec V_i$, $i = 1, ..., n$, such that $\Sigma_{i=1}^n h_i = 1$ on supp$f$ and $h_i f \prec y_i h_i$ for all $i$. From this we get

$$\begin{aligned} Tf = \Sigma_{i=1}^n T(h_i f) &\leq \Sigma_{i=1}^n y_i T h_i \leq \Sigma_{i=1}^n y_i \mu(V_i) \\ &\leq \Sigma_{i=1}^n y_i \mu(E_i) + \Sigma_{i=1}^n y_i \frac{\varepsilon}{n} \\ &\leq \Sigma_{i=1}^n (y_i - \varepsilon)\mu(E_i) + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon \\ &\leq \Sigma_{i=1}^n \int_{E_i} f d\mu + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon \\ &= \int_X f d\mu + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we get

$$Tf \leq \int_X f d\mu.$$

This proves Theorem 3.2.1.

It is now simple to show the existence of volume measure in $\mathbf{R}^n$. For pedagogical reasons we first discuss the so called volume measure in the unit cube $Q = [0, 1]^n$ in $\mathbf{R}^n$.

The Riemann integral

$$\int_Q f(x)dx,$$

is a positive linear functional as a function of $f \in C(Q)$. Moreover, $T1 = 1$ and the Riesz Representation Theorem gives us a Borel probability measure $\mu$ in $Q$ such that

$$\int_Q f(x)dx = \int_Q f d\mu.$$

Suppose $A \subseteq Q$ is a closed $n$-cell and $i \in \mathbf{N}_+$. Then

$$\mathrm{vol}(A) \leq \int_Q \Pi^Q_{A,2^{-i}}(x)dx \leq \mathrm{vol}(A_{2^{-i}})$$

and

$$\Pi^Q_{A,2^{-i}}(x) \to \chi_A(x) \text{ as } i \to \infty$$

for every $x \in \mathbf{R}^n$. Thus

$$\mu(A) = \mathrm{vol}(A).$$

The measure $\mu$ is called the volume measure in the unit cube. In the special case $n = 2$ it is called the area measure in the unit square and if $n = 1$ it is called the linear measure in the unit interval.

PROOF OF THEOREM 1.1.1. Let $\hat{\mathbf{R}}=\mathbf{R}\cup\{-\infty,\infty\}$ be the two-point compactification of $\mathbf{R}$ introduced in Example 3.1.3 and let $\hat{\mathbf{R}}^n$ denote the product of $n$ copies of the metric space $\hat{\mathbf{R}}$. Clearly,

$$\mathcal{B}(\mathbf{R}^n) = \left\{ A \cap \mathbf{R}^n;\ A \in \mathcal{B}(\hat{\mathbf{R}}^n) \right\}.$$

Moreover, let $w : \mathbf{R}^n \to ]0,\infty[$ be a continuous map such that

$$\int_{\mathbf{R}^n} w(x)dx = 1.$$

Now we define

$$Tf = \int_{\mathbf{R}^n} f(x)w(x)dx,\ f \in C(\hat{\mathbf{R}}^n).$$

Note that $T1 = 1$. The function $T$ is a positive linear functional on $C(\hat{\mathbf{R}}^n)$ and the Riesz Representation Theorem gives us a Borel probability measure $\mu$ on $\hat{\mathbf{R}}^n$ such that

$$\int_{\mathbf{R}^n} f(x)w(x)dx = \int_{\hat{\mathbf{R}}^n} f d\mu,\ f \in C(\hat{\mathbf{R}}^n).$$

As above we get

$$\int_A w(x)dx = \mu(A)$$

for each compact $n$-cell in $\mathbf{R}^n$. Thus

$$\mu(\mathbf{R}^n) = \lim_{i\to\infty} \int_{[-i,i]^n} w(x)dx = 1$$

and we conclude that $\mu$ is concentrated on $\mathbf{R}^n$. Set $\mu_0(A) = \mu(A)$, $A \in \mathcal{B}(\mathbf{R}^n)$, and

$$dm_n = \frac{1}{w} d\mu_0.$$

Then, if $f \in C_c(\mathbf{R}^n)$,

$$\int_{\mathbf{R}^n} f(x)w(x)dx = \int_{\mathbf{R}^n} f d\mu_0$$

and by replacing $f$ by $f/w$,

$$\int_{\mathbf{R}^n} f(x)dx = \int_{\mathbf{R}^n} f dm_n.$$

From this $m_n(A)$ =vol$(A)$ for every compact $n$-cell $A$ and it follows that $m_n$ is the volume measure on $\mathbf{R}^n$. Theorem 1.1.1 is proved.

↑↑↑

**3.3 q-Adic Expansions of Numbers in the Unit Interval**

To begin with in this section we will discuss so called q-adic expansions of real numbers and give some interesting consequences. As an example of an application, we construct a one-to-one real-valued Borel map $f$ defined on a proper interval such that the range of $f$ is a Lebesgue null set. Another example exhibits an increasing continuous function $G$ on the unit interval with the range equal to the unit interval such that the derivative of $G$ is equal to zero almost everywhere with respect to Lebesgue measure. In the next section we will give more applications of q-adic expansions in connection with infinite product measures.

To simplify notation let $(\Omega, P, \mathcal{F}) = ([0,1[\,, \mathcal{B}([0,1[), v_{1|[0,1[})$. Furthermore, let $q \geq 2$ be an integer and define a function $h : \mathbf{R} \to \{0,1,2,...,q-1\}$ of period one such that

$$h(x) = k,\ \frac{k}{q} \leq x < \frac{k+1}{q},\ k = 0,...,q-1.$$

Furthermore, set for each $n \in \mathbf{N}_+$,

$$\xi_n(\omega) = h(q^{n-1}\omega),\ 0 \leq \omega < 1.$$

Then

$$P\left[\xi_n = k\right] = \frac{1}{q},\ k = 0,...,q-1.$$

Moreover, if $k_1,...,k_n \in \{0,1,2,...,q-1\}$, it becomes obvious on drawing a figure that

$$P\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}\right] = \Sigma_{i=0}^{q-1} P\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}, \xi_n = i\right]$$

where each term in the sum in the right-hand side has the same value. Thus

$$P\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}\right] = qP\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}, \xi_n = k_n\right]$$

and

$$P\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}, \xi_n = k_n\right] = P\left[\xi_1 = k_1,...,\xi_{n-1} = k_{n-1}\right] P\left[\xi_n = k_n\right].$$

By repetition,

$$P\left[\xi_1 = k_1,...\xi_{n-1} = k_{n-1}, \xi_n = k_n\right] = \Pi_{i=1}^{n} P\left[\xi_i = k_i\right].$$

From this we get

$$P\left[\xi_1 \in A_1,...\xi_{n-1} \in A_{n-1}, \xi_n \in A_n\right] = \Pi_{i=1}^{n} P\left[\xi_i \in A_i\right]$$

for all $A_1, ..., A_n \subseteq \{0, 1, 2, ..., q-1\}$.

Note that each $\omega \in [0, 1[$ has a so called $q$-adic expansion

$$\omega = \Sigma_{i=1}^{\infty} \frac{\xi_i(\omega)}{q^i}.$$

If necessary, we write $\xi_n = \xi_n^{(q)}$ to indicate $q$ explicitly.

Let $k_0 \in \{0, 1, 2, ..., q-1\}$ be fixed and consider the event $A$ that a number in $[0, 1[$ does not have $k_0$ in its $q$-adic expansion. The probability of $A$ equals

$$P[A] = P[\xi_i \neq k_0,\ i = 1, 2, ...] = \lim_{n \to \infty} P[\xi_i \neq k_0,\ i = 1, 2, ..., n]$$

$$= \lim_{n \to \infty} \Pi_{i=1}^{n} P[\xi_i \neq k_0] = \lim_{n \to \infty} (\frac{q-1}{q})^n = 0.$$

In particular, if

$$D_n = \left\{ \omega \in [0, 1[\,;\ \xi_i^{(3)} \neq 1,\ i = 1, ..., n \right\}.$$

then, $D = \cap_{n=1}^{\infty} D_n$ is a $P$-zero set.

Set

$$f(\omega) = \Sigma_{i=1}^{\infty} \frac{2\xi_i^{(2)}(\omega)}{3^i},\ 0 \leq \omega < 1.$$

We claim that $f$ is one-to-one. If $0 \leq \omega, \omega' < 1$ and $\omega \neq \omega'$ let $n$ be the least $i$ such that $\xi_i^{(2)}(\omega) \neq \xi_i^{(2)}(\omega')$; we may assume that $\xi_n^{(2)}(\omega) = 0$ and $\xi_n^{(2)}(\omega') = 1$. Then

$$f(\omega') \geq \Sigma_{i=1}^{n} \frac{2\xi_i^{(2)}(\omega')}{3^i} = \Sigma_{i=1}^{n-1} \frac{2\xi_i^{(2)}(\omega')}{3^i} + \frac{2}{3^n}$$

$$= \Sigma_{i=1}^{n-1} \frac{2\xi_i^{(2)}(\omega)}{3^i} + \Sigma_{i=n+1}^{\infty} \frac{4}{3^i} > \Sigma_{i=1}^{\infty} \frac{2\xi_i^{(2)}(\omega)}{3^i} = f(\omega).$$

Thus $f$ is one-to-one. We next prove that $f(\Omega) = D$. To this end choose $y \in D$. If $\xi_i^{(3)}(y) = 2$ for all $i \in \mathbf{N}_+$, then $y = 1$ which is a contradiction. If $k \geq 1$ is fixed and $\xi_k^{(3)}(y) = 0$ and $\xi_i^{(3)}(y) = 2,\ i \geq k+1$, then it is readily seen that $\xi_k^{(3)}(y) = 1$ which is a contradiction. Now define

$$\omega = \Sigma_{i=1}^{\infty} \frac{\frac{1}{2}\xi_i^{(3)}(y)}{2^i}$$

and we have $f(\omega) = y$.

Let $C_n = D_n^-$, $n \in \mathbf{N}_+$. The set $C = \cap_{n=1}^{\infty} C_n$, is called the Cantor set. The Cantor set is a compact Lebesgue zero set. The construction of the Cantor set may alternatively be described as follows. First $C_0 = [0,1]$. Then trisect $C_0$ and remove the middle interval $\left]\frac{1}{3}, \frac{2}{3}\right[$ to obtain $C_1 = C_0 \setminus \left]\frac{1}{3}, \frac{2}{3}\right[ = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. At the second stage subdivide each of the closed intervals of $C_1$ into thirds and remove from each one the middle open thirds. Then $C_2 = C_1 \setminus (\left]\frac{1}{9}, \frac{2}{9}\right[ \cup \left]\frac{7}{9}, \frac{8}{9}\right[)$. What is left from $C_{n-1}$ is $C_n$ defined above. The set $[0,1] \setminus C_n$ is the union of $2^n - 1$ intervals numbered $I_k^n$, $k = 1, ..., 2^n - 1$, where the interval $I_k^n$ is situated to the left of the interval $I_l^n$ if $k < l$.

Suppose $n$ is fixed and let $G_n : [0,1] \to [0,1]$ be the unique monotone increasing continuous function, which satisfies $G_n(0) = 0, G_n(1) = 1, G_n(x) = k2^{-n}$ for $x \in I_k^n$ and which is affine on each interval of $C_{n.}$ It is clear that $G_n = G_{n+1}$ on each interval $I_k^n$, $k = 1, ..., 2^n - 1$. Moreover, $\mid G_n - G_{n+1} \mid \leq 2^{-n-1}$ and thus

$$\mid G_n - G_{n+k} \mid \leq \Sigma_{k=n}^{n+k} \mid G_k - G_{k+1} \mid \leq 2^{-n}.$$

Let $G(x) = \lim_{n \to \infty} G_n(x)$, $0 \leq x \leq 1$. The continuous and increasing function $G$ is constant on each removed interval and it follows that $G' = 0$ a.e. with respect to linear measure in the unit interval.The function $G$ is called the Cantor function or Cantor-Lebesgue function.

Next we introduce the following convention, which is standard in Lebesgue integration. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose $A \in \mathcal{M}$ and $\mu(A^c) = 0$. If two functions $g, h \in \mathcal{L}^1(\mu)$ agree on $A$,

$$\int_X g d\mu = \int_X h d\mu.$$

If a function $f : A \to \mathbf{R}$ is the restriction to $A$ of a function $g \in \mathcal{L}^1(\mu)$ we define

$$\int_X f d\mu = \int_X g d\mu.$$

Now suppose $F : \mathbf{R} \to \mathbf{R}$ is a right continuous increasing function and let $\mu$ be the unique positive Borel such that

$$\mu(]a, x]) = F(x) - F(a) \text{ if } a, x \in \mathbf{R} \text{ and } a < x.$$

If $h \in L^1(\mu)$ and $E \in \mathcal{R}$, the so called Stieltjes integral

$$\int_E h(x) dF(x)$$

is by definition equal to

$$\int_E h d\mu.$$

If $a, b \in \mathbf{R}$, $a < b$, and $F$ is continuous at the points $a$ and $b$, we define

$$\int_a^b h(x)dF(x) = \int_I h d\mu$$

where $I$ is any interval with boundary points $a$ and $b$.

The reader should note that the integral

$$\int_{\mathbf{R}} h(x)dF(x)$$

in general is different from the integral

$$\int_{\mathbf{R}} h(x)F'(x)dx.$$

For example, if $G$ is the Cantor function and $G$ is extended so that $G(x) = 0$ for negative $x$ and $G(x) = 1$ for $x$ larger than 1, clearly

$$\int_{\mathbf{R}} h(x)G'(x)dx = 0$$

since $G'(x) = 0$ a.e. $[m]$. On the other hand, if we choose $h = \chi_{[0,1]}$,

$$\int_{\mathbf{R}} h(x)dG(x) = 1.$$

### 3.4. Product Measures

Suppose $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are two measurable spaces. If $A \in \mathcal{M}$ and $B \in \mathcal{N}$, the set $A \times B$ is called a measurable rectangle in $X \times Y$. The product $\sigma$-algebra $\mathcal{M} \otimes \mathcal{N}$ is, by definition, the $\sigma$-algebra generated by all measurable rectangles in $X \times Y$. If we introduce the projections

$$\pi_X(x, y) = x,\ (x, y) \in X \times Y$$

and

$$\pi_Y(x, y) = y,\ (x, y) \in X \times Y,$$

the product $\sigma$-algebra $\mathcal{M} \otimes \mathcal{N}$ is the least $\sigma$-algebra $\mathcal{S}$ of subsets of $X \times Y$, which makes the maps $\pi_X : (X \times Y, \mathcal{S}) \to (X, \mathcal{M})$ and $\pi_Y : (X \times Y, \mathcal{S}) \to (Y, \mathcal{N})$ measurable, that is $\mathcal{M} \otimes \mathcal{N} = \sigma(\pi_X^{-1}(\mathcal{M}) \cup \pi_Y^{-1}(\mathcal{N}))$..

Suppose $\mathcal{E}$ generates $\mathcal{M}$, where $X \in \mathcal{E}$, and $\mathcal{F}$ generates $\mathcal{N}$, where $Y \in \mathcal{F}$. We claim that the class

$$\mathcal{E} \boxtimes \mathcal{F} = \{E \times F; E \in \mathcal{E} \text{ and } F \in \mathcal{F}\}$$

generates the $\sigma$-algebra $\mathcal{M} \otimes \mathcal{N}$. First it is clear that

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) \subseteq \mathcal{M} \otimes \mathcal{N}.$$

Moreover, the class

$$\{E \in \mathcal{M}; E \times Y \in \sigma(\mathcal{E} \boxtimes \mathcal{F})\ \} = \mathcal{M} \cap \left\{E \subseteq X;\ \pi_X^{-1}(E) \in \sigma(\mathcal{E} \boxtimes \mathcal{F})\ \right\}$$

is a $\sigma$-algebra, which contains $\mathcal{E}$ and therefore equals $\mathcal{M}$. Thus $A \times Y \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $A \in \mathcal{M}$ and, in a similar way, $X \times B \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $B \in \mathcal{N}$ and we conclude that $A \times B = (A \times Y) \cap (X \times B) \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $A \in \mathcal{M}$ and all $B \in \mathcal{N}$. This proves that

$$\mathcal{M} \otimes \mathcal{N} \subseteq \sigma(\mathcal{E} \boxtimes \mathcal{F})$$

and it follows that

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) = \mathcal{M} \otimes \mathcal{N}.$$

Thus

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) = \sigma(\mathcal{E}) \otimes \sigma(\mathcal{F}) \text{ if } X \in \mathcal{E} \text{ and } Y \in \mathcal{F}.$$

Since the $\sigma$-algebra $\mathcal{R}_n$ is generated by all open $n$-cells in $\mathbf{R}^n$, we conclude that

$$\mathcal{R}_{k+n} = \mathcal{R}_k \otimes \mathcal{R}_n.$$

Given $E \subseteq X \times Y$, define

$$E_x = \{y;\ (x, y) \in E\} \text{ if } x \in X$$

and

$$E^y = \{x;\ (x, y) \in E\} \text{ if } y \in Y.$$

If $f : X \times Y \to Z$ is a function and $x \in X$, $y \in Y$, let

$$f_x(y) = f(x, y), \text{ if } y \in Y$$

and

$$f^y(x) = f(x, y), \text{ if } x \in X.$$

**Theorem 3.4.1** (a) *If $E \in \mathcal{M} \otimes \mathcal{N}$, then $E_x \in \mathcal{N}$* and *$E^y \in \mathcal{M}$ for every $x \in X$ and $y \in Y$.*

(b) *If $f : (X \times Y, \mathcal{M} \otimes \mathcal{N}) \to (Z, \mathcal{O})$ is measurable, then $f_x$ is $(\mathcal{N}, \mathcal{O})$-measurable for each $x \in X$ and $f^y$ is $(\mathcal{M}, \mathcal{O})$-measurable for each $y \in Y$.*

**Proof.** (a) Choose $y \in Y$ and define $\varphi : X \to X \times Y$ by $\varphi(x) = (x, y)$. Then

$$\mathcal{M} = \sigma(\varphi^{-1}(\mathcal{M} \boxtimes \mathcal{N})) = \varphi^{-1}(\sigma(\mathcal{M} \boxtimes \mathcal{N})) = \varphi^{-1}(\mathcal{M} \otimes \mathcal{N})$$

and it follows that $E^y \in \mathcal{M}$. In a similar way $E_x \in \mathcal{N}$ for every $x \in X$.

(b) For any set $V \in \mathcal{O}$,

$$(f^{-1}(V))_x = (f_x)^{-1}(V)$$

and

$$(f^{-1}(V))^y = (f^y)^{-1}(V).$$

Part (b) now follows from (a).

Below an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable or $(\mathcal{M}, \mathcal{R})$-measurable function is simply called $\mathcal{M}$-measurable.

**Theorem 3.4.2.** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are positive $\sigma$-finite measurable spaces and suppose $E \in \mathcal{M} \otimes \mathcal{N}$. If*

$$f(x) = \nu(E_x) \text{ and } g(y) = \mu(E^y)$$

*for every $x \in X$ and $y \in Y$, then $f$ is $\mathcal{M}$-measurable, $g$ is $\mathcal{N}$-measurable, and*

$$\int_X f d\mu = \int_Y g d\nu.$$

**Proof.** We first assume that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are finite positive measure spaces.

Let $\mathcal{D}$ be the class of all sets $E \in \mathcal{M} \otimes \mathcal{N}$ for which the conclusion of the theorem holds. It is clear that the class $\mathcal{G}$ of all measurable rectangles in $X \times Y$ is a subset of $\mathcal{D}$ and $\mathcal{G}$ is a $\pi$-system. Furthermore, the Beppo Levi Theorem shows that $\mathcal{D}$ is a $\sigma$-additive class. Therefore, using Theorem 1.2.2, $\mathcal{M} \otimes \mathcal{N} = \sigma(\mathcal{G}) \subseteq \mathcal{D}$ and it follows that $\mathcal{D} = \mathcal{M} \otimes \mathcal{N}$.

In the general case, choose a denumerable disjoint collection $(X_k)_{k=1}^{\infty}$of members of $\mathcal{M}$ and a denumerable disjoint collection $(Y_n)_{n=1}^{\infty}$of members of $\mathcal{N}$ such that

$$\cup_{k=1}^{\infty} X_k = X \text{ and } \cup_{n=1}^{\infty} Y_n = Y.$$

Set

$$\mu_k = \chi_{X_k} \mu,\ k = 1, 2, ...$$

and

$$\nu_n = \chi_{Y_n} \nu,\ n = 1, 2, ... \ .$$

Then, by the Beppo Levi Theorem, the function

$$f(x) = \int_Y \Sigma_{n=1}^{\infty} \chi_E(x, y) \chi_{Y_n}(y) d\nu(y)$$

$$= \Sigma_{n=1}^{\infty} \int_Y \chi_E(x, y) \chi_{Y_n}(y) d\nu(y) = \Sigma_{n=1}^{\infty} \nu_n(E_x)$$

is $\mathcal{M}$-measurable. Again, by the Beppo Levi Theorem,

$$\int_X f d\mu = \Sigma_{k=1}^{\infty} \int_X f d\mu_k$$

and

$$\int_X f d\mu = \Sigma_{k=1}^{\infty} (\Sigma_{n=1}^{\infty} \int_X \nu_n(E_x) d\mu_k(x)) = \Sigma_{k,n=1}^{\infty} \int_X \nu_n(E_x) d\mu_k(x).$$

In a similar way, the function $g$ is $\mathcal{N}$-measurable and

$$\int_Y g d\nu = \Sigma_{n=1}^{\infty} (\Sigma_{k=1}^{\infty} \int_Y \mu_k(E^y) d\nu_n(y)) = \Sigma_{k,n=1}^{\infty} \int_Y \mu_k(E^y) d\nu_n(y).$$

Since the theorem is true for finite positive measure spaces, the general case follows.

**Definition 3.4.1.** If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are positive $\sigma$-finite measurable spaces and $E \in \mathcal{M} \otimes \mathcal{N}$, define

$$(\mu \times \nu)(E) = \int_X \nu(E_x) d\mu(x) = \int_Y \mu(E^y) d\nu(y).$$

The function $\mu \times \nu$ is called the product of the measures $\mu$ and $\nu$.

Note that Beppo Levi's Theorem ensures that $\mu \times \nu$ is a positive measure.

Before the next theorem we recall the following convention. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose $A \in \mathcal{M}$ and $\mu(A^c) = 0$. If two functions $g, h \in \mathcal{L}^1(\mu)$ agree on $A$,

$$\int_X g d\mu = \int_X h d\mu.$$

If a function $f : A \to \mathbf{R}$ is the restriction to $A$ of a function $g \in \mathcal{L}^1(\mu)$ we define

$$\int_X f d\mu = \int_X g d\mu.$$

**Theorem 3.4.3.** *Let* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *be positive* $\sigma$*-finite measurable spaces.*

(a) **(Tonelli's Theorem)** *If* $h : X \times Y \to [0, \infty]$ *is* $(\mathcal{M} \otimes \mathcal{N})$*-measurable and*

$$f(x) = \int_Y h(x, y) d\nu(y) \text{ and } g(y) = \int_X h(x, y) d\mu(x)$$

*for every* $x \in X$ *and* $y \in Y$*, then* $f$ *is* $\mathcal{M}$*-measurable,* $g$ *is* $\mathcal{N}$*-measurable, and*

$$\int_X f d\mu = \int_{X \times Y} h d(\mu \times \nu) = \int_Y g d\nu$$

(b) **(Fubini's Theorem)**

$(i)$ *If* $h : X \times Y \to \mathbf{R}$ *is* $(\mathcal{M} \otimes \mathcal{N})$*-measurable and*

$$\int_X \left( \int_Y \mid h(x, y) \mid d\nu(y) \right) d\mu(x) < \infty$$

*then* $h \in L^1(\mu \times \nu)$. *Moreover,*

$$\int_X (\int_Y h(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} hd(\mu\times\nu) = \int_Y (\int_X h(x,y)d\mu(x))d\nu(y)$$

(*ii*) *If* $h \in L^1((\mu\times\nu)^-)$, *then* $h_x \in L^1(\nu)$ *for* $\mu$*-almost all* $x$ *and*

$$\int_{X\times Y} hd(\mu\times\nu) = \int_X (\int_Y h(x,y)d\nu(y))d\mu(x)$$

(*iii*) *If* $h \in L^1((\mu\times\nu)^-)$, *then* $h^y \in L^1(\mu)$ *for* $\nu$*-almost all* $y$ *and*

$$\int_{X\times Y} hd(\mu\times\nu) = \int_Y (\int_X h(x,y)d\mu(x))d\nu(y)$$

PROOF. (a) The special case when $h$ is a non-negative $(\mathcal{M}\otimes\mathcal{N})$-measurable simple function follows from Theorem 3.4.2. Remembering that any non-negative measurable function is the pointwise limit of an increasing sequence of simple measurable functions, the Lebesgue Monotone Convergence Theorem implies the Tonelli Theorem.

(b) PART $(i)$ : By Part (a)

$$\infty > \int_X (\int_Y h^+(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^+ d(\mu\times\nu)$$

$$= \int_Y (\int_X h^+(x,y)d\mu(x))d\nu(y)$$

and

$$\infty > \int_X (\int_Y h^-(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^- d(\mu\times\nu)$$

$$= \int_Y (\int_X h^-(x,y)d\mu(x))d\nu(y).$$

In particular, $h = h^+ - h^- \in L^1(\mu\times\nu)$. Let

$$A = \left\{x \in X;\ (h^+)_x, (h^-)_x \in L^1(\nu)\right\}.$$

Then $A^c$ is a $\mu$-null set and we get

$$\int_A (\int_Y h^+(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^+ d(\mu\times\nu)$$

and

$$\int_A (\int_Y h^-(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^- d(\mu\times\nu).$$

Thus

$$\int_A (\int_Y h(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} hd(\mu\times\nu)$$

and, hence,

$$\int_X (\int_Y h(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} hd(\mu\times\nu).$$

The other case can be treated in a similar way. The theorem is proved.

