# Contents

1 Introduction and Preliminaries 3
   1 Introduction: Riemann to Lebesgue ........................................ 3
   2 Preliminaries ................................................................. 6

2 Measurable Sets and Measurable Functions 11
   3 \( \sigma \)-Algebras ............................................................... 11
   4 Measurable Functions ...................................................... 16

3 Measures and Integrals 22
   5 Measures ................................................................. 22
   6 Integral of Non-Negative Simple Functions ............................. 25
   7 Integral of Non-Negative Extended Real-Valued Functions .......... 28
   8 Integral of Extended Real-Valued Functions ............................ 33
   9 Almost Everywhere ......................................................... 36
  10 Fatou's Lemma and Dominated Convergence Theorem .................. 40

4 Spaces of Functions 44
   11 Normed Spaces and Banach Spaces ...................................... 44
   12 \( L^p \): Definitions and Basic Properties ............................ 46
   13 \( L^p \): Completeness ......................................................... 49
   14 \( B(X) \) and \( C(X) \) .......................................................... 51
5 Some Generalizations

16 Complex-Valued Measurable Functions

6 Construction of Measures

17 Outline

18 Set Functions

19 Caratheodory’s Theorem: Outer Measures to Measures

20 Construction of Outer Measures

21 Existence and Uniqueness of Extensions

7 Lebesgue Measure on $\mathbb{R}$

22 Lebesgue Measure on $\mathbb{R}$

23 Complete Measures

24 Completion of Measures (Optional)

25 Riemann Integral

26 Comparison of Lebesgue Integral and Riemann Integral

8 Product Measures and Fubini’s Theorem

27 Measurable Rectangles and Product $\sigma$-Algebras

28 Product Measures

29 Fubini’s Theorem

30 Product Spaces With More Than Two Factors

31 Lebesgue Measure on $\mathbb{R}^n$
Chapter 1

Introduction and Preliminaries

1 Introduction: Riemann to Lebesgue

You are probably familiar with the Riemann integral from calculus and undergraduate analysis. If \( f \) is a non-negative real-valued function defined on an interval \([a, b]\), then, roughly speaking, the Riemann integral of \( f \) is a limit of Riemann sums

\[
\int_a^b f(x)\,dx = \lim_{n \to \infty} \sum_{i=1}^n f(x_i^*) (x_i - x_{i-1}),
\]

where \([a, b]\) is divided into subintervals \([x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]\) with \( a = x_0 \leq x_1 \leq \ldots \leq x_n = b \), the (sample) points \( x_i^* \in [x_{i-1}, x_i] \) are arbitrary, and \((x_i - x_{i-1})\) is the length of \([x_{i-1}, x_i]\. We say \( f \) is Riemann integrable on \([a, b]\) when its Riemann integral is defined.

In this course, we will study Lebesgue’s theory of integration with respect to a measure. Roughly, a measure \( \mu \) on a set \( X \) is a function which takes subsets \( A \subseteq X \) as inputs and gives non-negative real numbers \( \mu(A) \) as outputs. You can think of \( \mu(A) \) as some general notion of the size of \( A \). The most important measure is the Lebesgue measure on \( \mathbb{R} \), denoted by \( \lambda \). For each interval \([a, b]\), \( \lambda([a, b]) \) is the length \( b - a \). For other sets \( A \subseteq \mathbb{R} \), \( \lambda(A) \) is the length of \( A \) (we will see how to define this later.) There are many other useful measures. If \( f \) is a non-negative real-valued function defined on a set \( X \), then, roughly speaking, the integral of \( f \) with respect to \( \mu \) is

\[
\int_X f(x)d\mu(x) = \lim_{n \to \infty} \sum_{i=1}^n f(x_i^*) \mu(A_i)
\]

Here the set \( X \) is partitioned into subsets \( A_1, \ldots, A_n \), we choose sample points \( x_i^* \in A_i \), and \( \mu(A_i) \) is the measure of the set \( A_i \). When the integral is defined, we say \( f \) is \( \mu \)-integrable. When \( \mu \) is the Lebesgue measure \( \lambda \) on \( \mathbb{R} \), the integral is simply called the Lebesgue integral.

Lebesgue’s theory of integration has several advantages over Riemann’s theory.

(1) Lebesgue’s theory allows us to integrate functions defined on arbitrary sets, whereas the Riemann integral is restricted to functions defined on \( \mathbb{R} \) or \( \mathbb{R}^d \). This is especially important for certain
applications. For example, Lebesgue’s theory of measure and integration forms the rigorous foundation for modern probability theory, where the functions are random variables and we integrate over the sample space of possible outcomes.

(2) The Lebesgue integral (i.e., the integral with respect to Lebesgue measure) is defined for a larger class of functions on \( \mathbb{R} \), though it still agrees with the Riemann integral whenever the latter is defined. For example, the indicator function of the rationals

\[
1_Q = \begin{cases} 
1 & \text{if } x \in \mathbb{Q} \\
0 & \text{if } x \notin \mathbb{Q}
\end{cases}
\]

is not Riemann integrable on any interval \([a, b]\), but it is Lebesgue integrable on any such interval. We will see the details later.

(3) Lebesgue’s theory possesses better convergence theorems, which lead to more general elegant results and more useful spaces of functions.

Here is the type of convergence theorem we are interested in:

**Prototype Convergence Theorem.** If \((f_n)\) is a sequence of integrable functions defined on \([a, b]\) and \((f_n)\) converges to \(f\) “nicely”, then \(f\) is integrable and \(\lim_{n \to \infty} \int_a^b f_n = \int_a^b f\).

The problem is, of course, to figure out what “nicely” might mean and to define the integral in such a way that theorems like this one will be widely applicable. These were important unresolved issues in the late nineteenth century; they arose, for example, in the study of Fourier series. The Lebesgue theory was developed, in large part, to address this. Let us look at this problem in a bit more detail.

**Definition 1.1.** A sequence \((f_n)\) of functions \(f_n : X \to \mathbb{R}\) is said to **converge pointwise** to a function \(f : X \to \mathbb{R}\) if

\[
\lim_{n \to \infty} f_n(x) = f(x) \quad \text{for each } x \in X.
\]

In this case, we write \(f_n \to f\) pointwise.

**Example 1.2.** \(f_n = \) tent function = graph is triangle with vertices \((0, 0), (1/n, n), (2/n, 0)\), equals zero elsewhere. Then \(f_n \to f = 0\) pointwise, \(f_n\) is Riemann integrable on \([a, b]\), and \(f = 0\) is Riemann integrable on \([a, b]\), but \(\lim_{n \to \infty} \int_a^b f_n = 1 \neq 0 = \int_a^b f\). So we don’t get the desired conclusion of the Prototype Convergence Theorem.