PART $(ii)$ : We first use Theorem 2.2.3 and write $h = \varphi + \psi$ where $\varphi \in L^1(\mu\times\nu)$, $\psi$ is $(\mathcal{M}\otimes\mathcal{N})^-$-measurable and $\psi = 0$ a.e. $[\mu\times\nu]$. Set

$$A = \left\{x \in X;\ (\varphi^+)_x, (\varphi^-)_x \in L^1(\nu)\right\}.$$

Furthermore, suppose $E \supseteq \{(x,y);\ \psi(x,y) \neq 0\}$, $E \in \mathcal{M}\otimes\mathcal{N}$ and

$$(\mu\times\nu)(E) = 0.$$

Then, by Tonelli's Theorem

$$0 = \int_X \nu(E_x)d\mu(x).$$

Let $B = \{x \in X;\ \nu(E_x) \neq 0\}$ and note that $B \in \mathcal{M}$. Moreover $\mu(B) = 0$ and if $x \notin B$, then $\psi_x = 0$ a.e. $[\nu]$ that is $h_x = \varphi_x$ a.e. $[\nu]$. Now, by Part $(i)$

$$\int_{X\times Y} hd(\mu\times\nu)^- = \int_{X\times Y} \varphi d(\mu\times\nu) = \int_A (\int_Y \varphi(x,y)d\nu(y))d\mu(x)$$

$$= \int_{A\cap B^c} (\int_Y \varphi(x,y)d\nu(y))d\mu(x) = \int_{A\cap B^c} (\int_Y h(x,y)d\nu(y))d\mu(x)$$

$$= \int_X (\int_Y h(x,y)d\nu(y))d\mu(x).$$

Part $(iii)$ is proved in the same manner as Part $(ii)$. This concludes the proof of the theorem.

If $(X_i, \mathcal{M}_i)$, $i = 1, ..., n$, are measurable spaces, the product $\sigma$-algebra $\mathcal{M}_1 \otimes ... \otimes \mathcal{M}_n$ is, by definition, the $\sigma$-algebra generated by all sets of the form

$$A_1 \times ... \times A_n$$

where $A_i \in \mathcal{M}_i$, $i = 1, ..., n$. Now assume $(X_i, \mathcal{M}_i, \mu_i)$, $i = 1, ..., n$, are $\sigma$-finite positive measure spaces. By induction, we define $\nu_1 = \mu_1$ and $\nu_k = \nu_{k-1} \times \mu_k$, $k = 1, 2, ..., n$. The measure, $\nu_n$ is called the product of the measures $\mu_1, ..., \mu_n$ and is denoted by $\mu_1 \times ... \times \mu_n$. It is readily seen that

$$\mathcal{R}_n = \mathcal{R}_1 \otimes ... \otimes \mathcal{R}_1 \text{ ($n$ factors)}$$

and

$$v_n = v_1 \times ... \times v_1 \text{ ($n$ factors).}$$

Moreover,

$$\mathcal{R}_n^- \supseteq (\mathcal{R}_1^-)^n =_{def} \mathcal{R}_1^- \otimes ... \otimes \mathcal{R}_1^- \quad \text{($n$ factors).}$$

If $A \in \mathcal{P}(\mathbf{R}) \setminus \mathcal{R}_1^-$, by the Tonelli Theorem, the set $A \times \{0, ..., 0\}$ ($n-1$ zeros) is an $m_n$-null set, which, in view of Theorem 3.4.1, cannot belong to the $\sigma$-algebra $(\mathcal{R}_1^-)^n$. Thus the Axiom of Choice implies that

$$\mathcal{R}_n^- \neq (\mathcal{R}_1^-)^n.$$

Clearly, the completion of the measure $m_1 \times ... \times m_1$ ($n$ factors) equals $m_n$.

Sometimes we prefer to write

$$\int_{A_1 \times ... \times A_n} f(x_1, ..., x_n)dx_1...dx_n$$

instead of

$$\int_{A_1 \times ... \times A_n} f(x)dm_n(x)$$

or

$$\int_{A_1 \times ... \times A_n} f(x)dx.$$

Moreover, the integral

$$\int_{A_n} ... \int_{A_1} f(x_1, ..., x_n)dx_1...dx_n$$

is the same as

$$\int_{A_1 \times ... \times A_n} f(x_1, ..., x_n)dx_1...dx_n.$$

**Definition 3.4.2**. (a) The measure

$$\gamma_1(A) = \int_A e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}, \ A \in \mathcal{R}$$

is called the standard Gauss measure in $\mathbf{R}$.

(b) The measure

$$\gamma_n = \gamma_1 \times ... \times \gamma_1 \ (n \text{ factors})$$

is called the standard Gauss measure in $\mathbf{R}^n$. Thus, if

$$| x |= \sqrt{x_1^2 + ... + x_n^2}, \ x = (x_1, ..., x_n) \in \mathbf{R}^n$$

we have

$$\gamma_n(A) = \int_A e^{-\frac{|x|^2}{2}} \frac{dx}{\sqrt{2\pi}^n}, \ A \in \mathcal{R}_n.$$

(c) A Borel measure $\mu$ in $\mathbf{R}$ is said to be a centred Gaussian measure if $\mu = f(\gamma_1)$ for some linear map $f : \mathbf{R} \to \mathbf{R}$.

(d) A real-valued random variable $\xi$ is said to be a centred Gaussian random variable if its probability law is a centred Gaussian measure in $\mathbf{R}$. Stated otherwise, $\xi$ is a real-valued centred Gaussian random variable if either

$$\mathcal{L}(\xi) = \delta_0 \ (\text{abbreviated } \xi \in N(0, 0))$$

or there exists a $\sigma > 0$ such that

$$\mathcal{L}(\frac{\xi}{\sigma}) = \gamma_1 \text{ (abbreviated } \xi \in N(0,\sigma)).$$

(e) A family $(\xi_t)_{t\in T}$ of real-valued random variables is said to be a centred real-valued Gaussian process if for all $t_1, ..., t_n \in T$, $\alpha_1, ..., \alpha_n \in \mathbf{R}$ and every $n \in \mathbf{N}_+$, the sum

$$\xi = \Sigma_{k=1}^n \alpha_k \xi_{t_k}$$

is a centred Gaussian random variable.

**Example 3.4.1** Suppose $\mid x \mid = \sqrt{x_1^2 + ... + x_n^2}$ if $x = (x_1, ..., x_n) \in \mathbf{R}^n$. We claim that

$$\lim_{k\to\infty} \int_{\mathbf{R}^n} \prod_{i=1}^n (1 + \frac{x_i + x_i^2}{4k})^k d\gamma_n(x) = \sqrt{2}^n e^{\frac{n}{16}}$$

To prove this claim we first use that $e^t \geq 1 + t$ for every real $t$ and have for each fixed $i \in \{1, ..., n\}$,

$$1 + \frac{x_i + x_i^2}{4k} \leq e^{\frac{x_i + x_i^2}{4k}}.$$

Moreover, if $k \in \mathbf{N}_+$, then

$$1 + \frac{x_i + x_i^2}{4k} = \frac{1}{4k}((x_i + \frac{1}{2})^2 + 4k - \frac{1}{4}) \geq 0$$

and we conclude that

$$(1 + \frac{x_i + x_i^2}{4k})^k \leq e^{\frac{x_i + x_i^2}{4}}.$$

Thus, for any $k \in \mathbf{N}_+$,

$$0 \leq f_k(x) =_{def} \prod_{i=1}^n (1 + \frac{x_i + x_i^2}{4k})^k \leq \prod_{i=1}^n e^{\frac{x_i + x_i^2}{4}} =_{def} g(x)$$

where $g \in \mathcal{L}^1(\gamma_n)$ since

$$\int_{\mathbf{R}^n} g(x) d\gamma_n(x) = \int_{\mathbf{R}^n} \prod_{i=1}^n e^{\frac{x_i - x_i^2}{4}} \frac{dx}{\sqrt{2\pi}^n} = \{\text{Tonelli's Theorem}\} =$$

$$\prod_{i=1}^{n}\int_{\mathbf{R}} e^{\frac{x_i - x_i^2}{4}}\frac{dx_i}{\sqrt{2\pi}} = \sqrt{2}^n e^{\frac{n}{16}}.$$

Moreover,

$$\lim_{k\to\infty} f_k(x) = g(x)$$

and by dominated convergence we get

$$\lim_{k\to\infty}\int_{\mathbf{R}^n}\prod_{i=1}^{n}(1+\frac{x_i+x_i^2}{4k})^k d\gamma_n(x) = \int_{\mathbf{R}^n} g(x)d\gamma_n(x) = \sqrt{2}^n e^{\frac{n}{16}}.$$

**Exercises**

1. Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be two $\sigma$-finite positive measure spaces. Let $f\in L^1(\mu)$ and $g\in L^1(\nu)$ and define $h(x,y)=f(x)g(y)$, $(x,y)\in X\times Y$. Prove that $h\in L^1(\mu\times\nu)$ and

$$\int_{X\times Y} hd(\mu\times\nu) = \int_X fd\mu\int_Y gd\nu.$$

2. Let $(X,\mathcal{M},\mu)$ be a $\sigma$-finite positive measure space and $f: X\to[0,\infty[$ a measurable function. Prove that

$$\int_X fd\mu = (\mu\times m)(\{(x,y);\ 0<y<f(x),\ x\in X\}).$$

3. Let $(X,\mathcal{M},\mu)$ be a $\sigma$-finite positive measure space and $f: X\to\mathbf{R}$ a measurable function. Prove that $(\mu\times m)(\{(x,f(x));\ x\in X\ \})=0$.

4. Let $E\in\mathcal{R}_2^-$ and $E\subseteq[0,1]\times[0,1]$. Suppose $m(E_x)\leq\frac{1}{2}$ for $m$-almost all $x\in[0,1]$. Show that

$$m(\{y\in[0,1]\,;m(E^y)=1\})\leq\frac{1}{2}.$$

5. Let $c$ be the counting measure on $\mathbf{R}$ restricted to $\mathcal{R}$ and

$$D = \{(x,x);\ x \in \mathbf{R}\}.$$

Define for every $A \in (\mathcal{R} \boxtimes \mathcal{R}) \cup \{D\}$,

$$\mu(A) = \int_{\mathbf{R}} (\int_{\mathbf{R}} \chi_A(x,y)dv_1(x))dc(y)$$

and

$$\nu(A) = \int_{\mathbf{R}} (\int_{\mathbf{R}} \chi_A(x,y)dc(y))dv_1(x).$$

(a) Prove that $\mu$ and $\nu$ agree on $\mathcal{R} \boxtimes \mathcal{R}$.
(b) Prove that $\mu(D) \neq \nu(D)$.

6. Let $I = ]0,1[$ and

$$h(x,y) = \frac{x^2 - y^2}{(x^2+y^2)^2},\ (x,y) \in I \times I.$$

Prove that

$$\int_I (\int_I h(x,y)dy)dx = \frac{\pi}{4},$$
$$\int_I (\int_I h(x,y)dx)dy = -\frac{\pi}{4}$$

and

$$\int_{I\times I} \mid h(x,y) \mid dxdy = \infty.$$

7. For $t > 0$ and $x \in \mathbf{R}$ let

$$g(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

and

$$h(t,x) = \frac{\partial g}{\partial t}.$$

Given $a > 0$, prove that

$$\int_{-\infty}^{\infty} (\int_a^{\infty} h(t,x)dt)dx = -1$$

and

$$\int_a^\infty (\int_{-\infty}^\infty h(t,x)dx)dt = 0$$

and conclude that

$$\int_{[a,\infty[\times \mathbf{R}} \mid h(t,x) \mid dtdx = \infty.$$

(Hint: First prove that

$$\int_{-\infty}^\infty g(t,x)dx = 1$$

and

$$\frac{\partial g}{\partial t} = \frac{1}{2}\frac{\partial^2 g}{\partial x^2}.)$$

8. Given $f \in L^1(m)$, let

$$g(x) = \frac{1}{2}\int_{x-1}^{x+1} f(t)dt,\ x \in \mathbf{R}.$$

Prove that

$$\int_{\mathbf{R}} \mid g(x) \mid dx \leq \int_{\mathbf{R}} \mid f(x) \mid dx.$$

9. Let $I = [0,1]$ and suppose $f : I \to \mathbf{R}$ is a Lebesgue measurable function such that

$$\int_{I\times I} \mid f(x) - f(y) \mid dxdy < \infty.$$

Prove that

$$\int_I \mid f(x) \mid dx < \infty.$$

10. Suppose $A \in \mathcal{R}^-$ and $f \in L^1(m)$. Set

$$g(x) = \int_{\mathbf{R}} \frac{d(y,A)f(y)}{\mid x-y \mid^2}dy,\ x \in \mathbf{R}.$$

Prove that

$$\int_A \mid g(x) \mid dx < \infty.$$

11. Suppose that the functions $f, g : \mathbf{R} \to [0, \infty[$ are Lebesgue measurable and introduce $\mu = fm$ and $\nu = gm$. Prove that the measures $\mu$ and $\nu$ are $\sigma$-finite and

$$(\mu \times \nu)(E) = \int_E f(x)g(y)dxdy \text{ if } E \in \mathcal{R}^- \otimes \mathcal{R}^-.$$

12. Suppose $\mu$ is a finite positive Borel measure on $\mathbf{R}^n$ and $f : \mathbf{R}^n \to \mathbf{R}$ a Borel measurable function. Set $g(x, y) = f(x) - f(y)$, $x, y \in \mathbf{R}^n$. Prove that $f \in L^1(\mu)$ if and only if $g \in L^1(\mu \times \mu)$.

13. A random variable $\xi$ is non-negative and possesses the distribution function $F(x) = P[\xi \leq x]$. Prove that$E[\xi] = \int_0^\infty (1 - F(x))dx$.

14. Let $(X, d)$ be a metric space and suppose $Y \in \mathcal{B}(X)$. Then $Y$ equipped with the metric $d_{|Y \times Y}$ is a metric space. Prove that

$$\mathcal{B}(Y) = \{A \cap Y;\ A \in \mathcal{B}(X)\}.$$

15. The continuous bijection $f : (X, d) \to (Y, e)$ has a continuous inverse. Prove that $f(A) \in \mathcal{B}(Y)$ if $A \in \mathcal{B}(X)$

16. A real-valued function $f(x, y), x, y \in \mathbf{R}$, is a Borel function of $x$ for every fixed $y$ and a continuous function of $y$ for every fixed $x$. Prove that $f$ is a Borel function. Is the same conclusion true if we only assume that $f(x, y)$ is a real-valued Borel function in each variable separately?

17. Suppose $a > 0$ and

$$\mu_a = e^{-a} \sum_{n=0}^{\infty} \frac{a^n}{n!} \delta_n$$

where $\delta_n(A) = \chi_A(n)$ if $n \in \mathbf{N} = \{0, 1, 2, ...\}$ and $A \subseteq \mathbf{N}$. Prove that

$$(\mu_a \times \mu_b)s^{-1} = \mu_{a+b}$$

for all $a, b > 0$, if $s(x, y) = x + y$, $x, y \in \mathbf{N}$.

18. Suppose

$$f(t) = \int_0^\infty \frac{xe^{-x^2}}{x^2 + t^2} dx, \ t > 0.$$

Compute

$$\lim_{t \to 0+} f(t) \text{ and } \int_0^\infty f(t)dt.$$

Finally, prove that $f$ is differentiable.

19. Suppose

$$f(t) = \int_0^\infty e^{-tx} \frac{\ln(1 + x)}{1 + x} dx, \ t > 0.$$

a) Show that $\int_0^\infty f(t)dt < \infty$.
b) Show that $f$ is infinitely many times differentiable.

20. Suppose $f$ is Lebesgue integrabel on $]0, 1[$. (a) Show that the function $g(x) = \int_x^1 t^{-1} f(t)dt$, $0 < x < 1$, is continuous. (b) Prove that $\int_0^1 g(x)dx = \int_0^1 f(x)dx$.

### 3.5 Change of Variables in Volume Integrals

If $T$ is a non-singular $n$ by $n$ matrix with real entries, we claim that

$$T(v_n) = \frac{1}{\mid \det T \mid} v_n$$

(here $T$ is viewed as a linear map of $\mathbf{R}^n$ into $\mathbf{R}^n$). Remembering Corollary 3.1.3 this means that the following linear change of variables formula holds, viz.

$$\int_{\mathbf{R}^n} f(Tx)dx = \frac{1}{\mid \det T \mid} \int_{\mathbf{R}^n} f(x)dx \text{ all } f \in C_c(\mathbf{R}^n).$$

The case $n = 1$ is obvious. Moreover, by Fubini's Theorem the linear change of variables formula is true for arbitrary $n$ in the following cases:

(a) $Tx = (x_{\pi(1)}, ..., x_{\pi(n)})$, where $\pi$ is a permutation of the numbers $1, ..., n$.

(b) $Tx = (\alpha x_1, x_2, ..., x_n)$, where $\alpha$ is a non-zero real number.

(c) $Tx = (x_1 + x_2, x_2, ..., x_n)$.

Recall from linear algebra that every non-singular $n$ by $n$ matrix $T$ can be row-reduced to the identity matrix, that is $T$ can by written as the product of finitely many transformations of the types in (a),(b), and (c). This proves the above linear change of variables formula.

Our main objective in this section is to prove a more general change of variable formula. To this end let $\Omega$ and $\Gamma$ be open subsets of $\mathbf{R}^n$ and $G : \Omega \to \Gamma$ a $C^1$ diffeomorphism, that is $G = (g_1, ..., g_n)$ is a bijective continuously differentiable map such that the matrix $G'(x) = (\frac{\partial g_i}{\partial x_j}(x))_{1 \leq i,j \leq n}$ is non-singular for each $x \in \Omega$. The inverse function theorem implies that $G^{-1} : \Gamma \to \Omega$ is a $C^1$ diffeomorphism $[DI]$.

**Theorem 3.5.1.** *If $f$ is a non-negative Borel function in $\Omega$, then*

$$\int_\Gamma f(x)dx = \int_\Omega f(G(x)) \mid \det G'(x) \mid dx.$$

The proof of Theorem 3.5.1 is based on several lemmas.

Throughout, $\mathbf{R}^n$ is equipped with the metric

$$d_n(x, y) = \max_{1 \leq k \leq n} \mid x_k - y_k \mid .$$

Let $K$ be a compact convex subset of $\Omega$. Then if $x, y \in K$ and $1 \leq i \leq n$,

$$g_i(x) - g_i(y) = \int_0^1 \frac{d}{dt} g_i(y + t(x - y))dt$$

$$= \int_0^1 \Sigma_{k=1}^n \frac{\partial g_i}{\partial x_k}(y + t(x - y))(x_k - y_k)dt$$

and we get

$$d_n(G(x), G(y)) \leq M(G,K) d_n(x,y)$$

where

$$M(G,K) = \max_{1 \leq i \leq n} \Sigma_{k=1}^n \max_{z \in K} \mid \frac{\partial g_i}{\partial x_k}(z) \mid .$$

Thus if $\bar{B}(a;r)$ is a closed ball contained in $K$,

$$G(\bar{B}(a;r)) \subseteq \bar{B}(G(a); M(G,K)r).$$

**Lemma 3.5.1.** *Let $(Q_k)_{k=1}^{\infty}$ be a sequence of closed balls contained in $\Omega$ such that*

$$Q_{k+1} \subseteq Q_k$$

*and*

$$\text{diam } Q_k \to 0 \text{ as } k \to \infty.$$

*Then, there is a unique point $a$ belonging to each $Q_k$ and*

$$\limsup_{n \to \infty} \frac{v_n(G(Q_k))}{v_n(Q_k)} \leq \mid \det G'(a) \mid .$$

PROOF. The existence of a point $a$ belonging to each $Q_k$ is an immediate consequence of Theorem 3.1.2. The uniqueness is also obvious since the diameter of $Q_k$ converges to 0 as $k \to \infty$. Set $T = G'(a)$ and $F = T^{-1}G$. Then, if $Q_k = \bar{B}(x_k; r_k)$,

$$v_n(G(Q_k)) = v_n(T(T^{-1}G(Q_k))) = \mid \det T \mid \ v_n(T^{-1}G(\bar{B}(x_k; r_k)))$$

$$\leq \mid \det T \mid v_n(\bar{B}(T^{-1}G(x_k); M(T^{-1}G; Q_k)r_k) = \mid \det T \mid \ M(T^{-1}G; Q_k)^n v_n(Q_k).$$

Since

$$\lim_{k \to \infty} M(T^{-1}G; Q_k) = 1$$

the lemma follows at once.

**Lemma 3.5.2.** *Let $Q$ be a closed ball contained in $\Omega$. Then*

$$v_n(G(Q)) \leq \int_Q \mid \det G'(x) \mid dx.$$

PROOF. Suppose there is a closed ball $Q$ contained in $\Omega$ such that

$$v_n(G(Q)) > \int_Q \mid \det G'(x) \mid dx.$$

This will lead us to a contradiction as follows.

Choose $\varepsilon > 0$ such that

$$v_n(G(Q)) \geq (1+\varepsilon) \int_Q \mid \det G'(x) \mid dx.$$

Let $Q = \cup_1^{2^n} B_k$ where $B_1, ..., B_{2^n}$ are mutually almost disjoint closed balls with the same volume. If

$$v_n(G(B_k)) < (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx, \;\; k = 1, ..., 2^n$$

we get

$$v_n(G(Q)) \leq \Sigma_{k=1}^{2^n} v_n(G(B_k))$$

$$< \Sigma_{k=1}^{2^n} (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx = (1+\varepsilon) \int_Q \mid \det G'(x) \mid dx$$

which is a contradiction. Thus

$$v_n(G(B_k)) \geq (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx$$

for some $k$. By induction we obtain a sequence $(Q_k)_{k=1}^{\infty}$ of closed balls contained in $\Omega$ such that

$$Q_{k+1} \subseteq Q_k,$$

$$\text{diam } Q_k \to 0 \text{ as } k \to \infty$$

and

$$v_n(G(Q_k)) \geq (1+\varepsilon) \int_{Q_k} \mid \det G'(x) \mid dx.$$

But applying Lemma 3.5.1 we get a contradiction.

PROOF OF THEOREM 3.5.1. Let $U \subseteq \Omega$ be open and write $U = \cup_{i=1}^{\infty} Q_i$ where the $Q_i's$ are almost disjoint cubes as in Theorem 3.1.5. Then

$$v_n(G(U)) \leq \Sigma_{i=1}^{\infty} v_n(G(Q_i)) \leq \Sigma_{i=1}^{\infty} \int_{Q_i} \mid \det G'(x) \mid dx$$

$$= \int_U \mid \det G'(x) \mid dx.$$

Using Theorem 3.1.3 we now have that

$$v_n(G(E)) \leq \int_E \mid \det G'(x) \mid dx$$

for each Borel set $E \subseteq \Omega$. But then

$$\int_\Gamma f(x)dx \leq \int_\Omega f(G(x)) \mid \det G'(x) \mid dx$$

for each simple Borel measurable function $f \geq 0$ and, accordingly from this and monotone convergence, the same inequality holds for each non-negative Borel function $f$. But the same line of reasoning applies to $G$ replaced by $G^{-1}$ and $f$ replaced by $f(G) \mid \det G' \mid$, so that

$$\int_\Omega f(G(x)) \mid \det G'(x) \mid dx \leq \int_\Gamma f(x) \mid \det G'(G^{-1}(x)) \mid\mid \det(G^{-1})'(x) \mid dx$$

$$= \int_\Gamma f(x)dx.$$

This proves the theorem.

**Example 3.5.1.** If $f : \mathbf{R}^2 \rightarrow [0, \infty]$ is $(\mathcal{R}_2, \mathcal{R}_{0,\infty})$-measurable and $0 < \varepsilon < R < \infty$, the substitution

$$G(r, \theta) = (r\cos\theta, r\sin\theta)$$

yields

$$\int_{\varepsilon < \sqrt{x_1^2 + x_2^2} < R} f(x_1, x_2)dx_1dx_2 = \lim_{\delta \rightarrow 0^+} \int_\varepsilon^R \int_\delta^{2\pi} f(r\cos\theta, r\sin\theta)rd\theta dr$$

$$= \int_{\varepsilon}^{R} \int_{0}^{2\pi} f(r\cos\theta, r\sin\theta) r d\theta dr$$

and by letting $\varepsilon \to 0$ and $R \to \infty$, we have

$$\int_{\mathbf{R}^2} f(x_1, x_2) dx_1 dx_2 = \int_0^\infty \int_0^{2\pi} f(r\cos\theta, r\sin\theta) r d\theta dr.$$

The purpose of the example is to show an analogue formula for volume measure in $\mathbf{R}^n$.

Let $S^{n-1} = \{x \in \mathbf{R}^n ; \mid x \mid = 1\}$ be the unit sphere in $\mathbf{R}^n$. We will define a so called surface area Borel measure $\sigma_{n-1}$ on $S^{n-1}$ such that

$$\int_{\mathbf{R}^n} f(x) dx = \int_0^\infty \int_{S^{n-1}} f(r\omega) r^{n-1} d\sigma_{n-1}(\omega) dr$$

for any $(\mathcal{R}_n, \mathcal{R}_{0,\infty})$-measurable function $f : \mathbf{R}^n \to [0, \infty]$ . To this end define $G : \mathbf{R}^n \setminus \{0\} \to ]0, \infty[ \times S^{n-1}$ by setting $G(x) = (r, \omega)$, where

$$r = \mid x \mid \text{ and } \omega = \frac{x}{\mid x \mid}.$$

Note that $G^{-1} : ]0, \infty[ \times S^{n-1} \to \mathbf{R}^n \setminus \{0\}$ is given by the equation

$$G^{-1}(r, \omega) = r\omega.$$

Moreover,

$$G^{-1}(]0, a] \times E) = aG^{-1}(]0, 1] \times E) \text{ if } a > 0 \text{ and } E \subseteq S^{n-1}.$$

If $E \in \mathcal{B}(S^{n-1})$ we therefore have that

$$v_n(G^{-1}(]0, a] \times E)) = a^n v_n(G^{-1}(]0, 1] \times E)).$$

We now define

$$\sigma_{n-1}(E) = n v_n(G^{-1}(]0, 1] \times E)) \text{ if } E \in \mathcal{B}(S^{n-1})$$

and

$$\rho(A) = \int_A r^{n-1} dr \text{ if } A \in \mathcal{B}(]0, \infty[).$$

Below, by abuse of language, we write $v_{n|\mathbf{R}^n\setminus\{0\}} = v_n$. Then, if $0 < a \leq b < \infty$ and $E \in \mathcal{B}(S^{n-1})$,

$$G(v_n)(]0,a] \times E) = \rho(]0,a])\sigma_{n-1}(E)$$

and

$$G(v_n)(]a,b] \times E) = \rho(]a,b])\sigma_{n-1}(E).$$

Thus, by Theorem 1.2.3,

$$G(v_n) = \rho \times \sigma_{n-1}$$

and the claim above is immediate.

To check the normalization constant in the definition of $\sigma_{n-1}$, first note that

$$v_n(\mid x \mid < R) = \int_0^R \int_{S^{n-1}} r^{n-1} d\sigma(\omega) dr = \frac{R^n}{n}\sigma_{n-1}(S^{n-1})$$

and we get

$$\frac{d}{dR} v_n(\mid x \mid < R) = R^{n-1}\sigma_{n-1}(S^{n-1}).$$

**Exercises**

1. Extend Theorem 3.5.1 to Lebesgue measurable functions.

2. The function $f : \mathbf{R} \to [0, \infty[$ is Lebesgue measurable and $\int_{\mathbf{R}} f dm = 1$. Determine all non-zero real numbers $\alpha$ such that $\int_{\mathbf{R}} h dm < \infty$, where

$$h(x) = \Sigma_{n=0}^{\infty} f(\alpha^n x + n), \; x \in \mathbf{R}.$$

3. Suppose

$$\mu(A) = \int_A \mid x \mid^n e^{-|x|^n} dx, \; A \in \mathcal{B}(\mathbf{R}^n),$$

where $\mid x \mid = \sqrt{x_1^2 + ... + x_n^2}$ if $x = (x_1, ..., x_n) \in \mathbf{R}^n$. Compute the limit

$$\lim_{\rho \to \infty} \rho^{-n} \ln \mu(\{x \in \mathbf{R}^n; \mid x \mid \geq \rho\}).$$

4. Compute the $n$-dimensional Lebesgue integral

$$\int_{|x|<1} \ln(1- \mid x \mid)dx$$

where $\mid x \mid$ denotes the Euclidean norm of the vector $x \in \mathbf{R}^n$. (Hint: $\sigma(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$.)

5. Suppose $f \in L^1(m_2)$. Show that $\lim_{n\to\infty} f(nx) = 0$ for $m_2$-almost all $x \in \mathbf{R}^2$.

$\downarrow\downarrow\downarrow$

## 3.6. Independence in Probability

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. The random variables $\xi_k : (\Omega, P) \to (S_k, \mathcal{S}_k)$, $k = 1, ..., n$ are said to be independent if

$$P_{(\xi_1,...,\xi_n)} = \times_{k=1}^{n} P_{\xi_k}.$$

A family $(\xi_i)_{i\in I}$ of random variables is said to be independent if $\xi_{i_1}, ..., \xi_{i_n}$ are independent for any $i_1, ...i_n \in I$ with $i_k \neq i_l$ if $k \neq l$. A family of events $(A_i)_{i\in I}$ is said to be independent if $(\chi_{A_i})_{i\in I}$ is a family of independent random variables. Finally a family $(\mathcal{A}_i)_{i\in I}$ of sub-$\sigma$-algebras of $\mathcal{F}$ is said to be independent if, for any $A_i \in \mathcal{A}_i$, $i \in I$, the family $(A_i)_{i\in I}$ is a family of independent events.

**Example 3.6.1.** Let $q \geq 2$ be an integer. A real number $\omega \in [0,1[$ has a $q$-adic expansion

$$\omega = \Sigma_{k=1}^{\infty} \frac{\xi_k^{(q)}}{q^k}.$$

The construction of the Cantor set shows that $(\xi_k^{(q)})_{k=1}^{\infty}$ is a sequence of independent random variables based on the probability space

$$([0,1[\,, v_{1|[0,1[}, \mathcal{B}([0,1[)).$$

**Example 3.6.2.** Let $(X, \mathcal{M}, \mu)$ be a positive measure space and let $A_i \in \mathcal{M}$, $i \in \mathbf{N}_+$, be such that

$$\Sigma_{i=1}^{\infty} \mu(A_i) < \infty.$$

The first Borel-Cantelli Lemma asserts that $\mu$-almost all $x \in X$ lie in $A_i$ for at most finitely many $i$. This result is an immediate consequence of the Beppo Levi Theorem since

$$\int_X \Sigma_{i=1}^{\infty} \chi_{A_i} d\mu = \Sigma_{i=1}^{\infty} \int_X \chi_{A_i} d\mu < \infty$$

implies that

$$\Sigma_{i=1}^{\infty} \chi_{A_i} < \infty \text{ a.e. } [\mu].$$

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and let $(A_i)_{i=1}^{\infty}$ be independent events such that

$$\Sigma_{i=1}^{\infty} P\left[A_i\right] = \infty.$$

The second Borel-Cantelli Lemma asserts that almost surely $A_i$ happens for infinitely many $i$.

To prove this, we use the inequality

$$1 + x \leq e^x, \; x \in \mathbf{R}$$

to obtain

$$P\left[\cap_{i=k}^{k+n} A_i^c\right] = \Pi_{i=k}^{k+n} P\left[A_i^c\right]$$

$$= \Pi_{i=k}^{k+n} (1 - P\left[A_i\right]) \leq \Pi_{i=k}^{k+n} e^{-P[A_i]} = e^{-\Sigma_{i=k}^{k+n} P[A_i]}.$$

By letting $n \to \infty$,

$$P\left[\cap_{i=k}^{\infty} A_i^c\right] = 0$$

or

$$P\left[\cup_{i=k}^{\infty} A_i\right] = 1.$$

But then

$$P\left[\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} A_i\right] = 1$$

and the second Borel-Cantelli Lemma is proved.

**Theorem 3.6.1**. *Suppose* $\xi_1, ..., \xi_n$ *are independent random variables and* $\xi_k \in N(0,1)$, $k = 1, ..., n$. *If* $\alpha_1, ..., \alpha_n \in R$, *then*

$$\Sigma_{k=1}^n \alpha_k \xi_k \in N(0, \Sigma_{k=1}^n \alpha_k^2)$$

PROOF. The case $\alpha_1, ..., \alpha_n = 0$ is trivial so assume $\alpha_k \neq 0$ for some $k$. We have for each open interval $A$,

$$P\left[\Sigma_{k=1}^n \alpha_k \xi_k \in A\right] = \int_{\Sigma_{k=1}^n \alpha_k x_k \in A} d\gamma_1(x_1)...d\gamma_1(x_n)$$

$$\int_{\Sigma_{k=1}^n \alpha_k x_k \in A} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x_1^2+...+x_n^2)} dx_1...dx_n.$$

Set $\sigma = \sqrt{\alpha_1^2 + ... + \alpha_n^2}$ and let $y = Gx$ be an orthogonal transformation such that

$$y_1 = \frac{1}{\sigma}(\alpha_1 x_1 + ... + \alpha_n x_n).$$

Then, since $\det G = 1$,

$$P\left[\Sigma_{k=1}^\infty \alpha_k \xi_k \in A\right] = \int_{\sigma y_1 \in A} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(y_1^2+...+y_n^2)} dy_1...dy_n$$

$$= \int_{\sigma y_1 \in A} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2} dy_1$$

where we used Fubini's theorem in the last step. The theorem is proved.

Finally, in this section, we prove a basic result about the existence of infinite product measures. Let $\mu_k$, $k \in \mathbf{N}_+$ be Borel probability measures in $\mathbf{R}$. The space $\mathbf{R}^{\mathbf{N}_+}$ is, by definition, the set of all sequences $x = (x_k)_{k=1}^\infty$ of real numbers. For each $k \in \mathbf{N}_+$, set $\pi_k(x) = x_k$. The $\sigma$-algebra $\mathcal{R}^{\mathbf{N}_+}$ is the least $\sigma$-algebra $\mathcal{S}$ of subsets of $\mathbf{R}^{\mathbf{N}_+}$ which makes all the projections $\pi_k : (\mathbf{R}^{\mathbf{N}_+}, \mathcal{S}) \to (\mathbf{R}, \mathcal{R})$, $k \in \mathbf{N}_+$, measurable. Below, $(\pi_1, ..., \pi_n)$ denotes the mapping of $\mathbf{R}^{\mathbf{N}_+}$ into $\mathbf{R}^n$ defined by the equation

$$(\pi_1, ..., \pi_n)(x) = (\pi_1(x), ..., \pi_n(x)).$$

**Theorem 3.6.1.** *There is a unique probability measure $\mu$ on $\mathcal{R}^{\mathbf{N}_+}$ such that*

$$\mu_{(\pi_1,...,\pi_n)} = \mu_1 \times ... \times \mu_n$$

*for every $n \in \mathbf{N}_+$.*

The measure $\mu$ in Theorem 3.6.1 is called the product of the measures $\mu_k$, $k \in \mathbf{N}_+$, and is often denoted by

$$\times_{k=1}^{\infty}\mu_k.$$

PROOF OF THEOREM 3.6.1. Let $(\Omega, P, \mathcal{F}) = ([0,1[\,, v_{1|[0,1[}, \mathcal{B}([0,1[)$ and set

$$\eta(\omega) = \Sigma_{k=1}^{\infty}\frac{\xi_k^{(2)}(\omega)}{2^k},\ \omega \in \Omega.$$

We already know that $P_\eta = P$. Now suppose $(k_i)_{i=1}^{\infty}$ is a strictly increasing sequence of positive integers and introduce

$$\eta' = \Sigma_{i=1}^{\infty}\frac{\xi_{k_i}^{(2)}(\omega)}{2^i},\ \omega \in \Omega.$$

Note that for each fixed positive integer $n$, the $\mathbf{R}^n$-valued maps $(\xi_1^{(2)}, ..., \xi_n^{(2)})$ and $(\xi_{k_1}^{(2)}, ..., \xi_{k_n}^{(2)})$ are $P$-equimeasurable. Thus, if $f : \Omega \to \mathbf{R}$ is continuous,

$$\int_\Omega f(\eta)dP = \lim_{n\to\infty}\int_\Omega f(\Sigma_{k=1}^{n}\frac{\xi_k^{(2)}(\omega)}{2^k})dP(\omega)$$

$$= \lim_{n\to\infty}\int_\Omega f(\Sigma_{i=1}^{n}\frac{\xi_{k_i}^{(2)}(\omega)}{2^i})dP(\omega) = \int_\Omega f(\eta')dP$$

and it follows that $P_{\eta'} = P_\eta = P$.

By induction, we define for each $k \in \mathbf{N}_+$ an infinite subset $N_k$ of the set $\mathbf{N}_+ \setminus \cup_{i=1}^{k-1} N_i$ such that the set $\mathbf{N}_+ \setminus \cup_{i=1}^{k} N_i$ contains infinitely many elements and define

$$\eta_k = \Sigma_{i=1}^{\infty}\frac{\xi_{n_{ik}}^{(2)}(\omega)}{2^i}$$

where $(n_{ik})_{i=1}^{\infty}$ is an enumeration of $N_k$. The map

$$\Psi(\omega) = (\eta_k(\omega))_{k=1}^{\infty}$$

is a measurable map of $(\Omega, \mathcal{F})$ into $(\mathbf{R}^{\mathbf{N}_+}, \mathcal{R}^{\mathbf{N}_+})$ and

$$P_\Psi = \times_{k=1}^{\infty} \lambda_i$$

where $\lambda_i = P$ for each $i \in \mathbf{N}_+$.

For each $i \in \mathbf{N}_+$ there exists a measurable map $\varphi_i$ of $(\Omega, \mathcal{F})$ into $(\mathbf{R}, \mathcal{R})$ such that $P_{\varphi_i} = \mu_i$ (see Section 1.6). The map

$$\Gamma(x) = (\varphi_i(x_i))_{i=1}^{\infty}$$

is a measurable map of $(\mathbf{R}^{\mathbf{N}_+}, \mathcal{R}^{\mathbf{N}_+})$ into itself and we get $\mu = (P_\Psi)_\Gamma$. This completes the proof of Theorem 3.6.1.

$\uparrow\uparrow\uparrow$

# CHAPTER 4

# MODES OF CONVERGENCE

## Introduction

In this chapter we will treat a variety of different sorts of convergence notions in measure theory. So called $L^2$-convergence is of particular importance.

## 4.1. Convergence in Measure, in $L^1(\mu)$, and in $L^2(\mu)$

Let $(X, \mathcal{M}, \mu)$ be a positive measure space and denote by $\mathcal{F}(X)$ the class of measurable functions $f : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$. For any $f \in \mathcal{F}(X)$, set

$$\parallel f \parallel_1 = \int_X \mid f(x) \mid d\mu(x)$$

and

$$\parallel f \parallel_2 = \sqrt{\int_X f^2(x) d\mu(x)}.$$

The Cauchy-Schwarz inequality states that

$$\int_X \mid fg \mid d\mu \leq \parallel f \parallel_2 \parallel g \parallel_2 \text{ if } f, g \in \mathcal{F}(X).$$

To prove this, without loss of generality, it can be assumed that

$$0 < \parallel f \parallel_2 < \infty \text{ and } 0 < \parallel g \parallel_2 < \infty.$$

We now use the inequality

$$\alpha\beta \leq \frac{1}{2}(\alpha^2 + \beta^2), \ \alpha, \beta \in \mathbf{R}$$

to obtain

$$\int_X \frac{|f|}{\| f \|_2} \frac{|g|}{\| g \|_2} d\mu \leq \int \frac{1}{2}(\frac{f^2}{\| f \|_2^2} + \frac{g^2}{\| g \|_2^2}) d\mu = 1$$

and the Cauchy-Schwarz inequality is immediate.

If not otherwise stated, in this section $p$ is a number equal to 1 or 2. If it is important to emphasize the underlying measure $\| f \|_p$ is written $\| f \|_{p,\mu}$ .

We now define

$$\mathcal{L}^p(\mu) = \{f \in \mathcal{F}(X);\ \ \| f \|_p < \infty\}.$$

The special case $p = 1$ has been introduced earlier. We claim that the following so called triangle inequality holds, viz.

$$\| f + g \|_p \leq \| f \|_p + \| g \|_p \text{ if } f, g \in \mathcal{L}^p(\mu).$$

The case $p = 1$, follows by $\mu$-integration of the relation

$$| f + g | \leq | f | + | g | .$$

To prove the case $p = 2$, we use the Cauchy-Schwarz inequality and have

$$\| f + g \|_2^2 \leq \| | f | + | g | \|_2^2$$

$$= \| f \|_2^2 + 2 \int_X | fg | \, d\mu + \| g \|_2^2$$

$$\leq \| f \|_2^2 + 2 \| f \|_2 \| g \|_2 + \| g \|_2^2 = (\| f \|_2 + \| g \|_2)^2$$

and the triangle inequality is immediate.

Suppose $f, g \in \mathcal{L}^p(\mu)$. The functions $f$ and $g$ are equal almost everywhere with respect to $\mu$ if $\{f \neq g\} \in \mathcal{Z}_\mu$. This is easily seen to be an equivalence relation and the set of all equivalence classes is denoted by $L^p(\mu)$. Below we consider the elements of $L^p(\mu)$ as members of $\mathcal{L}^p(\mu)$ and two members of $L^p(\mu)$ are identified if they are equal a.e. $[\mu]$. From this convention it is straight-forward to define $f + g$ and $\alpha f$ for all $f, g \in L^p(\mu)$ and $\alpha \in \mathbf{R}$ and the function $d^{(p)}(f, g) = \| f - g \|_p$ is a metric on $L^p(\mu)$. Convergence in the metric space $L^p(\mu) = (L^p(\mu), d^{(p)})$ is called convergence in $L^p(\mu)$. A sequence $(f_k)_{k=1}^\infty$ in $\mathcal{F}(X)$ converges in measure to a function $f \in \mathcal{F}(X)$ if

$$\lim_{k\to\infty} \mu(| f_k - f | > \varepsilon) = 0 \text{ all } \varepsilon > 0.$$

If the sequence $(f_k)_{k=1}^{\infty}$ in $\mathcal{F}(X)$ converges in measure to a function $f \in \mathcal{F}(X)$ as well as to a function $g \in \mathcal{F}(X)$, then $f = g$ a.e. $[\mu]$ since

$$\{\mid f-g \mid > \varepsilon\} \subseteq \left\{\mid f-f_k \mid > \frac{\varepsilon}{2}\right\} \cup \left\{\mid f_k - g \mid > \frac{\varepsilon}{2}\right\}$$

and

$$\mu(\mid f-g \mid > \varepsilon) \leq \mu(\mid f-f_k \mid > \frac{\varepsilon}{2}) + \mu(\mid f_k - g \mid > \frac{\varepsilon}{2})$$

for every $\varepsilon > 0$ and positive integer $k$. A sequence $(f_k)_{k=1}^{\infty}$ in $\mathcal{F}(X)$ is said to be Cauchy in measure if for every $\varepsilon > 0$,

$$\mu(\mid f_k - f_n \mid > \varepsilon) \to 0 \text{ as } k, n \to \infty.$$

By the Markov inequality, a Cauchy sequence in $L^p(\mu)$ is Cauchy in measure.

**Example 4.1.1.** (a) If $f_k = \sqrt{k}\chi_{[0,\frac{1}{k}]}$, $k \in \mathbf{N}_+$, then

$$\| f_k \|_{2,m} = 1 \text{ and } \| f_k \|_{1,m} = \frac{1}{\sqrt{k}}.$$

Thus $f_k \to 0$ in $L^1(m)$ as $k \to \infty$ but $f_k \nrightarrow 0$ in $L^2(m)$ as $k \to \infty$.

(b) $L^1(m) \nsubseteq L^2(m)$ since

$$\chi_{[1,\infty[}(x)\frac{1}{\mid x \mid} \in L^2(m) \setminus L^1(m)$$

and $L^2(m) \nsubseteq L^1(m)$ since

$$\chi_{]0,1]}(x)\frac{1}{\sqrt{\mid x \mid}} \in L^1(m) \setminus L^2(m).$$

**Theorem 4.1.1.** *Suppose $p = 1$ or 2.*

(a) *Convergence in $L^p(\mu)$ implies convergence in measure.*

(b) *If $\mu(X) < \infty$, then $L^2(\mu) \subseteq L^1(\mu)$ and convergence in $L^2(\mu)$ implies convergence in $L^1(\mu)$.*

**Proof.** (a) Suppose the sequence $(f_n)_{n=1}^{\infty}$ converges to $f$ in $L^p(\mu)$ and let $\varepsilon > 0$. Then, by the Markov inequality,

$$\mu(\mid f_n - f \mid \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int_X \mid f_n - f \mid^p d\mu = \frac{1}{\varepsilon^p} \parallel f_n - f \parallel_p^p$$

and (a) follows at once.

(b) The Cauchy-Schwarz inequality gives for any $f \in \mathcal{F}(X)$,

$$(\int_X \mid f \mid \cdot 1 d\mu)^2 \leq \int_X f^2 d\mu \int_X 1 d\mu$$

or

$$\parallel f \parallel_1 \leq \parallel f \parallel_2 \sqrt{\mu(X)}$$

and Part (b) is immediate.

**Theorem 4.1.2.** *Suppose* $f_n \in \mathcal{F}(X)$, $n \in \mathbf{N}_+$.

(a) *If $(f_n)_{n=1}^{\infty}$ is Cauchy in measure, there is a measurable function $f : X \to \mathbf{R}$ such that $f_n \to f$ in measure as $n \to \infty$ and a strictly increasing sequence of positive integers $(n_j)_{j=1}^{\infty}$ such that $f_{n_j} \to f$ a.e. $[\mu]$ as $j \to \infty$.*

(b) *If $\mu$ is a finite positive measure and $f_n \to f \in \mathcal{F}(X)$ a.e. $[\mu]$ as $n \to \infty$, then $f_n \to f$ in measure.*

(c) **(Egoroff's Theorem)** *If $\mu$ is a finite positive measure and $f_n \to f \in \mathcal{F}(X)$ a.e. $[\mu]$ as $n \to \infty$, then for every $\varepsilon > 0$ there exists $E \in \mathcal{M}$ such that $\mu(E) < \varepsilon$* and

$$\sup_{\substack{k \geq n \\ x \in E^c}} \mid f_k(x) - f(x) \mid \to 0 \text{ as } n \to \infty.$$

PROOF. (a) For each positive integer $j$, there is a positive integer $n_j$ such that

$$\mu(\mid f_k - f_l \mid > 2^{-j}) < 2^{-j}, \text{ all } k, l \geq n_j.$$

There is no loss of generality to assume that $n_1 < n_2 < ...$ . Set

$$E_j = \left\{ \mid f_{n_j} - f_{n_{j+1}} \mid > 2^{-j} \right\}$$

and

$$F_k = \cup_{j=k}^{\infty} E_j.$$

If $x \in F_k^c$ and $i \geq j \geq k$

$$\mid f_{n_i}(x) - f_{n_j}(x) \mid \leq \sum_{j \leq l < i} \mid f_{n_{l+1}}(x) - f_{n_l}(x) \mid$$

$$\leq \sum_{j \leq l < i} 2^{-l} < 2^{-j+1}$$

and we conclude that $(f_{n_j}(x))_{j=1}^{\infty}$ is a Cauchy sequence for every $x \in F_k^c$. Let $G = \cup_{k=1}^{\infty} F_k^c$ and note that for every fixed positive integer $k$,

$$\mu(G^c) \leq \mu(F_k) < \sum_{j=k}^{\infty} 2^{-j} = 2^{-k+1}.$$

Thus $G^c$ is a $\mu$-null set. We now define $f(x) = \lim_{j \to \infty} f_{n_j}(x)$ if $x \in G$ and $f(x) = 0$ if $x \notin G$.

We next prove that the sequence $(f_n)_{n=1}^{\infty}$ converges to $f$ in measure. If $x \in F_k^c$ and $j \geq k$ we get

$$\mid f(x) - f_{n_j}(x) \mid \leq 2^{-j+1}.$$

Thus, if $j \geq k$

$$\mu(\mid f - f_{n_j} \mid > 2^{-j+1}) \leq \mu(F_k) < 2^{-k+1}.$$

Since

$$\mu(\mid f_n - f \mid > \varepsilon) \leq \mu(\mid f_n - f_{n_j} \mid > \frac{\varepsilon}{2}) + \mu(\mid f_{n_j} - f \mid > \frac{\varepsilon}{2})$$

if $\varepsilon > 0$, Part (a) follows at once.

(b) For each $\varepsilon > 0$,

$$\mu(\mid f_n - f \mid > \varepsilon) = \int_X \chi_{]\varepsilon,\infty[}(\mid f_n - f \mid) d\mu$$

and Part (c) follows from the Lebesgue Dominated Convergence Theorem.

(c) Set for fixed $k, n \in \mathbf{N}_+$,

$$E_{kn} = \cup_{j=n}^{\infty} \left\{ \mid f_j - f \mid > \frac{1}{k} \right\}.$$

We have

$$\cap_{n=1}^{\infty} E_{kn} \in Z_\mu$$

and since $\mu$ is a finite measure

$$\mu(E_{kn}) \to 0 \text{ as } n \to \infty.$$

Given $\varepsilon > 0$ pick $n_k \in \mathbf{N}_+$ such that $\mu(E_{kn_k}) < \varepsilon 2^{-k}$. Then, if $E = \cup_{k=1}^{\infty} E_{kn_k}$, $\mu(E) < \varepsilon$. Moreover, if $x \notin E$ and $j \geq n_k$

$$\mid f_j(x) - f(x) \mid \leq \frac{1}{k}.$$

The theorem is proved.

**Corollary 4.1.1.** *The spaces $L^1(\mu)$ and $L^2(\mu)$ are complete.*

PROOF. Suppose $p = 1$ or $2$ and let $(f_n)_{n=1}^{\infty}$ be a Cauchy sequence in $L^p(\mu)$. We know from the previous theorem that there exists a subsequence $(f_{n_j})_{j=1}^{\infty}$ which converges pointwise to a function $f \in \mathcal{F}(X)$ a.e. $[\mu]$. Thus, by Fatou's Lemma,

$$\int_X \mid f - f_k \mid^p d\mu \leq \liminf_{j \to \infty} \int_X \mid f_{n_j} - f_k \mid^p d\mu$$

and it follows that $f - f_k \in L^p(\mu)$ and, hence $f = (f - f_k) + f_k \in L^p(\mu)$. Moreover, we have that $\parallel f - f_k \parallel_p \to 0$ as $k \to \infty$. This concludes the proof of the theorem.