**Example 1.3.** Choose an enumeration \(\mathbb{Q} = \{r_1, r_2, r_3, \ldots\}\). Define \(f = 1_Q\) and define \(f_n\) by \(f_n(x) = 1\) if \(x \in \{r_1, \ldots, r_n\}\) and \(f_n(x) = 0\) otherwise. Then \(f_n \to f\) pointwise, and \(f_n\) is Riemann integrable on \([a, b]\), but \(f\) is not Riemann integrable on \([a, b]\). Again we don’t get the desired conclusion of the Prototype Convergence Theorem. Note that here the sequence \((f_n)\) is quite nice. Indeed, \((f_n)\) is bounded (\(|f_n| \leq 1\) for all \(n\)) and \((f_n)\) is increasing (\(f_n \leq f_{n+1}\) for all \(n\)). But it’s not enough.

**Example 1.4.** Later, we will construct a sequence of functions \(f_n : [0, 1] \to \mathbb{R}\) such that

- \(0 \leq f_n \leq 1\) for each \(n\)
- \(f_n\) is continuous for each \(n\)
• \((f_n)\) is an increasing sequence

• \((f_n)\) converges pointwise to a function \(f\) on \([0,1]\).

• \(f\) is not Riemann integrable on \([0,1]\).

This sequence is even nicer than the one in the previous example, but we still don’t get the desired conclusion of the Prototype Convergence Theorem.

The examples above show that pointwise convergence isn’t good enough for Riemann integration. By assuming more, we can get some convergence theorems for Riemann integration. We describe three such theorems below. Unfortunately, they are all somewhat unsatisfactory.

The first theorem requires uniform convergence.

**Definition 1.5.** A sequence \((f_n)\) of functions \(f_n : X \to \mathbb{R}\) is said to **converge uniformly** to a function \(f : X \to \mathbb{R}\) if

\[
\lim_{n \to \infty} \sup \{|f_n(x) - f(x)| : x \in X\} = 0.
\]

In this case, we write \(f_n \to f\) uniformly.

Note that uniform convergence implies pointwise convergence.

**Theorem 1.6.** If \(f_n\) is Riemann integrable on \([a,b]\) for each \(n\) and \(f_n \to f\) uniformly on \([a,b]\), then \(f\) is Riemann integrable on \([a,b]\) and \(\lim_{n \to \infty} \int_a^b f_n = \int_a^b f\).

The next two theorems require, respectively, the sequence of functions to be increasing or bounded.

**Theorem 1.7.** If \(f_n\) is Riemann integrable on \([a,b]\) for each \(n\), \(f_n \leq f_{n+1}\) for each \(n\), \(f_n \to f\) pointwise on \([a,b]\), and \(f\) is Riemann integrable on \([a,b]\), then \(\lim_{n \to \infty} \int_a^b f_n = \int_a^b f\).

**Theorem 1.8.** If \(f_n\) is Riemann integrable on \([a,b]\) for each \(n\), \(|f_n| \leq M\) for each \(n\), \(f_n \to f\) pointwise on \([a,b]\), and \(f\) is Riemann integrable on \([a,b]\), then \(\lim_{n \to \infty} \int_a^b f_n = \int_a^b f\).

The problem with the first of these three theorems is that uniform convergence is too strong. In many applications, we don’t have it. The problem with the last two theorems is that the Riemann integrability of \(f\) is part of the hypothesis, rather than part of the conclusion.

We will see that the Lebesgue theory gives us very powerful convergence theorems, which lead to some very impressive results and some very useful function spaces.
2 Preliminaries

Set Theory

Definition 2.1.

• If $A$ and $B$ are sets, their union and intersection are
  \[ A \cup B = \{ x : x \in A \text{ or } x \in B \} \]
  \[ A \cap B = \{ x : x \in A \text{ and } x \in B \} \]

• If \( \{A_i : i \in I\} \) is an indexed family of sets, the union and intersection of the family are
  \[ \bigcup_{i \in I} A_i = \{ x : x \in A_i \text{ for at least one } i \in I \} \]
  \[ \bigcap_{i \in I} A_i = \{ x : x \in A_i \text{ for all } i \in I \} \]

• If $A$ and $B$ are sets, the set difference of $A$ and $B$ (or the relative complement of $B$ in $A$) is
  \[ A \setminus B = \{ x : x \in A \text{ and } x \notin B \} \]
  It is read as “$A$ minus $B$” or “$A$ take away $B$”.

• If all the sets in a given context are subsets of a fixed set $X$, then the complement (or absolute complement) of a set $A \subseteq X$ is
  \[ A^c = X \setminus A = \{ x \in X : x \notin A \} \]

Properties of Complement:

Suppose all sets under consideration are subsets of a fixed set $X$. Let $A, B \subseteq X$ and let \( \{A_i : i \in I\} \) be an indexed family of subsets of $X$.

• $A \cup A^c = X$
• $A \cap A^c = \emptyset$
• $\emptyset^c = X$
• $X^c = \emptyset$
• $(A^c)^c = A$
• $A \setminus B = A \cap B^c = B^c \setminus A^c$
• If $A \subseteq B$, then $B^c \subseteq A^c$
• De Morgan’s Laws:
  \[ \left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c \quad \text{and} \quad \left( \bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c \]
In words, De Morgan’s Laws say that the complement of a union is the intersection of the complements, and the complement of an intersection is the union of the complements.

**Properties of Union, Intersection, and Set Difference**

- \( B \cup \bigcup_{i \in I} A_i = \bigcup_{i \in I} (B \cup A_i) \)
- \( B \cap \bigcap_{i \in I} A_i = \bigcap_{i \in I} (B \cap A_i) \)
- \( B \cup \bigcap_{i \in I} A_i = \bigcap_{i \in I} (B \cup A_i) \)
- \( B \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I} (B \cap A_i) \)
- \( A = (A \cap B) \cup (A \setminus B) \)
- \( A \cap B = A \setminus (A \setminus B) \)
- \( A \setminus (B \cup C) = (A \setminus B) \setminus C \)
- \( (A \cup B) \setminus C = (A \setminus B) \cup (A \setminus C) \)

**Definition 2.2.** For any set \( X \), the **power set** of \( X \), denoted by \( \mathcal{P}(X) \), is the collection of all subsets of \( X \):

\[ \mathcal{P}(X) = \{ A : A \subseteq X \} \]

**Definition 2.3.** Let \( X, Y \) be sets. A **function** (or **map**) \( f : X \to Y \) is a rule that assigns to each element \( x \in X \) a unique element \( f(x) \in Y \). The sets \( X \) and \( Y \) are called the **domain** and **codomain** of \( f \). The **image** (or **range**) of \( f \) is the set

\[ f(X) = \{ f(x) : x \in X \} \]

If \( A \subseteq X \), the **image of** \( A \) **under** \( f \) is the set

\[ f(A) = \{ f(x) : x \in A \} \]

If \( B \subseteq Y \), the **inverse image** (or **preimage**) of \( B \) under \( f \) is the set

\[ f^{-1}(B) = \{ x \in X : f(x) \in B \} \]

The inverse image commutes with unions, intersections, set differences, and complements:

- \( f^{-1} \left( \bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} f^{-1}(A_i) \)
- \( f^{-1} \left( \bigcap_{i \in I} A_i \right) = \bigcap_{i \in I} f^{-1}(A_i) \)
- \( f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B) \)
- \( f^{-1}(A^c) = (f^{-1}(A))^c \)