**Corollary 4.1.2.** *Suppose $\xi_n \in N(0, \sigma_n^2)$, $n \in \mathbf{N}_+$, and $\xi_n \to \xi$ in $L^2(P)$ as $n \to \infty$. Then $\xi$ is a centred Gaussian random variable.*

PROOF. We have that $\| \xi_n \|_2 = \sqrt{E\left[\xi_n^2\right]} = \sigma_n$ and $\| \xi_n \|_2 \to \| \xi \|_2 =_{def} \sigma$ as $n \to \infty$.

Suppose $f$ is a bounded continuous function on $\mathbf{R}$. Then, by dominated convergence,

$$E\left[f(\xi_n)\right] = \int_{\mathbf{R}} f(\sigma_n x) d\gamma_1(x) \to \int_{\mathbf{R}} f(\sigma x) d\gamma_1(x)$$

as $n \to \infty$. Moreover, there exists a subsequence $(\xi_{n_k})_{k=1}^{\infty}$ which converges to $\xi$ a.s. Hence, by dominated convergence

$$E\left[f(\xi_{n_k})\right] \to E\left[f(\xi)\right]$$

as $k \to \infty$ and it follows that

$$E\left[f(\xi)\right] = \int_{\mathbf{R}} f(\sigma x) d\gamma_1(x).$$

By using Corollary 3.1.3 the theorem follows at once.

**Theorem 4.1.3.** *Suppose $X$ is a standard space and $\mu$ a positive $\sigma$-finite Borel measure on $X$. Then the spaces $L^1(\mu)$ and $L^2(\mu)$ are separable.*

PROOF. Let $(E_k)_{k=1}^{\infty}$ be a denumerable collection of Borel sets with finite $\mu$-measures and such that $E_k \subseteq E_{k+1}$ and $\cup_{k=1}^{\infty} E_k = X$. Set $\mu_k = \chi_{E_k}\mu$ and first suppose that the set $D_k$ is at most denumerable and dense in $L^p(\mu_k)$ for every $k \in \mathbf{N}_+$. Without loss of generality it can be assumed that each member of $D_k$ vanishes off $E_k$. By monotone convergence

$$\int_X f d\mu = \lim_{k \to \infty} \int_X f d\mu_k,\ f \geq 0 \text{ measurable},$$

and it follows that the set $\cup_{k=1}^{\infty} D_k$ is at most denumerable and dense in $L^p(\mu)$.

From now on we can assume that $\mu$ is a finite positive measure. Let $A$ be an at most denumerable dense subset of $X$ and and suppose the subset $\{r_n;\ n \in \mathbf{N}_+,\}$ of $]0, \infty[$ is dense in $]0, \infty[$. Furthermore, denote by $\mathcal{U}$ the class of all open sets which are finite unions of open balls of the type $B(a, r_n)$, $a \in A$, $n \in \mathbf{N}_+$. If $U$ is any open subset of $X$

$$U = \cup [V : V \subseteq U \text{ and } V \in \mathcal{U}]$$

and, hence, by the Ulam Theorem

$$\mu(U) = \sup \{\mu(V);\ V \in \mathcal{U} \text{ and } V \subseteq U\}.$$

Let $\mathcal{K}$ be the class of all functions which are finite sums of functions of the type $\kappa\chi_U$, where $\kappa$ is a positive rational number and $U \in \mathcal{U}$. It follows that $\mathcal{K}$ is at most denumerable.

Suppose $\varepsilon > 0$ and that $f \in L^p(\mu)$ is non-negative. There exists a sequence of simple measurable functions $(\varphi_i)_{i=1}^{\infty}$ such that

$$0 \leq \varphi_i \uparrow f \text{ a.e. } [\mu].$$

Since $\mid f - \varphi_i \mid^p \leq f^p$, the Lebesgue Dominated Convergence Theorem shows that $\parallel f - \varphi_k \parallel_p < \frac{\varepsilon}{2}$ for an appropriate $k$. Let $\alpha_1, ..., \alpha_l$ be the distinct positive values of $\varphi_k$ and set

$$C = 1 + \Sigma_{k=1}^{l} \alpha_k.$$

Now for each fixed $j \in \{1, ..., l\}$ we use Theorem 3.1.3 to get an open $U_j \supseteq \varphi_k^{-1}(\{\alpha_j\})$ such that $\parallel \chi_{U_j} - \chi_{\varphi_k^{-1}(\{\alpha_j\})} \parallel_p < \frac{\varepsilon}{4C}$ and from the above we get a $V_j \in \mathcal{U}$ such that $V_j \subseteq U_j$ and $\parallel \chi_{U_j} - \chi_{V_j} \parallel_p < \frac{\varepsilon}{4C}$. Thus

$$\parallel \chi_{V_j} - \chi_{\varphi_k^{-1}(\{\alpha_j\})} \parallel_p < \frac{\varepsilon}{2C}$$

and

$$\parallel f - \Sigma_{k=1}^{l} \alpha_j \chi_{V_j} \parallel_p < \varepsilon$$

Now it is simple to find a $\psi \in \mathcal{K}$ such that $\parallel f - \psi \parallel_p < \varepsilon$. From this we deduce that the set

$$\mathcal{K} - \mathcal{K} = \{g - h;\ g, h \in \mathcal{K}\}$$

is at most denumerable and dense in $L^p(\mu)$.

The set of all real-valued and infinitely many times differentiable functions defined on $\mathbf{R}^n$ is denoted by $C^{(\infty)}(\mathbf{R}^n)$ and

$$C_c^{(\infty)}(\mathbf{R}^n) = \left\{ f \in C^{(\infty)}(\mathbf{R}^n);\ \mathrm{supp} f \text{ compact} \right\}.$$

Recall that the support $\mathrm{supp} f$ of a real-valued continuous function $f$ defined on $\mathbf{R}^n$ is the closure of the set of all $x$ where $f(x) \neq 0$. If

$$f(x) = \prod_{k=1}^{n} \left\{ \varphi(1 + x_k)\varphi(1 - x_k) \right\},\ x = (x_1, ..., x_n) \in \mathbf{R}^n$$

where $\varphi(t) = \exp(-t^{-1})$, if $t > 0$, and $\varphi(t) = 0$, if $t \leq 0$, then $f \in C_c^{\infty}(\mathbf{R}^n)$ .

The proof of the previous theorem also gives Part (a) of the following

**Theorem 4.1.4.** *Suppose $\mu$ is a positive Borel measure in $\mathbf{R}^n$ such that $\mu(K) < \infty$ for every compact subset $K$ of $\mathbf{R}^n$. The following sets are dense in $L^1(\mu)$, and $L^2(\mu)$ :*

(a) *the linear span of the functions*

$$\chi_I,\ I \text{ open bounded } n\text{-cell in } \mathbf{R}^n,$$

(b) $C_c^{(\infty)}(\mathbf{R}^n)$.

PROOF. a) The proof is almost the same as the proof of Theorem 4.1.3. First the $E_k$:s can be chosen to be open balls with their centres at the origin since each bounded set in $\mathbf{R}^n$ has finite $\mu$-measure. Moreover, as in the proof of Theorem 4.1.3 we can assume that $\mu$ is a finite measure. Now let $A$ be an at most denumerable dense subset of $\mathbf{R}^n$ and for each $a \in A$ let

$$R(a) = \{ r > 0;\ \mu(\{ x \in X;\ \mid x_k - a_k \mid = r \}) > 0 \text{ for some } k = 1, ..., n \}.$$

Then $\cup_{a \in A} R(a)$ is at most denumerable and there is a subset $\{ r_n;\ n \in \mathbf{N}_+ \}$ of $]0, \infty[ \setminus \cup_{a \in A} R(a)$ which is dense in $]0, \infty[$ . Finally, let $\mathcal{U}$ denote the class of all open sets which are finite unions of open balls of the type $B(a, r_n)$, $a \in A$, $n \in \mathbf{N}_+$, and proceed as in the proof of Theorem 4.1.3. The result follows by observing that the characteristic function of any member of $\mathcal{U}$ equals a finite sum of characteristic functions of open bounded $n$-cells a.e. $[\mu]$ .

Part (b) in Theorem 4.1.4 follows from Part (a) and the following

**Lemma 4.1.1.** *Suppose $K \subseteq U \subseteq \mathbf{R}^n$, where $K$ is compact and $U$ is open. Then there exists a function $f \in C_c^{(\infty)}(\mathbf{R}^n)$* such that

$$K \prec f \prec U$$

*that is*

$$\chi_K \leq f \leq \chi_U \text{ and } \mathrm{supp} f \subseteq U.$$

PROOF. Suppose $\rho \in C_c^\infty(\mathbf{R}^n)$ is non-negative, supp $\rho \subseteq B(0,1)$, and

$$\int_{\mathbf{R}^n} \rho dm_n = 1.$$

Moreover, let $\varepsilon > 0$ be fixed. For any $g \in L^1(v_n)$ we define

$$f_\varepsilon(x) = \varepsilon^{-n} \int_{\mathbf{R}^n} g(y)\rho(\varepsilon^{-1}(x-y))dy.$$

Since

$$| g | \max_{\mathbf{R}^n} | \frac{\partial^{k_1+...+k_n}\rho}{\partial x_1^{k_1}...\partial x_n^{k_n}} | \in L^1(v_n), \text{ all } k_1, ..., k_n \in \mathbf{N}$$

the Lebesgue Dominated Convergent Theorem shows that $f_\varepsilon \in C^\infty(\mathbf{R}^n)$. Here $f_\varepsilon \in C_c^\infty(\mathbf{R}^n)$ if $g$ vanishes off a bounded subset of $\mathbf{R}^n$. In fact,

$$\text{supp } f_\varepsilon \subseteq (\text{supp } g)_\varepsilon.$$

Now choose a positive number $\varepsilon \leq \frac{1}{2}d(K, U^c)$ and define $g = \chi_{K_\varepsilon}$. Since

$$f_\varepsilon(x) = \int_{\mathbf{R}^n} g(x - \varepsilon y)\rho(y)dy$$

we also have that $f_\varepsilon(x) = 1$ if $x \in K$. The lemma is proved.

**Example 4.1.2.** Suppose $f \in L^1(m_n)$ and let $g : \mathbf{R}^n \to \mathbf{R}$ be a bounded Lebesgue measurable function. Set

$$h(x) = \int_{\mathbf{R}^n} f(x-y)g(y)dy,\ x \in \mathbf{R}^n.$$

We claim that $h$ is continuous.

To see this first note that

$$h(x+\Delta x) - h(x) = \int_{\mathbf{R}^n} (f(x+\Delta x - y) - f(x-y))g(y)dy$$

and

$$\mid h(x+\Delta x) - h(x) \mid \leq K \int_{\mathbf{R}^n} \mid f(x+\Delta x - y) - f(x-y) \mid dy$$

$$= K \int_{\mathbf{R}^n} \mid f(\Delta x + y) - f(y) \mid dy$$

if $\mid g(x) \mid \leq K$ for every $x \in \mathbf{R}^n$. Now first choose $\varepsilon > 0$ and then $\varphi \in C_c(\mathbf{R}^n)$ such that

$$\parallel f - \varphi \parallel_1 < \varepsilon.$$

Using the triangle inequality, we get

$$\mid h(x+\Delta x) - h(x) \mid \leq K(2 \parallel f - \varphi \parallel_1 + \int_{\mathbf{R}^n} \mid \varphi(\Delta x + y) - \varphi(y) \mid dy)$$

$$\leq K(2\varepsilon + \int_{\mathbf{R}^n} \mid \varphi(\Delta x + y) - \varphi(y) \mid dy)$$

where the right hand side is smaller than $3K\varepsilon$ if $\mid \Delta x \mid$ is sufficiently small. This proves that $h$ is continuous.

**Example 4.1.3.** Suppose $A \in \mathcal{R}_n^-$ and $m_n(A) > 0$. We claim that the set

$$A - A = \{x - x; x \in A\}$$

contains a neighbourhood of the origin.

To show this there is no loss of generality to assume that $m_n(A) < \infty$. Set

$$f(x) = m_n(A \cap (A + x)),\ x \in \mathbf{R}^n.$$

Note that

$$f(x) = \int_{\mathbf{R}^n} \chi_A(y)\chi_A(y-x)dy$$

and Example 4.1.2 proves that $f$ is continuous. Since $f(0) > 0$ there exists a $\delta > 0$ such that $f(x) > 0$ if $\mid x \mid < \delta$. In particular, $A \cap (A+x) \neq \phi$ if $\mid x \mid < \delta$, which proves that

$$B(0,\delta) \subseteq A - A.$$

The following three examples are based on the Axiom of Choice.

**Example 4.1.4.** Let $NL$ be the non-Lebesgue measurable set constructed in Section 1.3. Furthermore, assume $A \subseteq \mathbf{R}$ is Lebesgue measurable and $A \subseteq NL$. We claim that $m(A) = 0$. If not, there exists a $\delta > 0$ such that $B(0,\delta) \subseteq A - A \subseteq NL - NL$. If $0 < r < \delta$ and $r \in \mathbf{Q}$, there exist $a, b \in NL$ such that

$$a = b + r.$$

But then $a \neq b$ and at the same time $a$ and $b$ belong to the same equivalence class, which is a contradiction. Accordingly from this, $m(A) = 0$.

**Example 4.1.5.** Suppose $A \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]$ is Lebesgue measurable and $m(A) > 0$. We claim there exists a non-Lebesgue measurable subset of $A$. To see this note that

$$A = \cup_{i=1}^{\infty}((r_i + NL) \cap A)$$

where $(r_i)_{i=1}^{\infty}$ is an enumeration of the rational numbers in the interval $[-1, 1]$. If each set $(r_i + NL) \cap A$, is Lebesgue measurable

$$m(A) = \Sigma_{i=1}^{\infty} m((r_i + NL) \cap A)$$

and we conclude that $m((r_i + NL) \cap A) > 0$ for an appropriate $i$. But then $m(NL \cap (A - r_i)) > 0$ and $NL \cap (A - r_i) \subseteq NL$, which contradicts Example 4.1.4. Hence $(r_i + NL) \cap A$ is non-Lebesgue measurable for an appropriate $i$.

If $A$ is a Lebesgue measurable subset of the real line of positive Lebesgue measure, we conclude that $A$ contains a non-Lebesgue measurable subset.

**Example 4.1.6.** Set $I = [0,1]$. We claim there exist a continuous function $f : I \to I$ and a Lebesgue measurable set $L \subseteq I$ such that $f^{-1}(L)$ is not Lebesgue measurable.

First recall from Section 3.3 the construction of the Cantor set $C$ and the Cantor function $G$. First $C_0 = [0,1]$. Then trisect $C_0$ and remove the middle interval $\left]\frac{1}{3}, \frac{2}{3}\right[$ to obtain $C_1 = C_0 \setminus \left]\frac{1}{3}, \frac{2}{3}\right[ = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. At the second stage subdivide each of the closed intervals of $C_1$ into thirds and remove from each one the middle open thirds. Then $C_2 = C_1 \setminus (\left]\frac{1}{9}, \frac{2}{9}\right[ \cup \left]\frac{7}{9}, \frac{8}{9}\right[)$. We repeat the process and what is left from $C_{n-1}$ is $C_n$. The set $[0,1] \setminus C_n$ is the union of $2^n - 1$ intervals numbered $I_k^n$, $k = 1, ..., 2^n - 1$, where the interval $I_k^n$ is situated to the left of the interval $I_l^n$ if $k < l$. The Cantor set $C = \cap_{n=1}^{\infty} C_n$.

Suppose $n$ is fixed and let $G_n : [0,1] \to [0,1]$ be the unique the monotone increasing continuous function, which satisfies $G_n(0) = 0, G_n(1) = 1, G_n(x) = k2^{-n}$ for $x \in I_k^n$ and which is linear on each interval of $C_{n.}$ It is clear that $G_n = G_{n+1}$ on each interval $I_k^n$, $k = 1, ..., 2^n - 1$. The Cantor function is defined by the limit $G(x) = \lim_{n\to\infty} G_n(x)$, $0 \le x \le 1$.

Now define

$$h(x) = \frac{1}{2}(x + G(x)),\ x \in I$$

where $G$ is the Cantor function. Since $h : I \to I$ is a strictly increasing and continuous bijection, the inverse function $f = h^{-1}$ is a continuous bijection from $I$ onto $I$. Set

$$A = h(I \setminus C)$$

and

$$B = h(C).$$

Recall from the definition of $G$ that $G$ is constant on each removed interval $I_k^n$ and that $h$ takes each removed interval onto an interval of half its length. Thus $m(A) = \frac{1}{2}$ and $m(B) = 1 - m(A) = \frac{1}{2}$.

By the previous example there exists a non-Lebesgue measurable subset $M$ of $B$. Put $L = h^{-1}(M)$. The set $L$ is Lebesgue measurable since $L \subseteq C$ and $C$ is a Lebesgue null set. However, the set $M = f^{-1}(L)$ is not Lebesgue measurable.

**Exercises**

1. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose $\varphi(t) = \min(t, 1)$, $t \geq 0$. Prove that $f_n \to f$ in measure if and only if $\varphi(\mid f_n - f \mid) \to 0$ in $L^1(\mu)$.

2. Let $\mu = m_{|[0,1]}$. Find measurable functions $f_n : [0,1] \to [0,1]\,,\ n \in \mathbf{N}_+$, such that $f_n \to 0$ in $L^2(\mu)$ as $n \to \infty$,

$$\liminf_{n\to\infty} f_n(x) = 0 \text{ all } x \in [0,1]$$

and

$$\limsup_{n\to\infty} f_n(x) = 1 \text{ all } x \in [0,1]\,.$$

3. If $f \in \mathcal{F}(X)$ set

$$\parallel f \parallel_0 = \inf \{\alpha \in [0,\infty]\,;\ \mu(\mid f \mid > \alpha) \leq \alpha\}\,.$$

Let

$$L^0(\mu) = \{f \in \mathcal{F}(X);\ \ \parallel f \parallel_0 < \infty\}$$

and identify functions in $L^0(\mu)$ which agree a.e. $[\mu]$.

(a) Prove that $d^{(0)} = \parallel f - g \parallel_0$ is a metric on $L^0(\mu)$ and that the corresponding metric space is complete.

(b) Show that $\mathcal{F}(X) = L^0(\mu)$ if $\mu$ is a finite positive measure.

4. Suppose $L^p(X, \mathcal{M}, \mu)$ is separable, where $p = 1$ or 2. Show that $L^p(X, \mathcal{M}^-, \bar{\mu})$ is separable.

5. Suppose $g$ is a real-valued, Lebesgue measurable, and bounded function of period one. Prove that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)g(nx)dx = \int_{-\infty}^{\infty} f(x)dx \int_0^1 g(x)dx$$

for every $f \in L^1(m)$.

6. Let $h_n(t) = 2 + \sin nt$, $0 \leq t \leq 1$, and $n \in \mathbf{N}_+$. Find real constants $\alpha$ and $\beta$ such that

$$\lim_{n \to \infty} \int_0^1 f(t)h_n(t)dt = \alpha \int_0^1 f(t)dt$$

and

$$\lim_{n \to \infty} \int_0^1 \frac{f(t)}{h_n(t)}dt = \beta \int_0^1 f(t)dt$$

for every real-valued Lebesgue integrable function $f$ on $[0, 1]$.

7. If $k = (k_1, ..., k_n) \in \mathbf{N}_+^n$, set $e_k(x) = \Pi_{i=1}^n \sin k_i x_i$, $x = (x_1, ..., x_n) \in \mathbf{R}^n$, and $\mid k \mid = (\Sigma_{i=1}^n k_i^2)^{\frac{1}{2}}$. Prove that

$$\lim_{|k| \to \infty} \int_{\mathbf{R}^n} f e_k dm_n = 0$$

for every $f \in L^1(m_n)$.

8. Suppose $f \in L^1(m_n)$, where $m_n$ denotes Lebesgue measure on $\mathbf{R}^n$. Compute the following limit and justify the calculations:

$$\lim_{|h| \to \infty} \int_{\mathbf{R}^n} \mid f(x+h) - f(x) \mid dx.$$

9. The set $A \subseteq \mathbf{R}$ has positive Lebesgue measure and

$$A + \mathbf{Q} = \{x + y;\ x \in A \text{ and } y \in \mathbf{Q}\}$$

where $\mathbf{Q}$ stands for the set of all rational numbers. Show that the set

$$\mathbf{R} \backslash (A + \mathbf{Q})$$

is a Lebesgue null set. (Hint: The function $f(x) = m(A\Delta(A - x))$, $x \in \mathbf{R}$, is continuous.)

## 4.2 Orthogonality

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. If $f, g \in L^2(\mu)$, let

$$\langle f, g \rangle =_{def} \int_X fg d\mu$$

be the so called scalar product of $f$ and $g$. The Cauchy-Schwarz inequality

$$| \langle f, g \rangle | \leq \| f \|_2 \| g \|_2$$

shows that the map $f \to \langle f, g \rangle$ of $L^2(\mu)$ into $\mathbf{R}$ is continuous. Observe that

$$\| f + g \|_2^2 = \| f \|_2^2 + 2\langle f, g \rangle + \| g \|_2^2$$

and from this we get the so called Parallelogram Law

$$\| f + g \|_2^2 + \| f - g \|_2^2 = 2(\| f \|_2^2 + \| g \|_2^2).$$

We will say that $f$ and $g$ are orthogonal (abbr. $f \perp g$) if $\langle f, g \rangle = 0$. Note that

$$\| f + g \|_2^2 = \| f \|_2^2 + \| g \|_2^2 \text{ if and only if } f \perp g.$$

Since $f \perp g$ implies $g \perp f$, the relation $\perp$ is symmetric. Moreover, if $f \perp h$ and $g \perp h$ then $(\alpha f + \beta g) \perp h$ for all $\alpha, \beta \in \mathbf{R}$. Thus $h^\perp =_{def} \{ f \in L^2(\mu);\ f \perp h \}$ is a subspace of $L^2(\mu)$, which is closed since the map $f \to \langle f, h \rangle$, $f \in L^2(\mu)$ is continuous. If $M$ is a subspace of $L^2(\mu)$, the set

$$M^\perp =_{def} \cap_{h \in M} h^\perp$$

is a closed subspace of $L^2(\mu)$. The function $f = 0$ if and only if $f \perp f$.

If $M$ is a subspace of $L^2(\mu)$ and $f \in L^2(\mu)$ there exists at most one point $g \in M$ such that $f - g \in M^\perp$. To see this, let $g_0, g_1 \in M$ be such that $f - g_k \in M^\perp$, $k = 0, 1$. Then $g_1 - g_0 = (f - g_0) - (f - g_1) \in M^\perp$ and hence $g_1 - g_0 \perp g_1 - g_0$ that is $g_0 = g_1$.

**Theorem 4.2.1.** *Let $M$ be a closed subspace in $L^2(\mu)$ and suppose $f \in L^2(\mu)$. Then there exists a unique point $g \in M$ such that*

$$\| f - g \|_2 \leq \| f - h \|_2 \ \text{ all } h \in M.$$

*Moreover,*

$$f - g \in M^{\perp}.$$

The function $g$ in Theorem 4.2.1 is called the projection of $f$ on $M$ and is denoted by $\mathrm{Proj}_M\, f$.

PROOF OF THEOREM 4.2.1. Set

$$d =_{def} d^{(2)}(f, M) = \inf_{g \in M} \| f - g \|_2 .$$

and let $(g_n)_{n=1}^{\infty}$ be a sequence in $M$ such that

$$d = \lim_{n \to \infty} \| f - g_n \|_2 .$$

Then, by the Parallelogram Law

$$\| (f-g_k)+(f-g_n) \|_2^2 + \| (f-g_k)-(f-g_n) \|_2^2 = 2(\| f-g_k \|_2^2 + \| f-g_n \|_2^2)$$

that is

$$4 \| f - \frac{1}{2}(g_k + g_n) \|_2^2 + \| g_n - g_k \|_2^2 = 2(\| f - g_k \|_2^2 + \| f - g_n \|_2^2)$$

and, since $\frac{1}{2}(g_k + g_n) \in M$, we get

$$4d^2 + \| g_n - g_k \|_2^2 \leq 2(\| f - g_k \|_2^2 + \| f - g_n \|_2^2).$$

Here the right hand converges to $4d^2$ as $k$ and $n$ go to infinity and we conclude that $(g_n)_{n=1}^{\infty}$ is a Cauchy sequence. Since $L^2(\mu)$ is complete and $M$ closed there exists a $g \in M$ such that $g_n \to g$ as $n \to \infty$. Moreover,

$$d = \| f - g \|_2 .$$

We claim that $f - g \in M^{\perp}$. To prove this choose $h \in M$ and $\alpha > 0$ arbitrarily and use the inequality

$$\| (f - g) + \alpha h \|_2^2 \geq \| f - g \|_2^2$$

to obtain

$$\| f - g \|_2^2 + 2\alpha \langle f - g, h \rangle + \alpha^2 \| h \|_2^2 \geq \| f - g \|_2^2$$

and

$$2\langle f-g,h\rangle+\alpha \parallel h \parallel_2^2 \geq 0.$$

By letting $\alpha \to 0$, $\langle f-g,h\rangle \geq 0$ and replacing $h$ by $-h$, $\langle f-g,h\rangle \leq 0$. Thus $f-g \in h^{\perp}$ and it follows that $f-g \in M^{\perp}$.