**Remark 2.4.** The image is not so well-behaved. See Exercise ???.

**Definition 2.5.** If \( f : X \to Y \) and \( g : Y \to Z \), the **composition** of \( f \) and \( g \) is the function

\[ g \circ f : X \to Z \]

defined by

\[ (g \circ f)(x) = g(f(x)) \quad \text{for every } x \in X. \]
Extended Real Numbers

The set of extended real numbers is the set obtained by adjoining the two symbols $-\infty$ and $+\infty$ to the set of real numbers. It is denoted by $\mathbb{R}$. Thus

$$\mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}.$$ 

We often write $\infty$ instead of $+\infty$. We extend the usual ordering on $\mathbb{R}$ to $\mathbb{R}$ by declaring that

$$-\infty < x < \infty \quad \text{for every } x \in \mathbb{R}.$$ 

The interval notation is extended in the natural way. For example,

$$(0, \infty] = (0, \infty) \cup \{\infty\} = \{x \in \mathbb{R} : x > 0\} = \{x \in \mathbb{R} : 0 < x \leq \infty\}.$$ 

We extend the usual arithmetic operations on $\mathbb{R}$ to $\mathbb{R}$ by declaring that

- $x + \infty = \infty$ and $x - \infty = -\infty$ for all $x \in \mathbb{R}$
- $\infty + \infty = \infty$ and $-\infty - \infty = -\infty$
- $x \cdot (\pm \infty) = \pm \infty$ for all $x \in (0, \infty]$
- $x \cdot (\pm \infty) = \mp \infty$ for all $x \in [-\infty, 0)$
- $0 \cdot (\pm \infty) = 0$
- $\frac{x}{\pm \infty} = 0$ for all $x \in \mathbb{R}$.

Expressions of the form $\infty - \infty$ and $\frac{\infty}{\infty}$ are left undefined.

In some other areas of mathematics, the products $0 \cdot (\pm \infty) = 0$ are left undefined, but not here.

The absolute values of $-\infty$ and $+\infty$ are

$$| - \infty | = | + \infty | = + \infty$$

Supremum and Infimum

Let $A \subseteq \mathbb{R}$.

An element $a_0 \in \mathbb{R}$ is called the **maximum** (or **greatest element**) of $A$ if $a_0 \in A$ and $a \leq a_0$ for every $a \in A$. The maximum of $A$ is denoted by $\max(A)$. It is easy to see that the maximum is unique if it exists, hence why we say “the maximum” instead of “a maximum”. The **minimum** (or **least element**) of $A$ is defined similarly; it is denoted by $\min(A)$ and is unique if it exists.
An element \( a_0 \in \mathbb{R} \) is called an **upper bound** of \( A \) if \( a \leq a_0 \) for every \( a \in A \). If \( A \) has an upper bound, it is said to be bounded above. A **lower bound** of \( A \) is defined similarly.

An element \( a_0 \in \mathbb{R} \) is called the **supremum** of \( A \) if it is the least upper bound of \( A \), i.e., if \( a_0 \) is an upper bound of \( A \) and \( a_0 \leq a'_0 \) for every upper bound \( a'_0 \) of \( A \). The supremum is denoted by \( \text{sup}(A) \). Since minimums are unique if they exist, the supremum is unique if it exists. The **infimum** of \( A \) is the greatest lower bound of \( A \); it is denoted by \( \text{inf}(A) \) and is unique if it exists.

If \( A \) is non-empty and bounded above in \( \mathbb{R} \), the order completeness axiom of the real numbers implies that \( A \) has a supremum \( \text{sup}(A) \) in \( \mathbb{R} \).

If \( A \) is non-empty but not bounded above in \( \mathbb{R} \), then \( \infty \) is the only upper bound for \( A \), and so \( \text{sup}(A) = \infty \).

If \( A \) is empty, then every extended real number is an upper bound for \( A \), and so \( \text{sup}(A) = -\infty \).

Every subset of \( \mathbb{R} \) has a supremum and an infimum in \( \mathbb{R} \).

Thus every subset of \( \mathbb{R} \) has a supremum in \( \mathbb{R} \). The same goes for infimum.

**Limits, Limsup, Liminf**

Let \((a_n)\) be a sequence in \( \mathbb{R} \).

**Definition 2.6.**

- Let \( a \in \mathbb{R} \). We write \( \lim a_n = a \) (and we say that \((a_n)\) converges to \( a \) and that \( a \) is the limit of \((a_n)\)) if for every real \( \epsilon > 0 \) there exists an \( N \in \mathbb{N} \) such that if \( n \geq N \) then \(|a_n - a| < \epsilon\).

- We write \( \lim a_n = +\infty \) (and we say that \((a_n)\) converges to \(+\infty \) and that \(+\infty \) is the limit of \((a_n)\)) if for every real \( M > 0 \) there exists an \( N \in \mathbb{N} \) such that if \( n \geq N \) then \( a_n > M \).

- We write \( \lim a_n = -\infty \) (and we say that \((a_n)\) converges to \(-\infty \) and that \(-\infty \) is the limit of \((a_n)\)) if for every real \( M > 0 \) there exists an \( N \in \mathbb{N} \) such that if \( n \geq N \) then \( a_n < -M \).

**Definition 2.7.** The **limsup** (or limit superior) and **liminf** (or limit inferior) are defined by

\[
\limsup a_n = \inf_{k \geq 1} \left( \sup_{n \geq k} a_n \right)
\]

\[
\liminf a_n = \sup_{k \geq 1} \left( \inf_{n \geq k} a_n \right)
\]

Note that
• The sequence $c_k = \left( \sup_{n \geq k} a_n \right)$ is a decreasing sequence

• The sequence $b_k = \left( \inf_{n \geq k} a_n \right)$ is an increasing sequence

• $\limsup a_n = \inf_{k \geq 1} \left( \sup_{n \geq k} a_n \right) = \lim_k \left( \sup_{n \geq k} a_n \right)$

• $\liminf a_n = \sup_{k \geq 1} \left( \inf_{n \geq k} a_n \right) = \lim_k \left( \inf_{n \geq k} a_n \right)$

• $b_k \leq a_k \leq c_k$ for all $k$

• $\liminf a_n \leq \limsup a_n$

Theorem 2.8. We have $\limsup a_n = \liminf a_n$ iff $\lim_n a_n = L$ for some $L \in \mathbb{R}$, in which case

$$\limsup a_n = \liminf a_n = \lim_n a_n.$$ 

Proof. Exercise. 

Infinite Series

Definition 2.9. Let $\sum_{n=1}^{\infty} a_n$ be an infinite series whose terms $a_n$ belong to $\mathbb{R}$. The series has a sum in $\mathbb{R}$ (i.e., the series converges in $\mathbb{R}$) if both

(a) $\infty$ and $-\infty$ do not both occur among the terms of $\sum_{n=1}^{\infty} a_n$, and

(b) the sequence $s_k = \sum_{n=1}^{k} a_n$ of partial sums of $\sum_{n=1}^{\infty} a_n$ has a limit in $\mathbb{R}$

The sum of the series $\sum_{n=1}^{\infty} a_n$ is then defined to be the limit of the partial sums:

$$\lim s_k = \lim_k \sum_{n=1}^{k} a_n = \sum_{n=1}^{\infty} a_n.$$ 

Remark 2.10. (i) The sum of the series may be $-\infty$ or $+\infty$.

(ii) Note that condition (a) above is needed to guarantee that the undefined expression $\infty - \infty$ does not occur in any of partial sums $s_k = \sum_{n=1}^{k} a_n$.

(iii) If $\infty$ is one of the terms of the series and $-\infty$ is not, then the sum of the series is $\infty$. Likewise if we swap the roles of $\infty$ and $-\infty$. 
Chapter 2

Measurable Sets and Measurable Functions

3 σ-Algebras

Definition 3.1. Let $X$ be any set. A \textit{σ-algebra} on $X$ is a collection $\mathcal{A} \subseteq \mathcal{P}(X)$ that satisfies the following properties.

(i) $\emptyset \in \mathcal{A}$

(ii) If $A \in \mathcal{A}$, then $A^c = X \setminus A \in \mathcal{A}$.

(iii) If $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair $(X, \mathcal{A})$ is called a \textit{measurable space}. The sets in $\mathcal{A}$ are called \textit{measurable} sets (or \textit{\mathcal{A}-measurable} sets to avoid ambiguity).

Remark: \(\sigma\) indicates “countable union”

Theorem 3.2. If $\mathcal{A}$ is a σ-algebra on a set $X$, then $\mathcal{A}$ has the following properties:

(a) $X \in \mathcal{A}$

(b) If $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

(c) If $A_1, \ldots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

(d) If $A_1, \ldots, A_n \in \mathcal{A}$, then $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$.

(e) If $A, B \in \mathcal{A}$, then $A \setminus B = A \cap B^c \in \mathcal{A}$.

Proof. In this proof, (i),(ii),(iii) refer to the properties in the definition of a σ-algebra.

(a): By (i) and (ii), $X = \emptyset^c \in \mathcal{A}$. 
(b): Let $A_1, A_2, \ldots \in A$. By De Morgan’s laws

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

By (ii), $A_1^c, A_2^c, \ldots \in A$. Then, by (iii), $\bigcup_{i=1}^{\infty} A_i^c \in A$. Finally, by (b) again, $\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c \in A$.

(c): Define $A_i = \emptyset$ for $i > n$ and apply (iii).

(d): Define $A_i = X$ for $i > n$ and apply (b). Alternatively, use De Morgan’s laws and apply (ii) and (c).

(e): Use (ii) and (d).

Example 3.3. Let $X$ be any set. The following are $\sigma$-algebras on $X$.

(a) $\mathcal{P}(X)$

(b) $\{\emptyset, X\}$

(c) $\{\emptyset, A_0, A_0^c, X\}$ for any fixed $A_0 \subseteq X$.

(d) $\{A \subseteq X : A$ is countable or $A^c$ is countable$\}$

Theorem 3.4. Let $X$ be any set. The intersection of any collection of $\sigma$-algebras on $X$ is a $\sigma$-algebra on $X$.

Proof. Let $\{A_j\}_{j \in J}$ be any collection of $\sigma$-algebras on $X$ and consider their intersection $\bigcap_{j \in J} A_j$. We need to check properties (i), (ii), and (iii) of the definition of a $\sigma$-algebra.

(i): We have $\emptyset \in A_j$ for every $j \in J$. So $\emptyset \in \bigcap_{j \in J} A_j$.

(ii): Suppose $A \in \bigcap_{j \in J} A_j$. Then $A \in A_j$ for every $j \in J$. Since $A_j$ is a $\sigma$-algebra for every $j \in J$, we have $A^c \in A_j$ for every $j \in J$. So $A \in \bigcap_{j \in J} A_j$.

(iii): Suppose $A_1, A_2, \ldots \in \bigcap_{j \in J} A_j$. Then $A_1, A_2, \ldots \in A_j$ for every $j \in J$. Since $A_j$ is a $\sigma$-algebra for every $j \in J$, we have $\bigcap_{i=1}^{\infty} A_i \in A_j$ for every $j \in J$. So $\bigcap_{i=1}^{\infty} A_i \in \bigcap_{j \in J} A_j$.

Definition 3.5. Let $X$ be any set and let $\mathcal{G} \subseteq \mathcal{P}(X)$. The intersection of all $\sigma$-algebras on $X$ that contain $\mathcal{G}$ is denoted by $\sigma(\mathcal{G})$. In other words,

$$\sigma(\mathcal{G}) = \bigcap \{\mathcal{A} : \mathcal{A}$ is a $\sigma$-algebra on $X, \mathcal{G} \subseteq \mathcal{A}\}.$$

We call $\sigma(\mathcal{G})$ the $\sigma$-algebra generated by $\mathcal{G}$.

The following theorem is immediate from the definition of $\sigma(\mathcal{G})$.

Theorem 3.6. Let $X$ be any set and let $\mathcal{G} \subseteq \mathcal{P}(X)$. Then $\sigma(\mathcal{G})$ is the smallest $\sigma$-algebra on $X$ that contains $\sigma(\mathcal{G})$; in other words,
(i) $\sigma(\mathcal{G})$ is a $\sigma$-algebra on $X$

(ii) $\mathcal{G} \subseteq \sigma(\mathcal{G})$

(iii) If $\mathcal{A}$ is any $\sigma$-algebra on $X$ and $\mathcal{G} \subseteq \mathcal{A}$, then $\sigma(\mathcal{G}) \subseteq \mathcal{A}$.

**Definition 3.7.**

- An open ball in $\mathbb{R}^d$ is a set of the form $B(x_0, r) = \{x \in X : |x - x_0| < r\}$, where $x_0 \in \mathbb{R}^d$ and $0 < r < \infty$. Note that the open balls in $\mathbb{R}$ are the open intervals of the form $(a, b)$ with $-\infty < a < b < \infty$. To see this, note $(a, b) = (x_0 - r, x_0 + r) = B(x_0, r)$, where $x_0 = (b + a)/2$ and $r = (b - a)/2$.

- An open set in $\mathbb{R}^d$ is a union open balls.

**Definition 3.8.** Suppose $X$ is $\mathbb{R}$ or $\mathbb{R}^d$ (or any topological space). The Borel $\sigma$-algebra on $X$, denoted by $\mathcal{B}(X)$, is the $\sigma$-algebra generated by the collection of all open subsets of $X$. In other words, if $\mathcal{T}$ denotes the collection of all open subsets of $X$, then $\mathcal{B}(X) = \sigma(\mathcal{T})$. The elements of $\mathcal{B}(X)$ are called Borel sets in $X$.

**Theorem 3.9.** The Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^d)$ is generated by the collection of all open balls in $\mathbb{R}^d$. In other words, if $\mathcal{G}$ is the collection of all open balls in $\mathbb{R}^d$, then $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{G})$.