The uniqueness in Theorem 4.2.1 follows from the remark just before the formulation of Theorem 4.2.1. The theorem is proved.

A linear mapping $T : L^2(\mu) \to \mathbf{R}$ is called a linear functional on $L^2(\mu)$. If $h \in L^2(\mu)$, the map $h \to \langle f,h\rangle$ of $L^2(\mu)$ into $\mathbf{R}$ is a continuous linear functional on $L^2(\mu)$. It is a very important fact that every continuous linear functional on $L^2(\mu)$ is of this type.

**Theorem 4.2.2.** *Suppose $T$ is a continuous linear functional on $L^2(\mu)$. Then there exists a unique $w \in L^2(\mu)$ such that*

$$Tf = \langle f,w\rangle \text{ all } f \in L^2(\mu).$$

PROOF. Uniqueness: If $w, w' \in L^2(\mu)$ and $\langle f,w\rangle = \langle f,w'\rangle$ for all $f \in L^2(\mu)$, then $\langle f,w-w'\rangle = 0$ for all $f \in L^2(\mu)$. By choosing $f = w-w'$ we get $f \perp f$ that is $w = w'$.

Existence: The set $M =_{def} T^{-1}(\{0\})$ is closed since $T$ is continuous and $M$ is a linear subspace of $L^2(\mu)$ since $T$ is linear. If $M = L^2(\mu)$ we choose $w = 0$. Otherwise, pick a $g \in L^2(\mu) \setminus M$. Without loss of generality it can be assumed that $Tg = 1$ by eventually multiplying $g$ by a scalar. The previous theorem gives us a vector $h \in M$ such that $u =_{def} g-h \in M^{\perp}$. Note that $0 < \parallel u \parallel_2^2 = \langle u, g-h\rangle = \langle u,g\rangle$.

To conclude the proof, let fixed $f \in L^2(\mu)$ be fixed, and use that $(Tf)g - f \in M$ to obtain

$$\langle (Tf)g-f,u\rangle = 0$$

or

$$(Tf)\langle g,u\rangle = \langle f,u\rangle.$$

By setting

$$w = \frac{1}{\| u \|_2^2} u$$

we are done.

$$\downarrow\downarrow\downarrow$$

**4.3. The Haar Basis and Wiener Measure**

In this section we will show the existence of Brownian motion with continuous paths as a consequence of the existence of linear measure $\lambda$ in the unit interval. The so called Wiener measure is the probability law on $C\,[0,1]$ of real-valued Brownian motion in the time interval $[0,1]$ . The Brownian motion process is named after the British botanist Robert Brown (1773-1858). It was suggested by Lous Bachelier in 1900 as a model of stock price fluctuations and later by Albert Einstein in 1905 as a model of the physical phenomenon Brownian motion. The existence of the mathematical Brownian motion process was first established by Norbert Wiener in the twenties. Wiener also proved that the model can be chosen such that the path $t \rightarrow W(t)$, $0 \leq t \leq 1$, is continuous a.s. Today Brownian motion is a very important concept in probability, financial mathematics, partial differential equations and in many other fields in pure and applied mathematics.

Suppose $n$ is a non-negative integer and set $I_n = \{0, ..., n\}$ . A sequence $(e_i)_{i \in I_n}$ in $L^2(\mu)$ is said to be orthonormal if $e_i \perp e_j$ for all $i \neq j$, $i, j \in I_n$ and $\| e_i \| = 1$ for each $i \in I_n$. If $(e_i)_{i \in I_n}$ is orthonormal and $f \in L^2(\mu)$,

$$f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \perp e_j \text{ all } j \in I$$

and Theorem 4.2.1 shows that

$$\| f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2 \leq \| f - \Sigma_{i \in I_n} \alpha_i e_i \|_2 \text{ all real } \alpha_1, ..., \alpha_n.$$

Moreover

$$\| f \|_2^2 = \| f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2^2 + \| \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2^2$$

and we get

$$\Sigma_{i \in I_n} \langle f, e_i \rangle^2 \leq \| f \|_2^2 .$$

We say that $(e_n)_{n\in I_n}$ is an orthonormal basis in $L^2(\mu)$ if it is orthonormal and

$$f = \Sigma_{i\in I_n}\langle f, e_i\rangle e_i \text{ all } f \in L^2(\mu).$$

A sequence $(e_i)_{i=0}^{\infty}$ in $L^2(\mu)$ is said to be orthonormal if $(e_i)_{i=0}^{n}$ is orthonormal for each non-negative integer $n$. In this case, for each $f \in L^2(\mu)$,

$$\Sigma_{i=0}^{\infty}\langle f, e_i\rangle^2 \leq \| f \|_2^2$$

and the series

$$\Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i$$

converges since the sequence

$$(\Sigma_{i=0}^{n}\langle f, e_i\rangle e_i)_{n=0}^{\infty}$$

of partial sums is a Cauchy sequence in $L^2(\mu)$. We say that $(e_i)_{i=0}^{\infty}$ is an orthonormal basis in $L^2(\mu)$ if it is orthonormal and

$$f = \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i \text{ for all } f \in L^2(\mu).$$

**Theorem 4.3.1.** *An orthonormal sequence $(e_i)_{i=0}^{\infty}$ in $L^2(\mu)$ is a basis of $L^2(\mu)$ if*

$$(\langle f, e_i\rangle = 0 \text{ all } i \in \mathbf{N}) \Rightarrow f = 0$$

**Proof.** Let $f \in L^2(\mu)$ and set

$$g = f - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i.$$

Then, for any $j \in \mathbf{N}$,

$$\langle g, e_j\rangle = \langle f - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i, e_j\rangle$$

$$= \langle f, e_j\rangle - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle\langle e_i, e_j\rangle = \langle f, e_j\rangle - \langle f, e_j\rangle = 0.$$

Thus $g = 0$ or

$$f = \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i.$$

The theorem is proved.

As an example of an application of Theorem 4.3.1 we next construct an orthonormal basis of $L^2(\lambda)$, where $\lambda$ is linear measure in the unit interval. Set

$$H(t) = \chi_{[0,\frac{1}{2}[}(t) - \chi_{[\frac{1}{2},1]}(t),\ t \in \mathbf{R}$$

Moreover, define $h_{00}(t) = 1,\ 0 \leq t \leq 1$, and for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$,

$$h_{jn}(t) = 2^{\frac{n-1}{2}} H(2^{n-1}t - j + 1),\ 0 \leq t \leq 1.$$

Stated otherwise, we have for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$

$$h_{jn}(t) = \begin{cases} 2^{\frac{n-1}{2}},\ \frac{j-1}{2^{n-1}} \leq t < \frac{j-\frac{1}{2}}{2^{n-1}}, \\ \\ -2^{\frac{n-1}{2}},\ \frac{j-\frac{1}{2}}{2^{n-1}} \leq t \leq \frac{j}{2^{n-1}}, \\ \\ 0, \text{ elsewhere in } [0,1]. \end{cases}$$

It is simple to show that the sequence $h_{00}, h_{jn}, j = 1, ..., 2^{n-1},\ n \geq 1$, is orthonormal in $L^2(\lambda)$. We will prove that the same sequence constitute an orthonormal basis of $L^2(\lambda)$. Therefore, suppose $f \in L^2(\lambda)$ is orthogonal to each of the functions $h_{00}, h_{jn}, j = 1, ..., 2^{n-1},\ n \geq 1$. Then for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$

$$\int_{\frac{j-1}{2^{n-1}}}^{\frac{j-\frac{1}{2}}{2^{n-1}}} f d\lambda = \int_{\frac{j-\frac{1}{2}}{2^{n-1}}}^{\frac{j}{2^{n-1}}} f d\lambda$$

and, hence,

$$\int_{\frac{j-1}{2^{n-1}}}^{\frac{j}{2^{n-1}}} f d\lambda = \frac{1}{2^{n-1}} \int_0^1 f d\lambda = 0$$

since

$$\int_0^1 f d\lambda = \int_0^1 f h_{00} d\lambda = 0.$$

Thus

$$\int_{\frac{j}{2^{n-1}}}^{\frac{k}{2^{n-1}}} f d\lambda = 0,\ 1 \leq j \leq k \leq 2^{n-1}$$

and we conclude that

$$\int_0^1 1_{[a,b]} f d\lambda = \int_a^b f d\lambda = 0,\ 0 \leq a \leq b \leq 1.$$

Accordingly from this, $f = 0$ and we are done.

The above basis $(h_k)_{k=0}^{\infty} = (h_{00,} h_{11}, h_{12}, h_{22}, h_{13}, h_{23}, h_{33}, h_{43}, ...)$ of $L^2(\lambda)$ is called the Haar basis.

Let $0 \leq t \leq 1$ and define for fixed $k \in \mathbf{N}$

$$a_k(t) = \int_0^1 \chi_{[0,t]}(x) h_k(x) dx = \int_0^t h_k d\lambda$$

so that

$$\chi_{[0,t]} = \Sigma_{k=0}^{\infty} a_k(t) h_k \text{ in } L^2(\lambda).$$

Then, if $0 \leq s, t \leq 1$,

$$\min(s,t) = \int_0^1 \chi_{[0,s]}(x) \chi_{[0,t]}(x) dx = \langle \Sigma_{k=1}^{\infty} a_k(s) h_k, \chi_{[0,t]} \rangle$$

$$= \Sigma_{k=0}^{\infty} a_k(s) \langle h_k, \chi_{[0,t]} \rangle = \Sigma_{k=0}^{\infty} a_k(s) a_k(t).$$

Note that

$$t = \Sigma_{k=0}^{\infty} a_k^2(t).$$

If $(G_k)_{k=0}^{\infty}$ is a sequence of $N(0,1)$ distributed random variables based on a probability space $(\Omega, \mathcal{F}, P)$ the series

$$\Sigma_{k=0}^{\infty} a_k(t) G_k$$

converges in $L^2(P)$ and defines a Gaussian random variable which we denote by $W(t)$. From the above it follows that $(W(t))_{0 \leq t \leq 1}$ is a real-valued centred Gaussian stochastic process with the covariance

$$E\left[W(s)W(t)\right] = \min(s,t).$$

Such a process is called a real-valued Brownian motion in the time interval $[0,1]$.

Recall that

$$(h_{00,} h_{11}, h_{12}, h_{22}, h_{13}, h_{23}, h_{33}, h_{43}, ...) = (h_k)_{k=0}^{\infty}.$$

We define

$$(a_{00,} a_{11}, a_{12}, a_{22}, a_{13}, a_{23}, a_{33}, a_{43}, ...) = (a_k)_{k=0}^{\infty}$$

and

$$(G_{00,} G_{11}, G_{12}, G_{22}, G_{13}, G_{23}, G_{33}, G_{43}, ...) = (G_k)_{k=0}^{\infty}.$$

It is important to note that for fixed $n$,

$$a_{jn}(t) = \int_0^t \chi_{[0,t]}(x) h_{jn}(x) dx \neq 0 \text{ for at most one } j.$$

Set

$$U_0(t) = a_{00}(t) G_{00}$$

and

$$U_n(t) = \Sigma_{j=1}^{2^{n-1}} a_{jn}(t) G_{jn},\ n \in \mathbf{N}_+.$$

We know that

$$W(t) = \Sigma_{n=0}^{\infty} U_n(t) \text{ in } L^2(P)$$

for fixed $t$.

The space $C[0,1]$ will from now on be equipped with the metric

$$d(x,y) = \parallel x - y \parallel_\infty$$

where $\parallel x \parallel_\infty = \max_{0 \leq t \leq 1} \mid x(t) \mid$ . Recall that every $x \in C[0,1]$ is uniformly continuous. From this, remembering that $\mathbf{R}$ is separable, it follows that the space $C[0,1]$ is separable. Since $\mathbf{R}$ is complete it is also simple to show that the metric space $C[0,1]$ is complete. Finally, if $x_n \in C[0,1]$, $n \in \mathbf{N}$, and

$$\Sigma_{n=0}^{\infty} \parallel x_n \parallel_\infty < \infty$$

the series

$$\Sigma_{n=0}^{\infty} x_n$$

converges since the partial sums

$$s_n = \Sigma_{k=0}^{n} x_k,\ k \in \mathbf{N}$$

forms a Cauchy sequence.

We now define

$$\Theta = \{\omega \in \Omega; \Sigma_{n=0}^{\infty} \parallel U_n \parallel_\infty < \infty\}.$$

Here $\Theta \in \mathcal{F}$ since

$$\parallel U_n \parallel_\infty = \sup_{\substack{0 \leq t \leq 1 \\ t \in \mathbf{Q}}} \mid U_n(t) \mid$$

for each $n$. Next we prove that $\Omega \setminus \Theta$ is a null set.

To this end let $n \geq 1$ and note that

$$P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] \leq P\left[\max_{1\leq j\leq 2^{n-1}} (\| a_{jn} \|_\infty | G_{jn} |) > 2^{-\frac{n}{4}}\right].$$

But

$$\| a_{jn} \|_\infty = \frac{1}{2^{\frac{n+1}{2}}}$$

and, hence,

$$P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] \leq 2^{n-1} P\left[| G_{00} | > 2^{\frac{n}{4}+\frac{1}{2}}\right].$$

Since

$$x \geq 1 \Rightarrow P\left[| G_{00} | \geq x\right] \leq 2\int_x^\infty ye^{-y^2/2}\frac{dy}{x\sqrt{2\pi}} \leq e^{-x^2/2}$$

we get

$$P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] \leq 2^n e^{-2^{n/2}}$$

and conclude that

$$E\left[\sum_{n=0}^\infty 1_{\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right]}\right] = \sum_{n=0}^\infty P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] < \infty.$$

From this and the Beppo Levi Theorem (or the first Borel-Cantelli Lemma) $P[\Theta] = 1$.

The trajectory $t \to W(t,\omega)$, $0 \leq t \leq 1$, is continuous for every $\omega \in \Theta$. Without loss of generality, from now on we can therefore assume that all trajectories of Brownian motion are continuous (by eventually replacing $\Omega$ by $\Theta$).

Suppose

$$0 \leq t_1 < ... < t_n \leq 1$$

and let $I_1, ..., I_n$ be open subintervals of the real line. The set

$$S(t_1, ..., t_n; I_1, ..., I_n) = \{x \in C[0,1]\,;\ x(t_k) \in I_k,\ k = 1, ..., n\}$$

is called an open $n$-cell in $C[0,1]$. A set in $C[0,T]$ is called an open cell if there exists an $n \in \mathbf{N}_+$ such that it is an open $n$-cell. The $\sigma$-algebra generated by all open cells in $C[0,1]$ is denoted by $\mathcal{C}$. The construction above shows that the map

$$W : \Omega \to C[0,1]$$

which maps $\omega$ to the trajectory

$$t \to W(t, \omega),\ 0 \le t \le 1$$

is $(\mathcal{F}, \mathcal{C})$-measurable. The image measure $P_W$ is called Wiener measure in $C\,[0,1]$.

The Wiener measure is a Borel measure on the metric space $C\,[0,1]$. We leave it as an excersice to prove that

$$\mathcal{C} = \mathcal{B}(C\,[0,1]).$$

$\uparrow\uparrow\uparrow$

# CHAPTER 5

# DECOMPOSITION OF MEASURES

## Introduction

In this section a version of the fundamental theorem of calculus for Lebesgue integrals will be proved. Moreover, the concept of differentiating a measure with respect to another measure will be developped. A very important result in this chapter is the so called Radon-Nikodym Theorem.

## 5.1. Complex Measures

Let $(X, \mathcal{M})$ be a measurable space. Recall that if $A_n \subseteq X$, $n \in \mathbf{N}_+$, and $A_i \cap A_j = \phi$ if $i \neq j$, the sequence $(A_n)_{n\in\mathbf{N}_+}$ is called a disjoint denumerable collection. The collection is called a measurable partition of $A$ if $A = \cup_{n=1}^{\infty} A_n$ and $A_n \in \mathcal{M}$ for every $n \in \mathbf{N}_+$.

A complex function $\mu$ on $\mathcal{M}$ is called a complex measure if

$$\mu(A) = \Sigma_{n=1}^{\infty} \mu(A_n)$$

for every $A \in \mathcal{M}$ and measurable partition $(A_n)_{n=1}^{\infty}$ of $A$. Note that $\mu(\phi) = 0$ if $\mu$ is a complex measure. A complex measure is said to be a real measure if it is a real function. The reader should note that a positive measure need not be a real measure since infinity is not a real number. If $\mu$ is a complex measure $\mu = \mu_{\mathrm{Re}} + i\mu_{\mathrm{Im}}$ , where $\mu_{\mathrm{Re}}$ =Re $\mu$ and $\mu_{\mathrm{Im}}$ =Im $\mu$ are real measures.

If $(X, \mathcal{M}, \mu)$ is a positive measure and $f \in L^1(\mu)$ it follows that

$$\lambda(A) = \int_A f d\mu, \ A \in \mathcal{M}$$

is a real measure and we write $d\lambda = f d\mu$.

A function $\mu : \mathcal{M} \to [-\infty, \infty]$ is called a signed measure measure if

(a) $\mu : \mathcal{M} \to ]-\infty, \infty]$ or $\mu : \mathcal{M} \to [-\infty, \infty[$
(b) $\mu(\phi) = 0$
and
(c) for every $A \in \mathcal{M}$ and measurable partition $(A_n)_{n=1}^{\infty}$ of $A$,

$$\mu(A) = \Sigma_{n=1}^{\infty} \mu(A_n)$$

where the latter sum converges absolutely if $\mu(A) \in \mathbf{R}$.

Here $-\infty - \infty = -\infty$ and $-\infty + x = -\infty$ if $x \in \mathbf{R}$. The sum of a positive measure and a real measure and the difference of a real measure and a positive measure are examples of signed measures and it can be proved that there are no other signed measures (see Folland $[F]$). Below we concentrate on positive, real, and complex measures and will not say more about signed measures here.

Suppose $\mu$ is a complex measure on $\mathcal{M}$ and define for every $A \in \mathcal{M}$

$$\mid \mu \mid (A) = \sup \Sigma_{n=1}^{\infty} \mid \mu(A_n) \mid,$$

where the supremum is taken over all measurable partitions $(A_n)_{n=1}^{\infty}$ of $A$. Note that $\mid \mu \mid (\phi) = 0$ and

$$\mid \mu \mid (A) \geq \mid \mu(B) \mid \text{ if } A, B \in \mathcal{M} \text{ and } A \supseteq B.$$

The set function $\mid \mu \mid$ is called the total variation of $\mu$ or the total variation measure of $\mu$. It turns out that $\mid \mu \mid$ is a positive measure. In fact, as will shortly be seen, $\mid \mu \mid$ is a finite positive measure.

**Theorem 5.1.1.** *The total variation $\mid \mu \mid$ of a complex measure is a positive measure.*

PROOF. Let $(A_n)_{n=1}^{\infty}$ be a measurable partition of $A$.

For each $n$, suppose $a_n <\mid \mu \mid (A_n)$ and let $(E_{kn})_{k=1}^{\infty}$ be a measurable partition of $A_n$ such that

$$a_n < \Sigma_{k=1}^{\infty} \mid \mu(E_{kn}) \mid .$$

Since $(E_{kn})_{k,n=1}^{\infty}$ is a partition of $A$ it follows that

$$\Sigma_{n=1}^{\infty} a_n < \Sigma_{k,n=1}^{\infty} \mid \mu(E_{kn}) \mid \leq \mid \mu \mid (A).$$

Thus

$$\Sigma_{n=1}^{\infty} \mid \mu \mid (A_n) \leq \mid \mu \mid (A).$$

To prove the opposite inequality, let $(E_k)_{k=1}^{\infty}$ be a measurable partition of $A$. Then, since $(A_n \cap E_k)_{n=1}^{\infty}$ is a measurable partition of $E_k$ and $(A_n \cap E_k)_{k=1}^{\infty}$ a measurable partition of $A_n$,

$$\Sigma_{k=1}^{\infty} \mid \mu(E_k) \mid = \Sigma_{k=1}^{\infty} \mid \Sigma_{n=1}^{\infty} \mu(A_n \cap E_k) \mid$$

$$\leq \Sigma_{k,n=1}^{\infty} \mid \mu(A_n \cap E_k) \mid \leq \Sigma_{n=1}^{\infty} \mid \mu \mid (A_n)$$

and we get

$$\mid \mu \mid (A) \leq \Sigma_{n=1}^{\infty} \mid \mu \mid (A_n).$$

Thus

$$\mid \mu \mid (A) = \Sigma_{n=1}^{\infty} \mid \mu \mid (A_n).$$

Since $\mid \mu \mid (\phi) = 0$, the theorem is proved.

**Theorem 5.1.2.** *The total variation $\mid \mu \mid$ of a complex measure $\mu$ is a finite positive measure.*

PROOF. Since

$$\mid \mu \mid \leq \mid \mu_{\mathrm{Re}} \mid + \mid \mu_{\mathrm{Im}} \mid$$

there is no loss of generality to assume that $\mu$ is a real measure.

Suppose $\mid \mu \mid (E) = \infty$ for some $E \in \mathcal{M}$. We first prove that there exist disjoint sets $A, B \in \mathcal{M}$ such that

$$A \cup B = E$$

and

$$| \mu(A) | > 1 \text{ and } | \mu | (B) = \infty.$$

To this end let $c = 2(1+ | \mu(E) |)$ and let $(E_k)_{k=1}^{\infty}$ be a measurable partition of $E$ such that

$$\Sigma_{k=1}^{n} | \mu(E_k) | > c$$

for some sufficiently large $n$. There exists a subset $N$ of $\{1, ..., n\}$ such that

$$| \Sigma_{k \in N} \mu(E_k) | > \frac{c}{2}.$$

Set $A = \cup_{k \in N} E_k$ and $B = E \setminus A$. Then $| \mu(A) | > \frac{c}{2} \geq 1$ and

$$| \mu(B) | = | \mu(E) - \mu(A) |$$

$$\geq | \mu(A) | - | \mu(E) | > \frac{c}{2} - | \mu(E) | = 1.$$

Since $\infty = | \mu | (E) = | \mu | (A) + | \mu | (B)$ we have $| \mu | (A) = \infty$ or $| \mu | (B) = \infty$. If $| \mu | (B) < \infty$ we interchange $A$ and $B$ and have $| \mu(A) | > 1$ and $| \mu | (B) = \infty$.

Suppose $| \mu | (X) = \infty$. Set $E_0 = X$ and choose disjoint sets $A_0, B_0 \in \mathcal{M}$ such that

$$A_0 \cup B_0 = E_0$$

and

$$| \mu(A_0) | > 1 \text{ and } | \mu | (B_0) = \infty.$$

Set $E_1 = B_0$ and choose disjoint sets $A_1, B_1 \in \mathcal{M}$ such that

$$A_1 \cup B_1 = E_1$$

and

$$| \mu(A_1) | > 1 \text{ and } | \mu | (B_1) = \infty.$$

By induction, we find a measurable partition $(A_n)_{n=0}^{\infty}$ of the set $A =_{def} \cup_{n=0}^{\infty} A_n$ such that $| \mu(A_n) | > 1$ for every $n$. Now, since $\mu$ is a complex measure,

$$\mu(A) = \Sigma_{n=0}^{\infty} \mu(A_n).$$

But this series cannot converge, since the general term does not tend to zero as $n \to \infty$. This contradiction shows that $| \mu |$ is a finite positive measure.

If $\mu$ is a real measure we define

$$\mu^+ = \frac{1}{2}(\mid \mu \mid + \mu)$$

and

$$\mu^- = \frac{1}{2}(\mid \mu \mid - \mu).$$

The measures $\mu^+$ and $\mu^-$ are finite positive measures and are called the positive and negative variations of $\mu$, respectively . The representation

$$\mu = \mu^+ - \mu^-$$

is called the Jordan decomposition of $\mu$.

**Exercises**

1. Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space and $d\lambda = f d\mu$, where $f \in L^1(\mu)$. Prove that $d \mid \lambda \mid = \mid f \mid d\mu$.

2. Suppose $\lambda, \mu$, and $\nu$ are real measures defined on the same $\sigma$-algebra and $\lambda \leq \mu$ and $\lambda \leq \nu$. Prove that

$$\lambda \leq \min(\mu, \nu)$$

where

$$\min(\mu, \nu) = \frac{1}{2}(\mu + \nu - \mid \mu - \nu \mid).$$

3. Suppose $\mu : \mathcal{M} \to \mathbf{C}$ is a complex measure and $f, g : X \to \mathbf{R}$ measurable functions. Show that

$$\mid \mu(f \in A) - \mu(g \in A) \mid \leq \mid \mu \mid (f \neq g)$$

for every $A \in \mathcal{R}$.

## 5.2. The Lebesque Decomposition and the Radon-Nikodym Theorem

Let $\mu$ be a positive measure on $\mathcal{M}$ and $\lambda$ a positive or complex measure on $\mathcal{M}$. The measure $\lambda$ is said to be absolutely continuous with respect to $\mu$ (abbreviated $\lambda << \mu$) if $\lambda(A) = 0$ for every $A \in \mathcal{M}$ for which $\mu(A) = 0$. If we define

$$\mathcal{Z}_\lambda = \{A \in \mathcal{M};\ \lambda(A) = 0\}$$

it follows that $\lambda << \mu$ if and only if

$$\mathcal{Z}_\mu \subseteq \mathcal{Z}_\lambda.$$

For example, $\gamma_n << v_n$ and $v_n << \gamma_n$.

The measure $\lambda$ is said to be concentrated on $E \in \mathcal{M}$ if $\lambda = \lambda^E$ , where $\lambda^E(A) =_{def} \lambda(E \cap A)$ for every $A \in \mathcal{M}$. This is equivalent to the hypothesis that $A \in \mathcal{Z}_\lambda$ if $A \in \mathcal{M}$ and $A \cap E = \phi$. Thus if $E_1, E_2 \in \mathcal{M}$, where $E_1 \subseteq E_2$, and $\lambda$ is concentrated on $E_1$, then $\lambda$ is concentrated on $E_2$. Moreover, if $E_1, E_2 \in \mathcal{M}$ and $\lambda$ is concentrated on both $E_1$ and $E_2$, then $\lambda$ is concentrated on $E_1 \cap E_2$. Two measures $\lambda_1$ and $\lambda_2$ are said to be mutually singular (abbreviated $\lambda_1 \perp \lambda_2$) if there exist disjoint measurable sets $E_1$ and $E_2$ such that $\lambda_1$ is concentrated on $E_1$ and $\lambda_2$ is concentrated on $E_2$.

**Theorem 5.2.1.** *Let $\mu$ be a positive measure and $\lambda, \lambda_1$, and $\lambda_2$ complex measures.*

(i) *If $\lambda_1 << \mu$ and $\lambda_2 << \mu$, then $(\alpha_1\lambda_1 + \alpha_2\lambda_2) << \mu$ for all complex numbers $\alpha_1$ and $\alpha_2$.*

(ii) *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $(\alpha_1\lambda_1 + \alpha_2\lambda_2) \perp \mu$ for all complex numbers $\alpha_1$ and $\alpha_2$.*

(iii) *If $\lambda << \mu$ and $\lambda \perp \mu$, then $\lambda = 0$.*

(iv) *If $\lambda << \mu$, then $\mid \lambda \mid << \mu$.*

PROOF. The properties (i) and (ii) are simple to prove and are left as exercises.

To prove (iii) suppose $E \in \mathcal{M}$ is a $\mu$-null set and $\lambda = \lambda^E$. If $A \in \mathcal{M}$, then $\lambda(A) = \lambda(A \cap E)$ and $A \cap E$ is a $\mu$-null set. Since $\lambda << \mu$ it follows that $A \cap E \in Z_\lambda$ and, hence, $\lambda(A) = \lambda(A \cap E) = 0$. This proves (iii)

To prove (iv) suppose $A \in \mathcal{M}$ and $\mu(A) = 0$. If $(A_n)_{n=1}^{\infty}$ is measurable partition of $A$, then $\mu(A_n) = 0$ for every $n$. Since $\lambda << \mu$, $\lambda(A_n) = 0$ for every $n$ and we conclude that $\mid \lambda \mid (A) = 0$. This proves (vi).

**Theorem 5.2.2.** *Let $\mu$ be a positive measure on $\mathcal{M}$ and $\lambda$ a complex measure on $\mathcal{M}$. Then the following conditions are equivalent:*

(a) $\lambda << \mu$.

(b) *To every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that $\mid \lambda(E) \mid < \varepsilon$ for all $E \in \mathcal{M}$ with $\mu(E) < \delta$.*

If $\lambda$ is a positive measure, the implication $(a) \Rightarrow (b)$ in Theorem 5.2.2 is, in general, wrong. To see this take $\mu = \gamma_1$ and $\lambda = v_1$. Then $\lambda << \mu$ and if we choose $A_n = [n, \infty[$, $n \in \mathbf{N}_+$, then $\mu(A_n) \to 0$ as $n \to \infty$ but $\lambda(A_n) = \infty$ for each $n$.

PROOF. (a)$\Rightarrow$(b). If (b) is wrong there exist an $\varepsilon > 0$ and sets $E_n \in \mathcal{M}$, $n \in \mathbf{N}_+$, such that $\mid \lambda(E_n) \mid \geq \varepsilon$ and $\mu(E_n) < 2^{-n}$. Set

$$A_n = \cup_{k=n}^{\infty} E_k \text{ and } A = \cap_{n=1}^{\infty} A_n.$$

Since $A_n \supseteq A_{n+1} \supseteq A$ and $\mu(A_n) < 2^{-n+1}$, it follows that $\mu(A) = 0$ and using that $\mid \lambda \mid (A_n) \geq \mid \lambda(E_n) \mid$, Theorem 1.1.2 (f) implies that

$$\mid \lambda \mid (A) = \lim_{n \to \infty} \mid \lambda \mid (A_n) \geq \varepsilon.$$

This contradicts that $\mid \lambda \mid << \mu$.

(b)$\Rightarrow$(a). If $E \in \mathcal{M}$ and $\mu(E) = 0$ then to each $\varepsilon > 0$, $\mid \lambda(E) \mid < \varepsilon$, and we conclude that $\lambda(E) = 0$. The theorem is proved.

**Theorem 5.2.3.** *Let $\mu$ be a $\sigma$-finite positive measure and $\lambda$ a real measure on $\mathcal{M}$.*

(a) **(The Lebesgue Decomposition of $\lambda$)** *There exists a unique pair of real measures $\lambda_a$ and $\lambda_s$ on$\mathcal{M}$ such that*

$$\lambda = \lambda_a + \lambda_s, \;\; \lambda_a << \mu, \text{ and } \lambda_s \perp \mu.$$

*If $\lambda$ is a finite positive measure, $\lambda_a$ and $\lambda_s$ are finite positive measures.*

(b) **(The Radon-Nikodym Theorem)** *There exits a unique $g \in L^1(\mu)$ such that*

$$d\lambda_a = gd\mu.$$

*If $\lambda$ is a finite positive measure, $g \geq 0$ a.e. $[\mu]$.*

The proof of Theorem 5.2.3 is based on the following

**Lemma 5.2.1.** *Let $(X, M, \mu)$ be a finite positive measure space and suppose $f \in L^1(\mu)$.*

(a) *If $a \in \mathbf{R}$ and*

$$\int_E f d\mu \leq a\mu(E), \text{ all } E \in \mathcal{M}$$

*then $f \leq a$ a.e. $[\mu]$.*

(b) *If $b \in \mathbf{R}$ and*

$$\int_E f d\mu \geq b\mu(E), \text{ all } E \in \mathcal{M}$$

*then $f \geq b$ a.e. $[\mu]$.*

PROOF. (a) Set $g = f - a$ so that

$$\int_E gd\mu \leq 0, \;\; \textit{all } E \in \mathcal{M}.$$

Now choose $E = \{g > 0\}$ to obtain

$$0 \geq \int_E gd\mu = \int_X \chi_E gd\mu \geq 0$$

as $\chi_E g \geq 0$ a.e. $[\mu]$. But then Example 2.1.2 yields $\chi_E g = 0$ a.e. $[\mu]$ and we get $E \in Z_\mu$. Thus $g \leq 0$ a.e. $[\mu]$ or $f \leq a$ a.e. $[\mu]$.

Part (b) follows in a similar way as Part (a) and the proof is omitted here.

PROOF. Uniqueness: (a) Suppose $\lambda_a^{(k)}$ and $\lambda_s^{(k)}$ are real measures on $\mathcal{M}$ such that

$$\lambda = \lambda_a^{(k)} + \lambda_s^{(k)},\ \lambda_a^{(k)} << \mu,\ \text{and}\ \lambda_s^{(k)} \perp \mu$$

for $k = 1, 2$. Then

$$\lambda_a^{(1)} - \lambda_a^{(2)} = \lambda_s^{(2)} - \lambda_s^{(1)}$$

and

$$\lambda_a^{(1)} - \lambda_a^{(2)} << \mu \text{ and } \lambda_a^{(1)} - \lambda_a^{(2)} \perp \mu.$$

Thus by applying Theorem 5.2.1, $\lambda_a^{(1)} - \lambda_a^{(2)} = 0$ and $\lambda_a^{(1)} = \lambda_a^{(2)}$. From this we conclude that $\lambda_s^{(1)} = \lambda_s^{(2)}$ .

(b) Suppose $g_k \in L^1(\mu), k = 1, 2,$ and

$$d\lambda_a = g_1 d\mu = g_2 d\mu.$$

Then $hd\mu = 0$ where $h = g_1 - g_2$. But then

$$\int_{\{h>0\}} hd\mu = 0$$

and it follows that $h \leq 0$ a.e. $[\mu]$. In a similar way we prove that $h \geq 0$ a.e. $[\mu]$. Thus $h = 0$ in $L^1(\mu)$, that is $g_1 = g_2$ in $L^1(\mu)$.

Existence: The beautiful proof that follows is due to von Neumann.

First suppose that $\mu$ and $\lambda$ are finite positive measures and set $\nu = \lambda + \mu$. Clearly, $L^1(\lambda) \supseteq L^1(\nu) \supseteq L^2(\nu)$. Moreover, if $f : X \to \mathbf{R}$ is measurable

$$\int_X \mid f \mid d\lambda \leq \int_X \mid f \mid d\nu \leq \sqrt{\int_X f^2 d\nu}\sqrt{\nu(X)}$$

and from this we conclude that the map

$$f \to \int_X f d\lambda$$

is a continuous linear functional on $L^2(\nu)$. Therefore, in view of Theorem 4.2.2, there exists a $g \in L^2(\nu)$ such that

$$\int_X f d\lambda = \int_X fg d\nu \text{ all } f \in L^2(\nu).$$

Suppose $E \in \mathcal{M}$ and put $f = \chi_E$ to obtain

$$0 \leq \lambda(E) = \int_E g d\nu$$

and, since $\nu \geq \lambda$,

$$0 \leq \int_E g d\nu \leq \nu(E).$$

But then Lemma 5.2.1 implies that $0 \leq g \leq 1$ a.e. $[\nu]$. Therefore, without loss of generality we can assume that $0 \leq g(x) \leq 1$ for all $x \in X$ and, in addition, as above

$$\int_X f d\lambda = \int_X fg d\nu \text{ all } f \in L^2(\nu)$$

that is

$$\int_X f(1-g) d\lambda = \int_X fg d\mu \text{ all } f \in L^2(\nu).$$

Put $A = \{0 \leq g < 1\}$, $S = \{g = 1\}$, $\lambda_a = \lambda^A$, and $\lambda_s = \lambda^S$. Note that $\lambda = \lambda^A + \lambda^S$. The choice $f = \chi_S$ gives $\mu(S) = 0$ and hence $\lambda_s \perp \mu$. Moreover, the choice

$$f = (1 + ... + g^n)\chi_E$$

where $E \in \mathcal{M}$, gives

$$\int_E (1 - g^{n+1}) d\lambda = \int_E (1 + ... + g^n) g d\mu.$$

By letting $n \to \infty$ and using monotone convergence

$$\lambda(E \cap A) = \int_E h d\mu.$$

where

$$h = \lim_{n \to \infty} (1 + ... + g^n) g.$$

Since $h$ is non-negative and

$$\lambda(A) = \int_X h d\mu$$

it follows that $h \in L^1(\mu)$. Moreover, the construction above shows that $\lambda = \lambda_a + \lambda_s$.

In the next step we assume that $\mu$ is a $\sigma$-finite positive measure and $\lambda$ a finite positive measure. Let $(X_n)_{n=1}^{\infty}$ be a measurable partition of $X$ such that $\mu(X_n) < \infty$ for every $n$. Let $n$ be fixed and apply Part (a) to the pair $\mu^{X_n}$ and $\lambda^{X_n}$ to obtain finite positive measures $(\lambda^{X_n})_a$ and $(\lambda^{X_n})_s$ such that

$$\lambda^{X_n} = (\lambda^{X_n})_a + (\lambda^{X_n})_s, \; (\lambda^{X_n})_a << \mu^{X_n}, \text{ and } (\lambda^{X_n})_s \perp \mu^{X_n}$$

and

$$d(\lambda^{X_n})_a = h_n d\mu^{X_n} \text{ (or } (\lambda^{X_n})_a = h_n \mu^{X_n})$$

where $0 \leq h_n \in L^1(\mu^{X_n})$. Without loss of generality we can assume that $h_n = 0$ off $X_n$ and that $(\lambda^{X_n})_s$ is concentrated on $A_n \subseteq X_n$ where $A_n \in \mathcal{Z}_\mu$. In particular, $(\lambda^{X_n})_a = h_n \mu$. Now

$$\lambda = h\mu + \Sigma_{n=1}^{\infty}(\lambda^{X_n})_s$$

where

$$h = \Sigma_{n=1}^{\infty} h_n$$

and

$$\int_X h d\mu \leq \lambda(X) < \infty.$$

Thus $h \in L^1(\mu)$. Moreover, $\lambda_s =_{def} \Sigma_{n=1}^{\infty}(\lambda^{X_n})_s$ is concentrated on $\cup_{n=1}^{\infty} A_n \in \mathcal{Z}_\mu$. Hence $\lambda_s \perp \mu$.

Finally if $\lambda$ is a real measure we apply what we have already proved to the positive and negative variations of $\lambda$ and we are done.

**Example 5.2.1.** Let $\lambda$ be Lebesgue measure in the unit interval and $\mu$ the counting measure in the unit interval restricted to the class of all Lebesgue measurable subsets of the unit interval. Clearly, $\lambda << \mu$. Suppose there is an

$h \in L^1(\mu)$ such that $d\lambda = hd\mu$. We can assume that $h \geq 0$ and the Markov inequality implies that the set $\{h \geq \varepsilon\}$ is finite for every $\varepsilon > 0$. But then

$$\lambda(h \in ]0,1]) = \lim_{n\to\infty} \lambda(h \geq 2^{-n}) = 0$$

and it follows that $1 = \lambda(h = 0) = \int_{\{h=0\}} hd\mu = 0$, which is a contradiction.

**Corollary 5.2.1.** *Suppose $\mu$ is a real measure. Then there exists*

$$h \in L^1(\mid \mu \mid)$$

*such that $\mid h(x) \mid= 1$ for all $x \in X$ and*

$$d\mu = hd \mid \mu \mid .$$

PROOF. Since $\mid \mu(A) \mid\leq\mid \mu \mid (A)$ for every $A \in \mathcal{M}$, the Radon-Nikodym Theorem implies that $d\mu = hd \mid \mu \mid$ for an appropriate $h \in L^1(\mid \mu \mid)$. But then $d \mid \mu \mid=\mid h \mid d \mid \mu \mid$ (see Exercise 1 in Chapter 5.1). Thus

$$\mid \mu \mid (E) = \int_E \mid h \mid d \mid \mu \mid, \text{ all } E \in \mathcal{M}$$

and Lemma 5.2.1 yields $h = 1$ a.e. $[\mid \mu \mid]$. From this the theorem follows at once.

**Theorem 5.2.4.** **(Hahn's Decomposition Theorem)** *Suppose $\mu$ is a real measure. There exists an $A \in \mathcal{M}$ such that*

$$\mu^+ = \mu^A \text{ and } \mu^- = -\mu^{A^c}.$$

PROOF. Let $d\mu = hd \mid \mu \mid$ where $\mid h \mid= 1$. Note that $hd\mu = d \mid \mu \mid$ . Set $A = \{h = 1\}$ . Then

$$d\mu^+ = \frac{1}{2}(d \mid \mu \mid +d\mu) = \frac{1}{2}(h + 1)d\mu = \chi_A d\mu$$

and

$$d\mu^- = d\mu^+ - d\mu = (\chi_A - 1)d\mu = -\chi_{A^c} d\mu.$$

The theorem is proved.

If a real measure $\lambda$ is absolutely continuous with respect to a $\sigma$-finite positive measure $\mu$, the Radon-Nikodym Theorem says that $d\lambda = fd\mu$ for an approprite $f \in L^1(\mu)$. We sometimes write

$$f = \frac{d\lambda}{d\mu}$$

and call $f$ the Radon-Nikodym derivate of $\lambda$ with respect to $\mu$.

**Exercises**

1. Let $\mu$ be a $\sigma$-finite positive measure on $(X, \mathcal{M})$ and $(f_n)_{n\in\mathbf{N}}$ a sequence of measurable functions which converges in $\mu$-measure to a measurable function $f$. Moreover, suppose $\nu$ is a finite positive measure on $(X, \mathcal{M})$ such that $\nu << \mu$. Prove that $f_n \to f$ in $\nu$-measure.

2. Suppose $\mu$ and $\nu_n, n \in\mathbf{N}$, are positive measures defined on the same $\sigma$-algebra and set $\theta = \Sigma_{n=0}^{\infty} \nu_n$. Prove that
   a) $\theta \perp \mu$ if $\nu_n \perp \mu$, all $n \in \mathbf{N}$.
   b) $\theta << \mu$ if $\nu_n << \mu$, all $n \in \mathbf{N}$.

3. Suppose $\mu$ is a real measure and $\mu = \lambda_1 - \lambda_2$, where $\lambda_1$ and $\lambda_2$ are finite positive measures. Prove that $\lambda_1 \geq \mu^+$ and $\lambda_2 \geq \mu^-$.

4. Let $\lambda_1$ and $\lambda_2$ be mutually singular complex measures on the same $\sigma$-algebra. Show that $\mid \lambda_1 \mid \perp \mid \lambda_2 \mid$ .

5. Let $(X, \mathcal{M}, \mu)$ be a $\sigma$-finite positive measure space and suppose $\lambda$ and $\tau$ are two probability measures defined on the $\sigma$-algebra $\mathcal{M}$ such that $\lambda << \mu$ and $\tau << \mu$. Prove that

$$\sup_{A \in \mathcal{M}} \mid \lambda(A) - \tau(A) \mid = \frac{1}{2} \int_X \mid \frac{d\lambda}{d\mu} - \frac{d\tau}{d\mu} \mid d\mu.$$

6. Let $(X, \mathcal{M})$ be a measurable space and suppose $\mu, \nu : \mathcal{M} \to \mathbf{R}$ and are real measures. Prove that

$$(\mu + \nu)^+ \leq \mu^+ + \nu^+.$$

## 5.3. The Wiener Maximal Theorem and the Lebesgue Differentiation Theorem

We say that a Lebesgue measurable function $f$ in $\mathbf{R}^n$ is locally Lebesgue integrable and belongs to the class $L^1_{loc}(m_n)$ if $f\chi_K \in L^1(m_n)$ for each compact subset $K$ of $\mathbf{R}^n$. In a similar way $f \in L^1_{loc}(v_n)$ if $f$ is a Borel function such that $f\chi_K \in L^1(v_n)$ for each compact subset $K$ of $\mathbf{R}^n$. If $f \in L^1_{loc}(m_n)$, we define the average $A_r f(x)$ of $f$ on the open ball $B(x, r)$ as

$$A_r f(x) = \frac{1}{m_n(B(x,r))} \int_{B(x,r)} f(y) dy.$$

It follows from dominated convergence that the map $(x, r) \to A_r f(x)$ of $\mathbf{R}^n \times \,]0, \infty[$ into $\mathbf{R}$ is continuous. The Hardy-Littlewood maximal function $f^*$ is, by definition, $f^* = \sup_{r>0} A_r \mid f \mid$ or, stated more explicitly,

$$f^*(x) = \sup_{r>0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) \mid dy, \; x \in \mathbf{R}^n.$$

The function $f^* : (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) \to ([0, \infty], \mathcal{R}_{0,\infty})$ is measurable since

$$f^* = \sup_{\substack{r>0 \\ r \in \mathbf{Q}}} A_r \mid f \mid.$$

**Theorem 5.3.1. (Wiener's Maximal Theorem)** *There exists a positive constant $C = C_n < \infty$ such that for all $f \in L^1(m_n)$,*

$$m_n(f^* > \alpha) \leq \frac{C}{\alpha} \parallel f \parallel_1 \text{ if } \alpha > 0.$$

The proof of the Wiener Maximal Theorem is based on the following remarkable result.

**Lemma 5.3.1.** *Let $\mathcal{C}$ be a collection of open balls in $\mathbf{R}^n$ and set $V = \cup_{B \in \mathcal{C}} B$. If $c < m_n(V)$ there exist pairwise disjoint $B_1, ..., B_k \in \mathcal{C}$ such that*

$$\Sigma_{i=1}^k m_n(B_i) > 3^{-n} c.$$

PROOF. Let $K \subseteq V$ be compact with $m_n(K) > c$, and suppose $A_1, ..., A_p \in \mathcal{C}$ cover $K$. Let $B_1$ be the largest of the $A_i's$ (that is, $B_1$ has maximal radius), let $B_2$ be the largest of the $A_i's$ which are disjoint from $B_1$, let $B_3$ be the largest of the $A_i's$ which are disjoint from $B_1 \cup B_2$, and so on until the process stops after $k$ steps. If $B_i = B(x_i, r_i)$ put $B_i^* = B(x_i, 3r_i)$. Then $\cup_{i=1}^k B_i^* \supseteq K$ and

$$c < \Sigma_{i=1}^k m_n(B_i^*) = 3^n \Sigma_{i=1}^k m_n(B_i).$$

The lemma is proved.

PROOF OF THEOREM 5.3.1. Set

$$E_\alpha = \{f^* > \alpha\}.$$

For each $x \in E_\alpha$ choose an $r_x > 0$ such that $A_{r_x} \mid f \mid (x) > \alpha$. If $c < m_n(E_\alpha)$, by Lemma 5.3.1 there exist $x_1, ..., x_k \in E_\alpha$ such that the balls $B_i = B(x_i, r_{x_i})$, $i = 1, ..., k$, are mutually disjoint and

$$\Sigma_{i=1}^k m_n(B_i) > 3^{-n} c.$$

But then

$$c < 3^n \Sigma_{i=1}^k m_n(B_i) < \frac{3^n}{\alpha} \Sigma_{i=1}^k \int_{B_i} \mid f(y) \mid dy \leq \frac{3^n}{\alpha} \int_{\mathbf{R}^n} \mid f(y) \mid dy.$$

The theorem is proved.

**Theorem 5.3.2.** *If* $f \in L^1_{loc}(m_n)$,

$$\lim_{r \to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} f(y)dy = f(x) \text{ a.e. } [m_n].$$

PROOF. Clearly, there is no loss of generality to assume that $f \in L^1(m_n)$. Suppose $g \in C_c(\mathbf{R}^n) =_{def} \{f \in C(\mathbf{R}^n);\ f(x) = 0 \text{ if } \mid x \mid \text{ large enough}\}$. Then

$$\lim_{r \to 0} A_r g(x) = g(x) \text{ all } x \in \mathbf{R}^n.$$

Since $A_r f - f = A_r(f-g) - (f-g) + A_r g - g$,

$$\overline{\lim_{r \to 0}} \mid A_r f - f \mid \leq (f-g)^* + \mid f - g \mid .$$

Now, for fixed $\alpha > 0$,

$$m_n(\overline{\lim_{r \to 0}} \mid A_r f - f \mid > \alpha)$$

$$\leq m_n((f-g)^* > \frac{\alpha}{2}) + m_n(\mid f - g \mid > \frac{\alpha}{2})$$

and the Wiener Maximal Theorem and the Markov Inequality give

$$m_n(\overline{\lim_{r \to 0}} \mid A_r f - f \mid > \alpha)$$

$$\leq (\frac{2C}{\alpha} + \frac{2}{\alpha}) \parallel f - g \parallel_1 .$$

Remembering that $C_c(\mathbf{R}^n)$ is dense in $L^1(m_n)$, the theorem follows at once.

If $f \in L^1_{loc}(m_n)$ we define the so called Lebesgue set $L_f$ to be

$$L_f = \left\{ x;\ \lim_{r \to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - f(x) \mid dy = 0 \right\}.$$

Note that if $q$ is real and

$$E_q = \left\{ x;\ \lim_{r \to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - q \mid dy = \mid f(x) - q \mid \right\}$$

then $m_n(\cup_{q\in\mathbf{Q}}E_q^c)=0$. If $x\in\cap_{q\in\mathbf{Q}}E_q$,

$$\overline{\lim_{r\to 0}}\frac{1}{m_n(B(x,r))}\int_{B(x,r)}\mid f(y)-f(x)\mid dy\leq 2\mid f(x)-q\mid$$

for all rational numbers $q$ and it follows that $m_n(L_f^c)=0$.

A family $\mathcal{E}_x=(E_{x,r})_{r>0}$ of Borel sets in $\mathbf{R}^n$ is said to shrink nicely to a point $x$ in $\mathbf{R}^n$ if $E_{x,r}\subseteq B(x,r)$ for each $r$ and there is a positive constant $\alpha$, independent of $r$, such that $m_n(E_{x,r})\geq\alpha m_n(B(x,r))$.

**Theorem 5.3.3. (The Lebesgue Differentiation Theorem)** *Suppose $f\in L^1_{loc}(m_n)$ and $x\in L_f$. Then*

$$\lim_{r\to 0}\frac{1}{m_n(E_{x,r})}\int_{E_{x,r}}\mid f(y)-f(x)\mid dy=0$$

and

$$\lim_{r\to 0}\frac{1}{m_n(E_{x,r})}\int_{E_{x,r}}f(y)dy=f(x).$$

PROOF. The result follows from the inequality

$$\frac{1}{m_n(E_{x,r})}\int_{E_{x,r}}\mid f(y)-f(x)\mid dy\leq\frac{1}{\alpha m_n(B(x,r))}\int_{B(x,r)}\mid f(y)-f(x)\mid dy.$$

**Theorem 5.3.4.** *Suppose $\lambda$ is a real or positive measure on $\mathcal{R}_n$ and suppose $\lambda\perp v_n$. If $\lambda$ is a positive measure it is assumed that $\lambda(K)<\infty$ for every compact subset of $\mathbf{R}^n$. Then*

$$\lim_{r\to 0}\frac{\lambda(E_{x,r})}{v_n(E_{x,r})}=0\text{ a.e. }[v_n]$$

If $E_{x,r}=B(x,r)$ and $\lambda$ is the counting measure $c_{\mathbf{Q}^n}$ restricted to $\mathcal{R}_n$ then $\lambda\perp v_n$ but the limit in Theorem 5.3.4 equals plus infinity for all $x\in\mathbf{R}^n$. The hypothesis "$\lambda(K)<\infty$ for every compact subset of $\mathbf{R}^n$" in Theorem 5.3.4 is not superflous.