**Proof.** Let $\mathcal{T}$ be the collection of all open sets in $\mathbb{R}^d$. Since $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{T})$, we must show $\sigma(\mathcal{G}) = \sigma(\mathcal{T})$. Since $\mathcal{G} \subseteq \mathcal{T}$, we have $\sigma(\mathcal{G}) \subseteq \sigma(\mathcal{T})$ by Theorem 3.6. It remains to show $\sigma(\mathcal{T}) \subseteq \sigma(\mathcal{G})$. If $A \in \mathcal{T}$, then $A$ is a union of open balls in $\mathbb{R}^d$. Thus, for each point in $x \in A$, there is an open ball $B_x$ that contains $x$ and is contained in $A$. Now for each point $x \in A$ choose an open ball $B'_x$ with the following four properties:

1. $B'_x$ contains $x$,
2. $B'_x$ is contained in $B_x$ and (consequently) in $A$,
3. The center of $B'_x$ is a point with rational coordinates,
4. The radius of $B'_x$ is rational.

Then $A$ is the union of the collection of the open balls $B_{x'}$. But, since the open balls $B_{x'}$ all have rationals center and radii, there can be only countably many of them. Thus $A$ is the union of a countable collection of open balls. Hence $A \in \sigma(\mathcal{G})$. Thus $\mathcal{T} \subseteq \sigma(\mathcal{G})$. Therefore $\sigma(\mathcal{T}) = \sigma(\mathcal{G})$.

**Theorem 3.10.** The Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$ is generated by each of the following collections of sets:

(i) $\mathcal{G}_1 = \{(a, b) : a, b \in \mathbb{R}, a < b\}$

(ii) $\mathcal{G}_2 = \{(a, b] : a, b \in \mathbb{R}, a < b\}$

(iii) $\mathcal{G}_3 = \{(-\infty, b] : b \in \mathbb{R}\}$

(iv) $\mathcal{G}_4 = \{(-\infty, b) : b \in \mathbb{R}\}$

(v) $\mathcal{G}_5 = \{[a, b] : a, b \in \mathbb{R}, a < b\}$

(vi) $\mathcal{G}_6 = \{[a, b) : a, b \in \mathbb{R}, a < b\}$
(vii) $G_7 = \{[a, \infty) : a \in \mathbb{R}\}$

(viii) $G_8 = \{(a, \infty) : a \in \mathbb{R}\}$

**Proof.** (i): The previous theorem implies $G_1$ generates $B(\mathbb{R})$, i.e., $\sigma(G_1) = B(\mathbb{R})$.

(ii): By writing,

$$(a, b) = \bigcap_{n=1}^{\infty} (a, b + 1/n)$$

we see that $G_2 \subseteq \sigma(G_1)$, hence $\sigma(G_2) \subseteq \sigma(G_1)$. By writing

$$(a, b) = \bigcup_{n=1}^{\infty} (a, b - 1/n)$$

we see that $G_1 \subseteq \sigma(G_2)$, hence $\sigma(G_1) \subseteq \sigma(G_2)$. This proves $G_2$ generates $B(\mathbb{R})$.

(iii): By writing

$$(-\infty, b] = \bigcup_{n=1}^{\infty} (-n, b]$$

we see that $G_3 \subseteq \sigma(G_2)$, hence $\sigma(G_3) \subseteq \sigma(G_2)$. By writing

$$(a, b] = (-\infty, b] \cap (a, \infty) = (-\infty, b] \cap (-\infty, a)^c$$

we see that $G_2 \subseteq \sigma(G_3)$, hence $\sigma(G_2) \subseteq \sigma(G_3)$. This proves $G_3$ generates $B(\mathbb{R})$.

(iv): By writing

$$(-\infty, b] = \bigcup_{n=1}^{\infty} (-\infty, b - 1/n]$$

we see that $G_4 \subseteq \sigma(G_3)$, hence $\sigma(G_4) \subseteq \sigma(G_3)$. By writing

$$(-\infty, b] = \bigcap_{n=1}^{\infty} (-\infty, b + 1/n)$$

we see that $G_3 \subseteq \sigma(G_4)$, hence $\sigma(G_3) \subseteq \sigma(G_4)$. This proves $G_3$ generates $B(\mathbb{R})$.

(v),(vi),(vii),(viii): Similar. □

**Definition 3.11.** Let $B(\overline{\mathbb{R}})$ denote the collection of all sets of the form $A, A \cup -\infty, A \cup \infty, A \cup -\infty \cup \infty$, where $A$ is a Borel set in $\mathbb{R}$. It is straightforward to check that $B(\overline{\mathbb{R}})$ is a $\sigma$-algebra on $\overline{\mathbb{R}}$. We call $B(\overline{\mathbb{R}})$ the **Borel $\sigma$-algebra on $\overline{\mathbb{R}}$**.

**Remark 3.12.** It is possible to define open sets in $\overline{\mathbb{R}}$ and to define the Borel $\sigma$-algebra on $\overline{\mathbb{R}}$ as the $\sigma$-algebra generated by the open sets in $\overline{\mathbb{R}}$. See Exercise X???

**Theorem 3.13.** The Borel $\sigma$-algebra $B(\overline{\mathbb{R}})$ is generated by each of the following collection of sets:

(i) $G'_3 = \{(-\infty, b] : b \in \mathbb{R}\}$

(ii) $G'_4 = \{(-\infty, b) : b \in \mathbb{R}\}$
(iii) \( G'_7 = \{ [a, \infty) : a \in \mathbb{R} \} \)
(iv) \( G'_8 = \{ [a, \infty) : a \in \mathbb{R} \} \)

**Proof.** (i) By the previous theorem,
\[
G'_{11} = \{ (-\infty) \cup (-\infty, b) : b \in \mathbb{R} \} \subseteq \mathcal{B}(\mathbb{R}),
\]
hence \( \sigma(G'_{11}) \subseteq \mathcal{B}(\mathbb{R}) \). Now we show \( \mathcal{B}(\mathbb{R}) \subseteq \sigma(G'_{11}) \). Let \( B \in \mathcal{B}(\mathbb{R}) \). Then \( B \) equals one of \( A \) or \( A \cup -\infty \) or \( A \cup \infty \) or \( A \cup -\infty \cup \infty \), for some \( A \in \mathcal{B}(\mathbb{R}) \). To show that \( B \in \sigma(G'_{11}) \), it suffices to show that \( \{ -\infty \} \), \( \{ \infty \} \in \sigma(G'_{11}) \) and \( \mathcal{B}(\mathbb{R}) \subseteq \sigma(G'_{11}) \).

By writing,
\[
\{ -\infty \} = \bigcap_{n=1}^{\infty} [-n, -\infty)
\]
we see that \( \{ -\infty \} \in \sigma(G'_{11}) \). By writing
\[
\{ \infty \} = \bigcap_{n=1}^{\infty} (n, \infty) = \bigcap_{n=1}^{\infty} [-\infty, n]^c
\]
we see that \( \{ \infty \} \in \sigma(G'_{11}) \). Recall the definition of \( G_2 \) from the previous theorem:
\[
G_2 = \{ (a, b) : a, b \in \mathbb{R}, a < b \}
\]
By writing
\[
(a, b) = [-\infty, b] \cap (a, \infty) = [-\infty, b] \cap [-\infty, a]^c
\]
we see that \( G_2 \subseteq \sigma(G'_{11}) \). Therefore \( \sigma(G_2) \subseteq \sigma(G'_{11}) \). It is tempting to cite the previous theorem and assert that \( \mathcal{B}(\mathbb{R}) = \sigma(G_2) \). But this is not true. Here \( \sigma(G_2) \) is the smallest \( \sigma \)-algebra on \( \mathbb{R} \) that contains \( G_2 \). But (by the previous theorem) \( \mathcal{B}(\mathbb{R}) \) is the smallest \( \sigma \)-algebra on \( \mathbb{R} \) that contains \( G_2 \). We need to work a bit harder. We showed \( G_2 \subseteq \sigma(G'_{11}) \). Since every element of \( G_2 \) is a subset of \( \mathbb{R} \), we have
\[
G_2 \subseteq \{ \mathbb{R} \cap B : B \in \sigma(G'_{11}) \}.
\]
The reader can verify that \( \{ \mathbb{R} \cap B : B \in \sigma(G'_{11}) \} \) is a \( \sigma \)-algebra on \( \mathbb{R} \). Therefore, since \( \mathcal{B}(\mathbb{R}) \) is the smallest \( \sigma \)-algebra on \( \mathbb{R} \) that contains \( G_2 \), we have
\[
\mathcal{B}(\mathbb{R}) \subseteq \{ \mathbb{R} \cap B : B \in \sigma(G'_{11}) \}.
\]
For each \( B \in \sigma(G'_{11}) \),
\[
\mathbb{R} \cap B = \mathbb{R} \cap (-\infty)^c \cap \{ \infty \}^c \cap B \in \sigma(G'_{11}),
\]
and so
\[
\{ B \cap \mathbb{R} : B \in \sigma(G'_{11}) \} \subseteq \sigma(G'_{11}).
\]
Therefore
\[
\mathcal{B}(\mathbb{R}) \subseteq \sigma(G'_{11}).
\]
(ii),(iii),(iv): Similar.
4 Measurable Functions

To motivate the definition of a measurable function, we ask the reader to recall the following theorem from undergraduate analysis about continuous functions.

**Theorem 4.1.** Let \( f : \mathbb{R} \to \mathbb{R} \). Then \( f \) is continuous iff
\[
  f^{-1}(V) \text{ is an open set in } \mathbb{R} \text{ whenever } V \text{ is an open set in } \mathbb{R}
\]
If \( \mathcal{T} \) is the collection of open sets in \( \mathbb{R} \), we can rewrite this as: \( f \) is continuous iff
\[
  f^{-1}(V) \in \mathcal{T} \text{ whenever } V \in \mathcal{T}
\]

The reader who has studied topology may recall the following more general theorem.

**Theorem 4.2.** Let \( X \) and \( Y \) be topological spaces and \( f : X \to Y \). Then \( f \) is continuous iff
\[
  f^{-1}(V) \text{ is an open set in } X \text{ whenever } V \text{ is an open set in } Y
\]
If \( \mathcal{T}_X \) is the collection of open sets in \( X \) and \( \mathcal{T}_Y \) is the collection of open sets in \( Y \), we can rewrite this as: \( f \) is continuous iff
\[
  f^{-1}(V) \in \mathcal{T}_X \text{ whenever } V \in \mathcal{T}_Y
\]

In summary, a continuous function is a function whose inverse image preserves open sets.

Now we state the definition of a measurable function.

**Definition 4.3.** Let \( f : X \to Y \). Let \( \mathcal{A} \) be a \( \sigma \)-algebra on \( X \). Let \( \mathcal{B} \) be a \( \sigma \)-algebra on \( Y \). We say that \( f \) is measurable (or \( (\mathcal{A}, \mathcal{B}) \)-measurable to avoid ambiguity) if
\[
  f^{-1}(B) \in \mathcal{A} \text{ for every } B \in \mathcal{B}.
\]

Now we turn to some special cases. If \( f : X \to \mathbb{R} \), the \( \sigma \)-algebra on \( \mathbb{R} \) is always assumed to be the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \), i.e., \( (Y, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). We say \( f : X \to \mathbb{R} \) is measurable (or \( \mathcal{A} \)-measurable to avoid ambiguity) if
\[
  f^{-1}(B) \in \mathcal{A} \text{ for every } B \in \mathcal{B}(\mathbb{R}).
\]

If \( f : \mathbb{R}^d \to \mathbb{R} \), we sometimes (but not always) consider the Borel \( \sigma \)-algebra on the domain \( \mathbb{R}^d \). We say \( f : \mathbb{R}^d \to \mathbb{R} \) is Borel measurable (or \( \mathcal{B}(\mathbb{R}^d) \)-measurable) if
\[
  f^{-1}(B) \in \mathcal{B}(\mathbb{R}^d) \text{ for every } B \in \mathcal{B}(\mathbb{R}).
\]

The next theorem says that to show a function \( f \) is measurable we only need to check \( f^{-1}(B) \) for \( B \) in a generating collection.

**Theorem 4.4.** Let \( (X, \mathcal{A}) \) and \( (Y, \mathcal{B}) \) be measurable spaces. Let \( f : X \to Y \). Let \( \mathcal{G} \) be any collection of subsets of \( Y \) that generates \( \mathcal{B} \), i.e., \( \sigma(\mathcal{G}) = \mathcal{B} \). Then \( f \) is measurable iff
\[
  f^{-1}(B) \in \mathcal{A} \text{ for every } B \in \mathcal{G}.
\]
Proof. \(\Rightarrow\): Since \(G \subseteq \sigma(G) = B\), if 

\[ f^{-1}(B) \in A \quad \text{for every} \ B \in B. \]

then 

\[ f^{-1}(B) \in A \quad \text{for every} \ B \in G. \]

\(\Leftarrow\): It is readily verified that \(\{B \subseteq Y : f^{-1}(B) \in A\}\) is a \(\sigma\)-algebra. If 

\[ f^{-1}(B) \in A \quad \text{for every} \ B \in G \]

then \(\{B \subseteq Y : f^{-1}(B) \in A\}\) contains \(\sigma\) Thus \(B = \sigma(G) \subseteq \{B \subseteq Y : f^{-1}(B) \in A\}\). So \(f^{-1}(B) \in A\) for every \(B \in B\). Therefore \(f\) is measurable.

By combining the above theorem and Theorem 3.13 (which gives generating sets for \(B(\mathbb{R})\)), we obtain:

**Theorem 4.5.** Let \((X, A)\) be a measurable space and let \(f : X \to \mathbb{R}\). Then the following are equivalent:

(i) \(f\) is measurable

(ii) \(\{x \in X : f(x) < b\} = f^{-1}([−\infty, b)) \in A\) for every \(b \in \mathbb{R}\)

(iii) \(\{x \in X : f(x) \leq b\} = f^{-1}([−\infty, b]) \in A\) for every \(b \in \mathbb{R}\)

(iv) \(\{x \in X : f(x) > a\} = f^{-1}((a, \infty]) \in A\) for every \(a \in \mathbb{R}\)

(v) \(\{x \in X : f(x) \geq a\} = f^{-1}([a, \infty]) \in A\) for every \(a \in \mathbb{R}\)

**Remark 4.6.** When verifying a function \(f : X \to \mathbb{R}\) is measurable, we usually use the theorem above, rather than the definition.