PROOF. Since $| \lambda(E) | \leq | \lambda | (E)$ if $E \in \mathcal{R}_n$, there is no restriction to assume that $\lambda$ is a positive measure (cf. Theorem 3.1.4). Moreover, since

$$\frac{\lambda(E_{x,r})}{v_n(E_{x,r})} \leq \frac{\lambda(B(x,r))}{\alpha v_n(B(x,r))}$$

it can be assumed that $E_{x,r} = B(x,r)$. Note that the function $\lambda(B(\cdot,r))$ is Borel measurable for fixed $r > 0$ and $\lambda(B(x,\cdot))$ left continuous for fixed $x \in \mathbf{R}^n$.

Suppose $A \in \mathcal{Z}_\lambda$ and $v_n = (v_n)^A$. Given $\delta > 0$, it is enough to prove that $F \in \mathcal{Z}_{v_n}$ where

$$F = \left\{ x \in A;\ \overline{\lim_{r \to 0}} \frac{\lambda(B(x,r))}{m_n(B(x,r))} > \delta \right\}$$

To this end let $\varepsilon > 0$ and use Theorem 3.1.3 to get an open $U \supseteq A$ such that $\lambda(U) < \varepsilon$. For each $x \in F$ there is an open ball $B_x \subseteq U$ such that

$$\lambda(B_x) > \delta v_n(B_x).$$

If $V = \cup_{x \in F} B_x$ and $c < v_n(V)$ we use Lemma 5.3.1 to obtain $x_1, ..., x_k$ such that $B_{x_1}, ..., B_{x_k}$ are pairwise disjoint and

$$c < 3^n \Sigma_{i=1}^k v_n(B_{x_i}) < 3^n \delta^{-1} \Sigma_{i=1}^k \lambda(B_{x_i})$$

$$\leq 3^n \delta^{-1} \lambda(U) < 3^n \delta^{-1} \varepsilon.$$

Thus $v_n(V) \leq 3^n \delta^{-1} \varepsilon$. Since $V \supseteq F \in \mathcal{R}_n$ and $\varepsilon > 0$ is arbitrary, $v_n(F) = 0$ and the theorem is proved.

**Corollary 5.3.1.** *Suppose $F : \mathbf{R} \to \mathbf{R}$ is an increasing function. Then $F'(x)$ exists for almost all $x$ with respect to linear measure.*

PROOF. Let $D$ be the set of all points of discontinuity of $F$. Suppose $-\infty < a < b < \infty$ and $\varepsilon > 0$. If $a < x_1 < ... < x_n < b$, where $x_1, ..., x_n \in D$ and

$$F(x_k+) - F(x_k-) \geq \varepsilon,\ \ k = 1, ..., n$$

then

$$n\varepsilon \leq \Sigma_{k=1}^{n}(F(x_k+) - F(x_k-)) \leq F(b) - F(a).$$

Thus $D \cap [a,b]$ is at most denumerable and it follows that $D$ is at most denumerable. Set $H(x) = F(x+) - F(x)$, $x \in \mathbf{R}$, and let $(x_j)_{j=0}^{N}$ be an enumeration of the members of the set $\{H > 0\}$. Moreover, for any $a > 0$,

$$\sum_{|x_j|<a} H(x_j) \leq \sum_{|x_j|<a} (F(x_j+) - F(x_j-))$$

$$\leq F(a) - F(-a) < \infty.$$

Now, if we introduce

$$\nu(A) = \Sigma_{j=0}^{N} H(x_j)\delta_{x_j}(A),\ A \in \mathcal{R}$$

then $\nu$ is a positive measure such that $\nu(K) < \infty$ for each compact subset $K$ of $\mathbf{R}$. Furthermore, if $h$ is a non-zero real number,

$$\mid \frac{1}{h}(H(x+h) - H(x) \mid \leq \frac{1}{\mid h \mid}(H(x+h) + H(x)) \leq 4\frac{1}{4 \mid h \mid}\nu(B(x, 2 \mid h \mid)$$

and Theorem 5.3.4 implies that $H'(x) = 0$ a.e. $[v_1]$. Therefore, without loss of generality it may be assumed that $F$ is right continuous and, in addition, there is no restriction to assume that $F(+\infty) - F(-\infty) < \infty$.

By Section 1.6 $F$ induces a finite positive Borel measure $\mu$ such that

$$\mu(]x,y]) = F(y) - F(x) \text{ if } x < y.$$

Now consider the Lebesgue decomposition

$$d\mu = f dv_1 + d\lambda$$

where $f \in L^1(v_1)$ and $\lambda \perp v_1$. If $x < y$,

$$F(y) - F(x) = \int_x^y f(t)dt + \lambda(]x,y])$$

and the previous two theorems imply that

$$\lim_{y \downarrow x} \frac{F(y) - F(x)}{y - x} = f(x) \text{ a.e. } [v_1]$$

If $y < x$,

$$F(x) - F(y) = \int_y^x f(t)dt + \lambda(]y, x])$$

and we get

$$\lim_{y \uparrow x} \frac{F(y) - F(x)}{y - x} = f(x) \text{ a.e. } [v_1].$$

The theorem is proved.

**Exercises**

1. Suppose $F : \mathbf{R} \to \mathbf{R}$ is increasing and let $f \in L^1_{loc}(v_1)$ be such that $F'(x) = f(x)$ a.e. $[v_1]$. Prove that

$$\int_x^y f(t)dt \leq F(y) - F(x) \text{ if} - \infty < x \leq y < \infty.$$

## 5.4. Absolutely Continuous Functions and Functions of Bounded Variation

Throughout this section $a$ and $b$ are reals with $a < b$ and to simplify notation we set $m_{a,b} = m_{|[a,b]}$. If $f \in L^1(m_{a,b})$ we know from the previous section that the function

$$(If)(x) =_{def} \int_a^x f(t)dt, \ a \leq x \leq b$$

has the derivative $f(x)$ a.e. $[m_{a,b}]$, that is

$$\frac{d}{dx} \int_a^x f(t)dt = f(x) \text{ a.e. } [m_{a,b}].$$

Our next main task will be to describe the range of the linear map $I$.

A function $F : [a, b] \to \mathbf{R}$ is said to be absolutely continuous if to every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\Sigma_{i=1}^n \mid b_i - a_i \mid < \delta \text{ implies } \Sigma_{i=1}^n \mid F(b_i) - F(a_i) \mid < \varepsilon$$

whenever $]a_1, b_1[\,, ..., ]a_n, b_n[$ are disjoint open subintervals of $[a, b]$. It is obvious that an absolutely continuous function is continuous. It can be proved that the Cantor function is not absolutely continuous.1.

**Theorem 5.4.1.** *If $f \in L^1(m_{a,b})$, then $If$ is absolutely continuous.*

PROOF. There is no restriction to assume $f \geq 0$. Set

$$d\lambda = f dm_{a,b}.$$

By Theorem 5.2.2, to every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\lambda(A) < \varepsilon$ for each Lebesgue set $A$ in $[a, b]$ such that $m_{a,b}(A) < \delta$. Now restricting $A$ to be a finite disjoint union of open intervals, the theorem follows.

Suppose $-\infty \leq \alpha < \beta \leq \infty$ and $F : ]\alpha, \beta[ \to \mathbf{R}$. For every $x \in ]\alpha, \beta[$ we define

$$T_F(x) = \sup \Sigma_{i=1}^n \mid F(x_i) - F(x_{i-1}) \mid$$

where the supremum is taken over all positive integers $n$ and all choices $(x_i)_{i=0}^n$ such that

$$\alpha < x_0 < x_1 < ... < x_n = x < \beta.$$

The function $T_F : ]\alpha, \beta[ \to [0, \infty]$ is called the total variation of $F$. Note that $T_F$ is increasing. If $T_F$ is a bounded function, $F$ is said to be of bounded variation. A bounded increasing function on $\mathbf{R}$ is of bounded variation. Therefore the difference of two bounded increasing functions on $\mathbf{R}$ is of bounded variation. Interestingly enough, the converse is true. In the special case $]\alpha, \beta[ = \mathbf{R}$ we write $F \in BV$ if $F$ is of bounded variation.

**Example 5.4.1.** Let $f$:$\mathbf{R} \to \mathbf{R}$ be a Lebesgue integrable function and define

$$g(x) = \int_{-\infty}^{\infty} e^{-|y|} f(x - y) dy \text{ if } x \in \mathbf{R}.$$

We claim that $g$ is a continuous function of bounded variation.

To prov this claim put $h(x) = e^{-|x|}$ if $x \in \mathbf{R}$ so that

$$g(x) = \int_{-\infty}^{\infty} h(x-y)f(y)dy.$$

We first prove that the function $h$ is continuous. To this end suppose $(a_n)_{n \in \mathbf{N}_+}$ is a sequence of real numbers which converges to a real number $a$. Then

$$\mid h(a_n - y)f(y)) \mid \in \mid f(y)) \mid \text{ if } n \in \mathbf{N}_+ \text{ and } y \in \mathbf{R}$$

and since $f \in \mathcal{L}^1(m)$ by dominated convergence,

$$\lim_{n \to \infty} g(a_n) = \int_{-\infty}^{\infty} \lim_{n \to \infty} h(a_n - y)f(y)dy =$$

$$\int_{-\infty}^{\infty} h(a-y)f(y)dy = g(a)$$

and it follows that $g$ is continuous.

We next prove that the function $h$ is of bounded variation. Recall that the total variation function $T_h(x)$ of $h$ at the point $x$ is the supremum of all sums of the type

$$\sum_{i=1}^{n} \mid h(x_i) - h(x_{i-1}) \mid$$

where

$$-\infty < x_0 < x_1 < ... < x_n = x < \infty.$$

We claim that $h$ is the difference of two bounded increasing functions. Setting

$$\psi(x) = e^{\min(0,x)}$$

and observing that

$$h(x) = \psi(x) + \psi(-x) - 1$$

the claim above is obvious and

$$C =_{def} \sup T_h < \infty.$$

Moreover, if $-\infty < x_0 < x_1 < ... < x_n < \infty$,

$$\sum_{i=1}^{n} \mid g(x_i) - g(x_{i-1}) \mid =$$

$$\sum_{i=1}^{n} \mid \int_{-\infty}^{\infty} h(x_i - y) f(y) dy - \int_{-\infty}^{\infty} h(x_{i-1} - y) f(y) dy \mid$$

$$\leq \sum_{i=1}^{n} \int_{-\infty}^{\infty} \mid h(x_i - y) - h(x_{i-1} - y) \mid\mid f(y) \mid dy$$

$$\int_{-\infty}^{\infty} \left( \sum_{i=1}^{n} \mid h(x_i - y) - h(x_{i-1} - y) \mid \right) \mid f(y) \mid dy$$

$$\leq \int_{-\infty}^{\infty} C \mid f(y) \mid dy = C \int_{-\infty}^{\infty} \mid f(y) \mid dy < \infty.$$

Hence $g$ is of bounded variation.

**Theorem 5.4.2.** *Suppose $F \in BV$.*

(a) *The functions $T_F + F$ and $T_F - F$ are increasing and*

$$F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F).$$

*In particular, $F$ is differentiable almost everywhere with respect to linear measure.*

(b) *If $F$ is right continuous, then so is $T_F$.*

PROOF. (a) Let $x < y$ and $\varepsilon > 0$. Choose $x_0 < x_1 < ... < x_n = x$ such that

$$\Sigma_{i=1}^{n} \mid F(x_i) - F(x_{i-1}) \mid \geq T_f(x) - \varepsilon.$$

Then

$$T_F(y) + F(y)$$
$$\geq \Sigma_{i=1}^{n} \mid F(x_i) - F(x_{i-1}) \mid + \mid F(y) - F(x) \mid + (F(y) - F(x)) + F(x)$$
$$\geq T_F(x) - \varepsilon + F(x)$$

and, since $\varepsilon > 0$ is arbitrary, $T_F(y) + F(y) \geq T_F(x) + F(x)$. Hence $T_F + F$ is increasing. Finally, replacing $F$ by $-F$ it follows that the function $T_F - F$ is increasing.

(b) If $c \in \mathbf{R}$ and $x > c$,

$$T_f(x) = T_F(c) + \sup \Sigma_{i=1}^{n} \mid F(x_i) - F(x_{i-1}) \mid$$

where the supremum is taken over all positive integers $n$ and all choices $(x_i)_{i=0}^{n}$ such that

$$c = x_0 < x_1 < ... < x_n = x.$$

Suppose $T_F(c+) > T_F(c)$ where $c \in \mathbf{R}$. Then there is an $\varepsilon > 0$ such that

$$T_F(x) - T_F(c) > \varepsilon$$

for all $x > c$. Now, since $F$ is right continuous at the point $c$, for fixed $x > c$ there exists a partition

$$c < x_{11} < ... < x_{1n_1} = x$$

such that

$$\Sigma_{i=2}^{n_1} \mid F(x_{1i}) - F(x_{1i-1}) \mid > \varepsilon.$$

But

$$T_F(x_{11}) - T_F(c) > \varepsilon$$

and we get a partition

$$c < x_{21} < ... < x_{2n_2} = x_{11}$$

such that

$$\Sigma_{i=2}^{n_2} \mid F(x_{2i}) - F(x_{2i-1}) \mid > \varepsilon.$$

Summing up we have got a partition of the interval $[x_{21}, x]$ with

$$\Sigma_{i=2}^{n_2} \mid F(x_{2i}) - F(x_{2i-1}) \mid + \Sigma_{i=2}^{n_1} \mid F(x_{1i}) - F(x_{1i-1}) \mid > 2\varepsilon.$$

By repeating the process the total variation of $F$ becomes infinite, which is a contradiction. The theorem is proved.

**Theorem 5.4.3.** *Suppose $F : [a, b] \to \mathbf{R}$ is absolutely continuous. Then there exists a unique $f \in L^1(m_{a,b})$ such that*

$$F(x) = F(a) + \int_a^x f(t)dt, \ a \leq x \leq b.$$

*In particular, the range of the map $I$ equals the set of all real-valued absolutely continuous maps on $[a, b]$ which vanish at the point $a$.*

PROOF. Set $F(x) = F(a)$ if $x < a$ and $F(x) = F(b)$ if $x > b$. There exists a $\delta > 0$ such that

$$\Sigma_{i=1}^{n} \mid b_i - a_i \mid < \delta \text{ implies } \Sigma_{i=1}^{n} \mid F(b_i) - F(a_i) \mid < 1$$

whenever $]a_1, b_1[, ..., ]a_n, b_n[$ are disjoint subintervals of $[a, b]$. Let $p$ be the least positive integer such that $a + p\delta \geq b$. Then $T_F \leq p$ and $F \in BV$. Let $F = G - H$, where $G = \frac{1}{2}(T_F + F)$ and $H = \frac{1}{2}(T_F - F)$. There exist finite positive Borel measures $\lambda_G$ and $\lambda_H$ such that

$$\lambda_G(]x, y]) = G(y) - G(x), \ x \leq y$$

and

$$\lambda_H(]x, y]) = H(y) - H(x), \ x \leq y.$$

If we define $\lambda = \lambda_G - \lambda_H$,

$$\lambda(]x, y]) = F(y) - F(x), \ x \leq y.$$

Clearly,

$$\lambda(]x, y[) = F(y) - F(x), \ x \leq y$$

since $F$ is continuous.

Our next task will be to prove that $\lambda << v_1$. To this end, suppose $A \in \mathcal{R}$ and $v_1(A) = 0$. Now choose $\varepsilon > 0$ and let $\delta > 0$ be as in the definition of the absolute continuity of $F$ on $[a, b]$. For each $k \in \mathbf{N}_+$ there exists an open set $V_k \supseteq A$ such that $v_1(V_k) < \delta$ and $\lim_{k\to\infty} \lambda(V_k) = \lambda(A)$. But each fixed $V_k$ is a disjoint union of open intervals $(]a_i, b_i[)_{i=1}^{\infty}$ and hence

$$\Sigma_{i=1}^{n} \mid b_i - a_i \mid < \delta$$

for every $n$ and, accordingly from this,

$$\Sigma_{i=1}^{\infty} \mid F(b_i) - F(a_i) \mid \leq \varepsilon$$

and

$$\mid \lambda(V_k) \mid \leq \Sigma_{i=1}^{\infty} \mid \lambda(]a_i, b_i[) \mid \leq \varepsilon.$$

Thus $\mid \lambda(A) \mid \leq \varepsilon$ and since $\varepsilon > 0$ is arbitrary, $\lambda(A) = 0$. From this $\lambda << v_1$ and the theorem follows at once.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. From now on we write $f \in L^1(\mu)$ if there exist a $g \in \mathcal{L}^1(\mu)$ and an $A \in \mathcal{M}$ such that $A^c \in \mathcal{Z}_\mu$ and $f(x) = g(x)$ for all $x \in A$. Furthermore, we define

$$\int_X f d\mu = \int_X g d\mu$$

(cf the discussion in Section 2). Note that $f(x)$ need not be defined for every $x \in X$.

**Corollary 5.4.1.** *A function* $f : [a, b] \to \mathbf{R}$ *is absolutely continuous if and only if the following conditions are true:*
*(i)* $f'(x)$ *exists for* $m_{a,b}$*-almost all* $x \in [a, b]$
*(ii)* $f' \in L^1(m_{a,b})$
*(iii)* $f(x) = f(a) + \int_a^x f'(t)dt$, all $x \in [a, b]$.

**Exercises**

1. Suppose $f : [0, 1] \to \mathbf{R}$ satisfies $f(0) = 0$ and

$$f(x) = x^2 \sin \frac{1}{x^2} \text{ if } 0 < x \leq 1.$$

Prove that $f$ is differentiable everywhere but $f$ is not absolutely continuous.

2. Suppose $\alpha$ is a positive real number and $f$ a function on $[0, 1]$ such that $f(0) = 0$ and $f(x) = x^\alpha \sin \frac{1}{x}$, $0 < x \leq 1$. Prove that $f$ is absolutely continuous if and only if $\alpha > 1$.

3. Suppose $f(x) = x\cos(\pi/x)$ if $0 < x < 2$ and $f(x) = 0$ if $x \in \mathbf{R} \backslash ]0,2[$. Prove that $f$ is not of bounded variation on $\mathbf{R}$.

4 A function $f : [a,b] \to \mathbf{R}$ is a Lipschitz function, that is there exists a positive real number $C$ such that

$$\mid f(x) - f(y) \mid \leq C \mid x - y \mid$$

for all $x, y \in [a,b]$. Show that $f$ is absolutely continuous and $\mid f'(x) \mid \leq C$ a.e. $[m_{a,b}]$.

5. Suppose $f : [a,b] \to \mathbf{R}$ is absolutely continuous. Prove that

$$T_g(x) = \int_a^x \mid f'(t) \mid dt, \ a < x < b$$

if $g$ is the restriction of $f$ to the open interval $]a,b[$.

6. Suppose $f$ and $g$ are real-valued absolutely continuous functions on the compact interval $[a,b]$. Show that the function $h = \max(f,g)$ is absolutely continuous and $h' \leq \max(f', g')$ a.e. $[m_{a,b}]$.

7. Suppose $(X, \mathcal{M}, \mu)$ is a finite positive measure space and $f \in L^1(\mu)$. Define

$$g(t) = \int_X \mid f(x) - t \mid d\mu(x), \ t \in \mathbf{R}.$$

Prove that $g$ is absolutely continuous and

$$g(t) = g(a) + \int_a^t (\mu(f \leq s) - \mu(f \geq s))ds \ \text{ if } a, t \in \mathbf{R}.$$

8. Let $\mu$ and $\nu$ be probability measures on $(X, \mathcal{M})$ such that $\mid \mu - \nu \mid (X) = 2$. Show that $\mu \perp \nu$.

$\downarrow\downarrow\downarrow$

### 5.5. Conditional Expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose $\xi \in L^1(P)$. Moreover, suppose $\mathcal{G} \subseteq \mathcal{F}$ is a $\sigma$-algebra and set

$$\mu(A) = P\left[A\right], \ A \in \mathcal{G}$$

and

$$\lambda(A) = \int_A \xi dP, \ A \in \mathcal{G}.$$

It is trivial that $\mathcal{Z}_\mu = \mathcal{Z}_P \cap \mathcal{G} \subseteq \mathcal{Z}_\lambda$ and the Radon-Nikodym Theorem shows there exists a unique $\eta \in L^1(\mu)$ such that

$$\lambda(A) = \int_A \eta d\mu \text{ all } A \in \mathcal{G}$$

or, what amounts to the same thing,

$$\int_A \xi dP = \int_A \eta dP \text{ all } A \in \mathcal{G}.$$

Note that $\eta$ is $(\mathcal{G}, \mathcal{R})$-measurable. The random variable $\eta$ is called the conditional expectation of $\xi$ given $\mathcal{G}$ and it is standard to write $\eta = E\left[\xi \mid \mathcal{G}\right]$.

A sequence of $\sigma$-algebras $(\mathcal{F}_n)_{n=1}^\infty$ is called a filtration if

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}.$$

If $(\mathcal{F}_n)_{n=1}^\infty$ is a filtration and $(\xi_n)_{n=1}^\infty$ is a sequence of real valued random variables such that for each $n$,

(a) $\xi_n \in L^1(P)$
(b) $\xi_n$ is $(\mathcal{F}_n, \mathcal{R})$-measurable
(c) $E\left[\xi_{n+1} \mid \mathcal{F}_n\right] = \xi_n$

then $(\xi_n, \mathcal{F}_n)_{n=1}^\infty$ is called a martingale. There are very nice connections between martingales and the theory of differentiation (see e.g Billingsley $[B]$ and Malliavin $[M]$).

$\uparrow\uparrow\uparrow$

# CHAPTER 6

# COMPLEX INTEGRATION

## Introduction

In this section, in order to illustrate the power of Lebesgue integration, we collect a few results, which often appear with uncomplete proofs at the undergraduate level.