**Theorem 4.7.** Let \((X, A)\) be a measurable space and let \(f, g : X \to \mathbb{R}\). If \(f\) is measurable, then \(\{x \in X : f(x) = c\} = f^{-1}\{c\} \in A\) for every \(c \in \mathbb{R}\).

**Proof.**

\[ \{x \in X : f(x) = c\} = \{x \in X : f(x) \leq c\} \cap \{x \in X : f(x) \geq c\} \in A. \]

**Notation.** For brevity, we write

\[ \{f < b\} = \{x \in X : f(x) < b\} f^{-1}([−\infty, b)), \]

\[ \{f = c\} = \{x \in X : f(x) = c\} = f^{-1}\{c\}, \]

and so on.

**Theorem 4.8.** Let \((X, A)\) be a measurable space. Let \(f, g : X \to \mathbb{R}\) be measurable functions. Then (i) \(\{f < g\}\), (ii) \(\{f \leq g\}\), and (iii) \(\{f = g\}\) are measurable sets.
Proof. (i): For each fixed \( x \in X \), we have \( f(x) < g(x) \) iff there is an \( r \in \mathbb{Q} \) such that \( f(x) < r < g(x) \). Thus

\[
\{ x \in X : f(x) < g(x) \} = \bigcup_{r \in \mathbb{Q}} \{ x \in X : f(x) < r \} \cap \{ x \in X : r < g(x) \} \in \mathcal{A}
\]

(ii): \( \{ f \leq g \} = \{ g < f \}^c \in \mathcal{A} \) by (i).

(iii): \( \{ f = g \} = \{ f \leq g \} \cap \{ f \geq g \} \in \mathcal{A} \) by (ii).

Now we give some examples of measurable functions.

**Theorem 4.9.** Let \((X, \mathcal{A})\) be a measurable space. Let \( f : X \to \mathbb{R} \). If \( f \) is constant, then \( f \) is measurable.

Proof. Assume \( f \) is constant. This means there exists a \( c \in \mathbb{R} \) such that \( f(x) = c \) for all \( x \in X \). Let \( a \in \mathbb{R} \). Then \( \{ x \in X : f(x) \geq a \} = \emptyset \) if \( a > c \), and \( \{ x \in X : f(x) \geq a \} = X \) if \( a \leq c \). Either way, \( \{ x \in X : f(x) < a \} \in \mathcal{A} \). Thus \( f \) is measurable.

**Theorem 4.10.** Let \((X, \mathcal{A})\) be a measurable space. Let \( A \subseteq X \). Then \( A \) is measurable (i.e., \( A \in \mathcal{A} \)) iff the indicator function \( 1_A \) is measurable.

Proof. \( \Rightarrow \): Assume \( A \in \mathcal{A} \). For every \( a \in \mathbb{R} \),

\[
\{ x \in X : 1_A(x) \geq a \} = \begin{cases} 
\emptyset & \text{if } a > 1 \\
A & \text{if } 0 \leq a \leq 1 \\
X & \text{if } a \leq 0 
\end{cases} \in \mathcal{A}.
\]

So \( 1_A \) is measurable.

\( \Leftarrow \): Assume \( 1_A \) is measurable. Then \( A = \{ x \in X : 1_A(x) > 0 \} \in \mathcal{A} \).

**Theorem 4.11.** Let \( f : \mathbb{R} \to \mathbb{R} \). If \( f \) is continuous, then \( f \) is Borel measurable.

Proof. Since \( f \) is continuous, \( f^{-1}(G) \) is open for every open set \( G \subseteq \mathbb{R} \). So \( \{ x \in X : f(x) < a \} = f^{-1}((\langle \infty, a \rangle)) \) is open for every \( a \in \mathbb{R} \). Therefore \( \{ x \in X : f(x) < a \} = f^{-1}((\langle -\infty, a \rangle)) \in \mathcal{B}(\mathbb{R}) \) for every \( a \in \mathbb{R} \).

**Theorem 4.12.** Let \( f : \mathbb{R} \to \mathbb{R} \). If \( f \) is increasing or decreasing, then \( f \) is Borel measurable.

Proof. Exercise.

**Theorem 4.13.** Let \((X, \mathcal{A}), (Y, \mathcal{B}), (Z, \mathcal{C})\) be a measurable spaces. Let \( f : X \to Y \) and \( g : Y \to Z \). If \( f \) is \((\mathcal{A}, \mathcal{B})\)-measurable and if \( g : Y \to Z \) is measurable \((\mathcal{B}, \mathcal{C})\)-measurable, then then the function \( g \circ f : X \to Z \) is \((\mathcal{A}, \mathcal{C})\) measurable.

Proof. Exercise.
Theorem 4.14. Let \((X, \mathcal{A})\) be a measurable space. Let \(f, g : X \to \mathbb{R}\) be measurable functions. Then the following functions are measurable:

(i) \(f + g\), if \(f(x) + g(x)\) is defined for every \(x \in X\).

(ii) \(fg\)

(iii) \(f/g\), if \(g(x) \neq 0\) for every \(x \in X\).

Proof. Let \(a \in \mathbb{R}\) be arbitrary.

(i):
\[
\{f + g < a\} = \{f < -g + a\} = \bigcup_{r \in \mathbb{Q}} (\{f < r\} \cap \{r < -g + a\}) = \bigcup_{r \in \mathbb{Q}} (\{f < r\} \cap \{g < -r + a\}) \in \mathcal{A}
\]

(ii): \(\{fg < a\} = A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4\), where \(A_i\) are as follows:

\[
A_0 = (\{f = 0\} \cup \{g = 0\}) \cap \{fg < a\}
\]
\[
A_1 = \{f > 0\} \cap \{g > 0\} \cap \{fg < a\}
\]
\[
A_2 = \{f > 0\} \cap \{g < 0\} \cap \{fg < a\}
\]
\[
A_3 = \{f < 0\} \cap \{g > 0\} \cap \{fg < a\}
\]
\[
A_4 = \{f < 0\} \cap \{g < 0\} \cap \{fg < a\}
\]