## 6.1. Complex Integrand

So far we have only treated integration of functions with their values in $\mathbf{R}$ or $[0, \infty]$ and it is the purpose of this section to discuss integration of complex valued functions.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure. Let $f, g \in L^1(\mu)$. We define

$$\int_X (f + ig)d\mu = \int_X f d\mu + i \int_X g d\mu.$$

If $\alpha$ and $\beta$ are real numbers,

$$\int_X (\alpha + i\beta)(f + ig)d\mu = \int_X ((\alpha f - \beta g) + i(\alpha g + \beta f))d\mu$$

$$= \int_X (\alpha f - \beta g)d\mu + i \int_X (\alpha g + \beta f)d\mu$$

$$= \alpha \int_X f d\mu - \beta \int_X g d\mu + i\alpha \int_X g d\mu + i\beta \int_X f d\mu$$

$$= (\alpha + i\beta)(\int_X f d\mu + i \int_X g d\mu)$$

$$= (\alpha + i\beta) \int_X (f + ig)d\mu.$$

We write $f \in L^1(\mu; \mathbf{C})$ if Re $f$, Im $f \in L^1(\mu)$ and have, for every $f \in L^1(\mu; \mathbf{C})$ and complex $\alpha$,

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu.$$

Clearly, if $f, g \in L^1(\mu; \mathbf{C})$, then

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Now suppose $\mu$ is a complex measure on $\mathcal{M}$. If

$$f \in L^1(\mu; \mathbf{C}) =_{def} L^1(\mu_{\mathrm{Re}}; \mathbf{C}) \cap L^1(\mu_{\mathrm{Im}}; \mathbf{C})$$

we define

$$\int_X f d\mu = \int_X f d\mu_{\mathrm{Re}} + i \int_X f d\mu_{\mathrm{Im}}.$$

It follows for every $f, g \in L^1(\mu; \mathbf{C})$ and $\alpha \in \mathbf{C}$ that

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu.$$

and

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$\downarrow\downarrow\downarrow$

**6.2. The Fourier Transform**

Below, if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n) \in \mathbf{R}^n$, we let

$$\langle x, y \rangle = \Sigma_{k=1}^n x_k y_k.$$

and

$$\mid x \mid = \sqrt{\langle x, y \rangle}.$$

If $\mu$ is a complex measure on $\mathcal{R}_n$ (or $\mathcal{R}_n^-$) the Fourier transform $\hat{\mu}$ of $\mu$ is defined by

$$\hat{\mu}(y) = \int_{\mathbf{R}^n} e^{-i\langle x, y \rangle} d\mu(x), \; y \in \mathbf{R}^n.$$

Note that

$$\hat{\mu}(0) = \mu(\mathbf{R}^n).$$

The Fourier transform of a function $f \in L^1(m_n; \mathbf{C})$ is defined by

$$\hat{f}(y) = \hat{\mu}(y) \text{ where } d\mu = f dm_n.$$

**Theorem 6.2.1.** *The canonical Gaussian measure $\gamma_n$ in $\mathbf{R}^n$ has the Fourier transform*

$$\hat{\gamma}_n(y) = e^{-\frac{|y|^2}{2}}.$$

PROOF. Since

$$\gamma_n = \gamma_1 \otimes ... \otimes \gamma_1 \ (n \text{ factors})$$

it is enough to consider the special case $n = 1$. Set

$$g(y) = \hat{\gamma}_1(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx.$$

Note that $g(0) = 1$. Since

$$\mid \frac{\cos x(y+h) - \cos xy}{h} \mid \leq \mid x \mid$$

the Lebesgue Dominated Convergence Theorem yields

$$g'(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} -xe^{-\frac{x^2}{2}} \sin xy dx$$

(Exercise: Prove this by using Example 2.2.1). Now, by partial integration,

$$g'(y) = \frac{1}{\sqrt{2\pi}} \left[ e^{-\frac{x^2}{2}} \sin xy \right]_{x=-\infty}^{x=\infty} - \frac{y}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx$$

that is

$$g'(y) + yg(y) = 0$$

and we get

$$g(y) = e^{-\frac{y^2}{2}}.$$

If $\xi = (\xi_1, ..., \xi_n)$ is an $\mathbf{R}^n$-valued random variable with $\xi_k \in L^1(P)$, $k = 1, ..., n$, the characteristic function $c_\xi$ of $\xi$ is defined by

$$c_\xi(y) = E\left[e^{i\langle \xi, y\rangle}\right] = \hat{P}_\xi(-y),\ y \in \mathbf{R}^n.$$

For example, if $\xi \in N(0, \sigma)$, then $\xi = \sigma G$, where $G \in N(0,1)$, and we get

$$c_\xi(y) = E\left[e^{i\langle G, \sigma y\rangle}\right] = \hat{\gamma}_1(-\sigma y)$$

$$= e^{-\frac{\sigma^2 y^2}{2}}.$$

Choosing $y = 1$ results in

$$E\left[e^{i\xi}\right] = e^{-\frac{1}{2}E\left[\xi^2\right]} \text{ if } \xi \in N(0, \sigma).$$

Thus if $(\xi_k)_{k=1}^n$ is a centred real-valued Gaussian process

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \exp(-\frac{1}{2}E\left[(\Sigma_{k=1}^n y_k \xi_k)^2\right]$$

$$= \exp(-\frac{1}{2}\Sigma_{k=1}^n y_k^2 E\left[\xi_k^2\right] - \Sigma_{1\leq j<k\leq n} y_j y_k E\left[\xi_j \xi_k\right]).$$

In particular, if

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k$$

we see that

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n e^{-\frac{y_k^2}{2}E\left[\xi_k^2\right]}$$

or

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n E\left[e^{i y_k \xi_k}\right].$$

Stated otherwise, the Fourier tranforms of the measures $P_{(\xi_1,...,\xi_n)}$ and $\times_{k=1}^n P_{\xi_k}$ agree. Below we will show that complex measures in $\mathbf{R}^n$ with the same Fourier transforms are equal and we get the following

**Theorem 6.2.2.** *Let $(\xi_k)_{k=1}^n$ be a centred real-valued Gaussian process with uncorrelated components, that is*

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k.$$

*Then the random variables $\xi_1, ..., \xi_n$ are independent.*

### 6.3 Fourier Inversion

**Theorem 6.3.1.** *Suppose $f \in L^1(m_n)$. If $\hat{f} \in L^1(m_n)$ and $f$ is bounded and continuous*

$$f(x) = \int_{\mathbf{R}^d} e^{i\langle y,x\rangle} \hat{f}(y) \frac{dy}{(2\pi)^n},\ x \in \mathbf{R}^n.$$

PROOF. Choose $\varepsilon > 0$. We have

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle y,x-u\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \frac{dy}{(2\pi)^n} \right\} du$$

where the right side equals

$$\int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle v,\frac{x-u}{\varepsilon}\rangle} e^{-\frac{1}{2}|v|^2} \frac{dv}{\sqrt{2\pi}^n} \right\} \frac{du}{\sqrt{2\pi}^n \varepsilon^n} = \int_{\mathbf{R}^n} f(u) e^{-\frac{1}{2\varepsilon^2}|u-x|^2} \frac{du}{\sqrt{2\pi}^n \varepsilon^n}$$

$$= \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

Thus

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

By letting $\varepsilon \to 0$ and using the Lebesgue Dominated Convergence Theorem, Theorem 6.3.1 follows at once.

Recall that $C_c^\infty(\mathbf{R}^n)$ denotes the class of all functions $f : \mathbf{R}^n \to \mathbf{R}$ with compact support which are infinitely many times differentiable. If $f \in C_c^\infty(\mathbf{R}^n)$ then $\hat{f} \in L^1(m_n)$. To see this, suppose $y_k \neq 0$ and use partial integration to obtain

$$\hat{f}(y) = \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f(x) dx = \frac{1}{iy_k} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f'_{x_k}(x) dx$$

and

$$\hat{f}(y) = \frac{1}{(iy_k)^l} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f^{(l)}_{x_k}(x) dx,\ l \in \mathbf{N}.$$

Thus

$$| y_k |^l | \hat{f}(y) | \leq \int_{\mathbf{R}^n} | f^{(l)}_{x_k}(x) | \, dx, \; l \in \mathbf{N}$$

and we conclude that

$$\sup_{y \in \mathbf{R}^n} (1+ | y |)^{n+1} | \hat{f}(y) | < \infty.$$

and, hence, $\hat{f} \in L^1(m_n)$.

**Corollary 6.3.1.** *If* $f \in C_c^\infty(\mathbf{R}^n)$, *then* $\hat{f} \in L^1(m_n)$ *and*

$$f(x) = \int_{\mathbf{R}^n} e^{i\langle y,x \rangle} \hat{f}(y) \frac{dy}{(2\pi)^n}, \; x \in \mathbf{R}^n.$$

**Corollary 6.3.2** *If* $\mu$ *is a complex Borel measure in* $\mathbf{R}^n$ *and* $\hat{\mu} = 0$, *then* $\mu = 0$.

PROOF. Choose $f \in C_c^\infty(\mathbf{R}^n)$. We multiply the equation $\hat{\mu}(-y) = 0$ by $\frac{\hat{f}(y)}{(2\pi)^n}$ and integrate over $\mathbf{R}^n$ with respect to Lebesgue measure to obtain

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = 0.$$

Since $f \in C_c^\infty(\mathbf{R}^n)$ is arbitrary it follows that $\mu = 0$. The theorem is proved.

### 6.4. Non-Differentiability of Brownian Paths

Let $ND$ denote the set of all real-valued continuous function defined on the unit interval which are not differentiable at any point. It is well known that $ND$ is non-empty. In fact, if $\nu$ is Wiener measure on $C\,[0,1]$, $x \in ND$ a.e. $[\nu]$. The purpose of this section is to prove this important property of Brownian motion.

Let $W = (W(t))_{0\le t\le 1}$ be a real-valued Brownian motion in the time interval $[0,1]$ such that every path $t \to W(t)$, $0 \le t \le 1$ is continuous. Recall that

$$E\,[W(t)] = 0$$

and

$$E\,[W(s)W(t)] = \min(s,t).$$

If

$$0 \le t_0 \le ... \le t_n \le 1$$

and $1 \le j < k \le n$

$$E\,[(W(t_k) - W(t_{k-1}))(W(t_j) - W(t_{j-1})]$$

$$= E\,[(W(t_k)W(t_j)] - E\,[W(t_k)W(t_{j-1})] - E\,[W(t_{k-1})W(t_j)] + E\,[W(t_{k-1})W(t_{j-1})]$$

$$= t_j - t_{j-1} - t_j + t_{j-1} = 0.$$

From the previous section we now infer that the random variables

$$W(t_1) - W(t_0), ..., W(t_n) - W(t_{n-1})$$

are independent.

**Theorem 7.** *The function* $t \to W(t)$, $0 \le t \le 1$ *is not differentiable at any point* $t \in [0,1]$ a.s. $[P]$.

PROOF. Without loss of generality we assume the underlying probability space is complete. Let $c, \varepsilon > 0$ and denote by $B(c,\varepsilon)$ the set of all $\omega \in \Omega$ such that

$$\mid W(t) - W(s) \mid < c \mid t - s \mid \text{ if } t \in [s-\varepsilon, s+\varepsilon] \cap [0,1]$$

for some $s \in [0,1]$. It is enough to prove that the set

$$\bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} B(j, \frac{1}{k}).$$

is of probability zero. From now on let $c, \varepsilon > 0$ be fixed. It is enough to prove $P\,[B(c,\varepsilon)] = 0$ .

Set

$$X_{n,k} = \max_{k \leq j < k+3} \mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid$$

for each integer $n > 3$ and $k \in \{0, ..., n-3\}$.

Let $n > 3$ be so large that

$$\frac{3}{n} \leq \varepsilon.$$

We claim that

$$B(c, \varepsilon) \subseteq \left[ \min_{0 \leq k \leq n-3} X_{n,k} \leq \frac{6c}{n} \right].$$

If $\omega \in B(c, \varepsilon)$ there exists an $s \in [0, 1]$ such that

$$\mid W(t) - W(s) \mid \leq c \mid t - s \mid \text{ if } t \in [s - \varepsilon, s + \varepsilon] \cap [0, 1].$$

Now choose $k \in \{0, ..., n-3\}$ such that

$$s \in \left[ \frac{k}{n}, \frac{k}{n} + \frac{3}{n} \right].$$

If $k \leq j < k + 3$,

$$\mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid \leq \mid W(\frac{j+1}{n}) - W(s) \mid + \mid W(s) - W(\frac{j}{n}) \mid$$

$$\leq \frac{6c}{n}$$

and, hence, $X_{n,k} \leq \frac{6c}{n}$. Now

$$B(c, \varepsilon) \subseteq \left[ \min_{0 \leq k \leq n-3} X_{n,k} \leq \frac{6c}{n} \right]$$

and it is enough to prove that

$$\lim_{n \to \infty} P \left[ \min_{0 \leq k \leq n-3} X_{n,k} \leq \frac{6c}{n} \right] = 0.$$

But

$$P \left[ \min_{0 \leq k \leq n-3} X_{n,k} \leq \frac{6c}{n} \right] \leq \sum_{k=0}^{n-3} P \left[ X_{n,k} \leq \frac{6c}{n} \right]$$

$$= (n-2)P\left[X_{n,0} \le \frac{6c}{n}\right] \le nP\left[X_{n,0} \le \frac{6c}{n}\right]$$

$$= n(P\left[\mid W(\frac{1}{n}) \mid \le \frac{6c}{n}\right])^3 = n(P(\mid W(1) \mid \le \frac{6c}{\sqrt{n}})^3$$

$$\le n(\frac{12c}{\sqrt{2\pi n}})^3.$$

where the right side converges to zero as $n \to \infty$. The theorem is proved.

Recall that a function of bounded variation possesses a derivative a.e. with respect to Lebesgue measure. Therefore, with probability one, a Brownian path is not of bounded variation. In view of this an integral of the type

$$\int_0^1 f(t)dW(t)$$

cannot be interpreted as an ordinary Stieltjes integral. Nevertheless, such an integral can be defined by completely different means and is basic in, for example, financial mathematics.

↑↑↑

# CHAPTER 6

# COMPLEX INTEGRATION

## Introduction

In this section, in order to illustrate the power of Lebesgue integration, we collect a few results, which often appear with uncomplete proofs at the undergraduate level.

## 6.1. Complex Integrand

So far we have only treated integration of functions with their values in **R** or $[0, \infty]$ and it is the purpose of this section to discuss integration of complex valued functions.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure. Let $f, g \in L^1(\mu)$. We define

$$\int_X (f + ig)d\mu = \int_X f d\mu + i \int_X g d\mu.$$

If $\alpha$ and $\beta$ are real numbers,

$$\int_X (\alpha + i\beta)(f + ig)d\mu = \int_X ((\alpha f - \beta g) + i(\alpha g + \beta f))d\mu$$

$$= \int_X (\alpha f - \beta g)d\mu + i \int_X (\alpha g + \beta f)d\mu$$

$$= \alpha \int_X f d\mu - \beta \int_X g d\mu + i\alpha \int_X g d\mu + i\beta \int_X f d\mu$$

$$= (\alpha + i\beta)(\int_X f d\mu + i \int_X g d\mu)$$

$$= (\alpha + i\beta) \int_X (f + ig)d\mu.$$

We write $f \in L^1(\mu; \mathbf{C})$ if $\mathrm{Re}\, f$, $\mathrm{Im}\, f \in L^1(\mu)$ and have, for every $f \in L^1(\mu; \mathbf{C})$ and complex $\alpha$,

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu.$$

Clearly, if $f, g \in L^1(\mu; \mathbf{C})$, then

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Now suppose $\mu$ is a complex measure on $\mathcal{M}$. If

$$f \in L^1(\mu; \mathbf{C}) =_{def} L^1(\mu_{\mathrm{Re}}; \mathbf{C}) \cap L^1(\mu_{\mathrm{Im}}; \mathbf{C})$$

we define

$$\int_X f d\mu = \int_X f d\mu_{\mathrm{Re}} + i \int_X f d\mu_{\mathrm{Im}}.$$

It follows for every $f, g \in L^1(\mu; \mathbf{C})$ and $\alpha \in \mathbf{C}$ that

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu.$$

and

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$\downarrow\downarrow\downarrow$

**6.2. The Fourier Transform**

Below, if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n) \in \mathbf{R}^n$, we let

$$\langle x, y \rangle = \Sigma_{k=1}^n x_k y_k.$$

and

$$\mid x \mid = \sqrt{\langle x, y \rangle}.$$

If $\mu$ is a complex measure on $\mathcal{R}_n$ (or $\mathcal{R}_n^-$) the Fourier transform $\hat{\mu}$ of $\mu$ is defined by

$$\hat{\mu}(y) = \int_{\mathbf{R}^n} e^{-i\langle x,y \rangle} d\mu(x),\ y \in \mathbf{R}^n.$$

Note that

$$\hat{\mu}(0) = \mu(\mathbf{R}^n).$$

The Fourier transform of a function $f \in L^1(m_n; \mathbf{C})$ is defined by

$$\hat{f}(y) = \hat{\mu}(y) \text{ where } d\mu = f dm_n.$$

**Theorem 6.2.1.** *The canonical Gaussian measure* $\gamma_n$ *in* $\mathbf{R}^n$ *has the Fourier transform*

$$\hat{\gamma}_n(y) = e^{-\frac{|y|^2}{2}}.$$

PROOF. Since

$$\gamma_n = \gamma_1 \otimes ... \otimes \gamma_1 \text{ ($n$ factors)}$$

it is enough to consider the special case $n = 1$. Set

$$g(y) = \hat{\gamma}_1(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx.$$

Note that $g(0) = 1$. Since

$$\mid \frac{\cos x(y+h) - \cos xy}{h} \mid \leq \mid x \mid$$

the Lebesgue Dominated Convergence Theorem yields

$$g'(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} -xe^{-\frac{x^2}{2}} \sin xy dx$$

(Exercise: Prove this by using Example 2.2.1). Now, by partial integration,

$$g'(y) = \frac{1}{\sqrt{2\pi}} \left[ e^{-\frac{x^2}{2}} \sin xy \right]_{x=-\infty}^{x=\infty} - \frac{y}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx$$

that is

$$g'(y) + yg(y) = 0$$

and we get

$$g(y) = e^{-\frac{y^2}{2}}.$$

If $\xi = (\xi_1, ..., \xi_n)$ is an $\mathbf{R}^n$-valued random variable with $\xi_k \in L^1(P)$, $k = 1, ..., n$, the characteristic function $c_\xi$ of $\xi$ is defined by

$$c_\xi(y) = E\left[e^{i\langle \xi, y\rangle}\right] = \hat{P}_\xi(-y),\ y \in \mathbf{R}^n.$$

For example, if $\xi \in N(0, \sigma)$, then $\xi = \sigma G$, where $G \in N(0,1)$, and we get

$$c_\xi(y) = E\left[e^{i\langle G, \sigma y\rangle}\right] = \hat{\gamma}_1(-\sigma y)$$

$$= e^{-\frac{\sigma^2 y^2}{2}}.$$

Choosing $y = 1$ results in

$$E\left[e^{i\xi}\right] = e^{-\frac{1}{2}E\left[\xi^2\right]} \text{ if } \xi \in N(0, \sigma).$$

Thus if $(\xi_k)_{k=1}^n$ is a centred real-valued Gaussian process

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \exp(-\frac{1}{2}E\left[(\Sigma_{k=1}^n y_k \xi_k)^2\right]$$

$$= \exp(-\frac{1}{2}\Sigma_{k=1}^n y_k^2 E\left[\xi_k^2\right] - \Sigma_{1\leq j<k\leq n} y_j y_k E\left[\xi_j \xi_k\right]).$$

In particular, if

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k$$

we see that

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n e^{-\frac{y_k^2}{2}E\left[\xi_k^2\right]}$$

or

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n E\left[e^{i y_k \xi_k}\right].$$

Stated otherwise, the Fourier tranforms of the measures $P_{(\xi_1,...,\xi_n)}$ and $\times_{k=1}^n P_{\xi_k}$ agree. Below we will show that complex measures in $\mathbf{R}^n$ with the same Fourier transforms are equal and we get the following

**Theorem 6.2.2.** *Let $(\xi_k)_{k=1}^n$ be a centred real-valued Gaussian process with uncorrelated components, that is*

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k.$$

*Then the random variables $\xi_1, ..., \xi_n$ are independent.*

### 6.3 Fourier Inversion

**Theorem 6.3.1.** *Suppose $f \in L^1(m_n)$. If $\hat{f} \in L^1(m_n)$ and $f$ is bounded and continuous*

$$f(x) = \int_{\mathbf{R}^d} e^{i\langle y,x\rangle} \hat{f}(y) \frac{dy}{(2\pi)^n},\ x \in \mathbf{R}^n.$$

PROOF. Choose $\varepsilon > 0$. We have

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle y,x-u\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \frac{dy}{(2\pi)^n} \right\} du$$

where the right side equals

$$\int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle v,\frac{x-u}{\varepsilon}\rangle} e^{-\frac{1}{2}|v|^2} \frac{dv}{\sqrt{2\pi}^n} \right\} \frac{du}{\sqrt{2\pi}^n \varepsilon^n} = \int_{\mathbf{R}^n} f(u) e^{-\frac{1}{2\varepsilon^2}|u-x|^2} \frac{du}{\sqrt{2\pi}^n \varepsilon^n}$$

$$= \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

Thus

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

By letting $\varepsilon \to 0$ and using the Lebesgue Dominated Convergence Theorem, Theorem 6.3.1 follows at once.

Recall that $C_c^\infty(\mathbf{R}^n)$ denotes the class of all functions $f : \mathbf{R}^n \to \mathbf{R}$ with compact support which are infinitely many times differentiable. If $f \in C_c^\infty(\mathbf{R}^n)$ then $\hat{f} \in L^1(m_n)$. To see this, suppose $y_k \neq 0$ and use partial integration to obtain

$$\hat{f}(y) = \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f(x) dx = \frac{1}{iy_k} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f'_{x_k}(x) dx$$

and

$$\hat{f}(y) = \frac{1}{(iy_k)^l} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f^{(l)}_{x_k}(x) dx,\ l \in \mathbf{N}.$$

Thus

$$| y_k |^l | \hat{f}(y) | \leq \int_{\mathbf{R}^n} | f^{(l)}_{x_k}(x) | \, dx, \; l \in \mathbf{N}$$

and we conclude that

$$\sup_{y \in \mathbf{R}^n} (1+ | y |)^{n+1} | \hat{f}(y) | < \infty.$$

and, hence, $\hat{f} \in L^1(m_n)$.

**Corollary 6.3.1.** *If* $f \in C_c^\infty(\mathbf{R}^n)$, *then* $\hat{f} \in L^1(m_n)$ *and*

$$f(x) = \int_{\mathbf{R}^n} e^{i\langle y,x \rangle} \hat{f}(y) \frac{dy}{(2\pi)^n}, \; x \in \mathbf{R}^n.$$

**Corollary 6.3.2** *If* $\mu$ *is a complex Borel measure in* $\mathbf{R}^n$ *and* $\hat{\mu} = 0$, *then* $\mu = 0$.

PROOF. Choose $f \in C_c^\infty(\mathbf{R}^n)$. We multiply the equation $\hat{\mu}(-y) = 0$ by $\frac{\hat{f}(y)}{(2\pi)^n}$ and integrate over $\mathbf{R}^n$ with respect to Lebesgue measure to obtain

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = 0.$$

Since $f \in C_c^\infty(\mathbf{R}^n)$ is arbitrary it follows that $\mu = 0$. The theorem is proved.

### 6.4. Non-Differentiability of Brownian Paths

Let $ND$ denote the set of all real-valued continuous function defined on the unit interval which are not differentiable at any point. It is well known that $ND$ is non-empty. In fact, if $\nu$ is Wiener measure on $C\,[0,1]$, $x \in ND$ a.e. $[\nu]$. The purpose of this section is to prove this important property of Brownian motion.

Let $W = (W(t))_{0 \leq t \leq 1}$ be a real-valued Brownian motion in the time interval $[0,1]$ such that every path $t \to W(t)$, $0 \leq t \leq 1$ is continuous. Recall that

$$E[W(t)] = 0$$

and

$$E[W(s)W(t)] = \min(s,t).$$

If

$$0 \leq t_0 \leq ... \leq t_n \leq 1$$

and $1 \leq j < k \leq n$

$$E[(W(t_k) - W(t_{k-1}))(W(t_j) - W(t_{j-1})]$$

$$= E[(W(t_k)W(t_j)] - E[W(t_k)W(t_{j-1})] - E[W(t_{k-1})W(t_j)] + E[W(t_{k-1})W(t_{j-1})]$$

$$= t_j - t_{j-1} - t_j + t_{j-1} = 0.$$

From the previous section we now infer that the random variables

$$W(t_1) - W(t_0), ..., W(t_n) - W(t_{n-1})$$

are independent.

**Theorem 7.** *The function* $t \to W(t)$, $0 \leq t \leq 1$ *is not differentiable at any point* $t \in [0,1]$ a.s. $[P]$.

PROOF. Without loss of generality we assume the underlying probability space is complete. Let $c, \varepsilon > 0$ and denote by $B(c,\varepsilon)$ the set of all $\omega \in \Omega$ such that

$$\mid W(t) - W(s) \mid < c \mid t - s \mid \text{ if } t \in [s-\varepsilon, s+\varepsilon] \cap [0,1]$$

for some $s \in [0,1]$. It is enough to prove that the set

$$\bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} B(j, \frac{1}{k}).$$

is of probability zero. From now on let $c, \varepsilon > 0$ be fixed. It is enough to prove $P[B(c,\varepsilon)] = 0$ .

Set

$$X_{n,k} = \max_{k \le j < k+3} \mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid$$

for each integer $n > 3$ and $k \in \{0, ..., n-3\}$.

Let $n > 3$ be so large that

$$\frac{3}{n} \le \varepsilon.$$

We claim that

$$B(c, \varepsilon) \subseteq \left[ \min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right].$$

If $\omega \in B(c, \varepsilon)$ there exists an $s \in [0, 1]$ such that

$$\mid W(t) - W(s) \mid \le c \mid t - s \mid \text{ if } t \in [s - \varepsilon, s + \varepsilon] \cap [0, 1].$$

Now choose $k \in \{0, ..., n-3\}$ such that

$$s \in \left[ \frac{k}{n}, \frac{k}{n} + \frac{3}{n} \right].$$

If $k \le j < k + 3$,

$$\mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid \le \mid W(\frac{j+1}{n}) - W(s) \mid + \mid W(s) - W(\frac{j}{n}) \mid$$

$$\le \frac{6c}{n}$$

and, hence, $X_{n,k} \le \frac{6c}{n}$. Now

$$B(c, \varepsilon) \subseteq \left[ \min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right]$$

and it is enough to prove that

$$\lim_{n \to \infty} P \left[ \min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] = 0.$$

But

$$P \left[ \min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] \le \sum_{k=0}^{n-3} P \left[ X_{n,k} \le \frac{6c}{n} \right]$$

$$= (n-2)P\left[X_{n,0} \leq \frac{6c}{n}\right] \leq nP\left[X_{n,0} \leq \frac{6c}{n}\right]$$

$$= n(P\left[\mid W(\frac{1}{n}) \mid \leq \frac{6c}{n}\right])^3 = n(P(\mid W(1) \mid \leq \frac{6c}{\sqrt{n}})^3$$

$$\leq n(\frac{12c}{\sqrt{2\pi n}})^3.$$

where the right side converges to zero as $n \to \infty$. The theorem is proved.

Recall that a function of bounded variation possesses a derivative a.e. with respect to Lebesgue measure. Therefore, with probability one, a Brownian path is not of bounded variation. In view of this an integral of the type

$$\int_0^1 f(t)dW(t)$$

cannot be interpreted as an ordinary Stieltjes integral. Nevertheless, such an integral can be defined by completely different means and is basic in, for example, financial mathematics.

↑↑↑

# REFERENCES

[$B$] Billingsley, P. (1995) Probability and Measure. Wiley&Sons.

[$BO$] Bogachev, V. I. (2007) Measure Theory, Springer.

[$D$] Dudley, R. M. (1989) Real Analysis and Probability. Wadsworth&Brooks.

[$DI$] Diedonné, J. (1960) Foundations of Modern Analysis. Academic Press.

[$F$] Folland, G. B. (1999) Real Analysis; Modern Techniques and Their Applications. Second Edition. Wiley&Sons.

[$M$] Malliavin, P. (1995) Integration and Probability. Springer.

[$R$] Rudin, W. (1966) Real and Complex Analysis. McGraw-Hill.

[$S$] Solovay R. M. (1970) A model of set theory in which every set of reals is Lebesgue measurable. Annals of Mathematics **92**, 1-56.