We have:
\[
A_0 = (\{f = 0\} \cup \{g = 0\}) \cap \{fg < a\} = \begin{cases} \emptyset & \text{if } a \leq 0 \\ \{f = 0\} \cup \{g = 0\} & \text{if } a \leq 0 \end{cases} \in \mathcal{A}
\]
\[
A_1 = \{f > 0\} \cap \{g > 0\} \cap \{fg < a\}
\]
\[
= \{f > 0, g > 0, fg < a\}
\]
\[
= \bigcup_{r \in \mathbb{Q}, r > 0} (\{f > 0, g > 0, f < r\} \cap \{f > 0, g > 0, r < a/g\})
\]
\[
= \bigcup_{r \in \mathbb{Q}, r > 0} (\{f > 0, g > 0, f < r\} \cap \{f > 0, g > 0, g < a/r\})
\]
\[
= \{f > 0\} \cap \{g > 0\} \bigcup_{r \in \mathbb{Q}, r > 0} (\{f < r\} \cap \{g < a/r\}) \in \mathcal{A}
\]
\[ A_2 = \{ f > 0 \} \cap \{ g < 0 \} \cap \{ fg < a \} = \{ f > 0, g < 0, fg < a \} = \{ f > 0, g < 0, f(-g) > -a \} = \{ f > 0, g < 0, f > (-a)/(-g) \} = \bigcup_{r \in \mathbb{Q}, r > 0} (\{ f > 0, g < 0, f > r \} \cap \{ f > 0, g < 0, r > (-a)/(-g) \}) = \bigcup_{r \in \mathbb{Q}, r > 0} (\{ f > 0, g < 0, f > r \} \cap \{ f > 0, g < 0, -g > (-a)/r \}) = \bigcup_{r \in \mathbb{Q}, r > 0} (\{ f > 0, g < 0, f > r \} \cap \{ f > 0, g < 0, g < a/r \}) = \{ f > 0 \} \cap \{ g < 0 \} \cup \bigcup_{r \in \mathbb{Q}, r > 0} (\{ f > r \} \cap \{ g < a/r \}) \in \mathcal{A} \]

Similarly, \( A_3, A_4 \in \mathcal{A} \). Therefore \( \{ fg < a \} \in \mathcal{A} \)

(iii): \( \{ f/g < a \} = \{ f < ag \} \in \mathcal{A} \) by (ii) and the previous theorem.

\[ \square \]

**Corollary 4.15.** Let \((X, \mathcal{A})\) be a measurable space. Let \( f, g : X \to \mathbb{R} \) be measurable functions and let \( c \in \mathbb{R} \). Then \( cf \) and \( f + g \) are measurable functions.

**Proof.** By Theorem 4.9 and Theorem 4.14(ii), \( cf \) is measurable. By Theorem 4.14(i), \( f + g \) is measurable. \( \square \)

**Theorem 4.16.** Let \((X, \mathcal{A})\) be a measurable space. Let \( f, g : X \to \mathbb{R} \) be measurable functions. The following functions are measurable:

(i) \( \max \{ f, g \} \)

(ii) \( \min f, g \)

(iii) \( |f| \)

**Proof.**

(i): \( \{ \max \{ f, g \} > a \} = \{ f > a \} \cup \{ g > a \} \in \mathcal{A} \)

(ii): \( \{ \min \{ f, g \} > a \} = \{ f > a \} \cap \{ g > a \} \in \mathcal{A} \)

(iii): \( \{|f| < a \} = \{-a < f < a \} = \{ f > -a \} \cap \{ f < a \} \in \mathcal{A} \). Alternatively, write \( |f| = \max(f, -f) \). \( \square \)

**Theorem 4.17.** Let \((X, \mathcal{A})\) be a measurable space. Let \( (f_n) \) be a sequence of measurable functions \( X \to \mathbb{R} \). The following functions are measurable:

(i) \( \sup_n f_n \)

(ii) \( \inf_n f_n \)
(iii) \( \limsup f_n = \inf_{k \geq 1} (\sup_{n \geq k} f_n) \)

(iv) \( \liminf f_n = \sup_{k \geq 1} (\inf_{n \geq k} f_n) \)

**Proof.** (i): For each \( a \in \mathbb{R} \),

\[
\left\{ \sup_n f_n > a \right\} = \left\{ x \in X : \sup_n f_n(x) > a \right\} = \bigcup_{n=1}^{\infty} \{ x \in X : f_n(x) > a \} \in \mathcal{A}.
\]

(ii): Similar to (i).

(iii): By (i), \( g_k = \sup_{n \geq k} f_n \) is measurable for each \( k \in \mathbb{N} \). By (ii), \( \limsup f_n = \inf_{k \geq 1} g_k \) is measurable.

(iv): Similar to (iii).

\( \square \)

**Corollary 4.18.** Let \((X, \mathcal{A})\) be a measurable space. Let \((f_n)\) be a sequence of measurable functions \(X \to [-\infty, \infty]\). If \( \lim_n f_n(x) \) converges to an extended real number for each \( x \in X \), then \( \lim_n f_n = \limsup_n f_n = \liminf_n f_n \), and so \( \lim_n f_n \) is measurable.
Chapter 3

Measures and Integrals

5 Measures

Definition 5.1. Let $\mathcal{A}$ be a $\sigma$-algebra on a set $X$. A measure on $\mathcal{A}$ is a function $\mu : \mathcal{A} \to [0, \infty]$ that has the following properties:

(a) $\mu(\emptyset) = 0$

(b) (Countable Additivity) If $E_1, E_2, \ldots$ is a sequence of disjoint sets in $\mathcal{A}$, then $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$.

Theorem 5.2 (Properties of Measures). Let $\mathcal{A}$ be a $\sigma$-algebra on a set $X$. If $\mu$ is a measure on $\mathcal{A}$, then $\mu$ has the following properties:

(i) (Finite Additivity) If $A, B$ are disjoint sets in $\mathcal{A}$, then $\mu(A \cup B) = \mu(A) + \mu(B)$.

(ii) (Subtraction) If $A, B \in \mathcal{A}$, $A \subseteq B$, and $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(iii) (Monotonicity) If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

(iv) (Countable Subadditivity) If $A_1, A_2, \ldots \in \mathcal{A}$, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

(v) (Continuity From Below) If $A_1, A_2, \ldots \in \mathcal{A}$ and $A_1 \subseteq A_2 \subseteq \ldots$, then $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n} \mu(A_n)$.

(vi) (Continuity From Above) If $A_1, A_2, \ldots \in \mathcal{A}$ and $A_1 \supseteq A_2 \supseteq \ldots$, and $\mu(A_1) < \infty$, then $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n} \mu(A_n)$.

Proof. (i): Define $A_1 = A$, $A_2 = B$, and $A_i = \emptyset$ for $i > 2$ and use countable additivity.

(ii): Since $B = (B \cap A) \cup (B \setminus A) = A \cup (B \setminus A)$, finite additivity gives

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

Since $\mu(A) < \infty$, we can subtract it from both sides to get the desired result.
(iii): Since \( B = (B \cap A) \cup (B \setminus A) = A \cup (B \setminus A) \), finite additivity and the fact that \( \mu(B \setminus A) \geq 0 \) gives

\[
\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)
\]

(iv): Define \( B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2) \), and so on. Then \( B_1, B_2, \ldots \) is sequence of disjoint sets in \( A \) such that \( \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \) and \( B_i \subseteq A_i \) for all \( i \in \mathbb{N} \). Therefore, by countable additivity and monotonicity,

\[
\mu\left( \bigcap_{i=1}^{\infty} A_i \right) = \mu\left( \bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i).
\]

(v): Define \( B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus A_2 \), and so on. Then \( B_1, B_2, \ldots \) is sequence of disjoint sets in \( A \) such that \( \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \) and \( \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n \) for all \( n \in \mathbb{N} \). Therefore, by countable additivity and finite additivity,

\[
\mu\left( \bigcap_{i=1}^{\infty} A_i \right) = \mu\left( \bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{\infty} \mu(B_i)
= \lim_{n \to \infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n \to \infty} \mu\left( \bigcup_{i=1}^{n} B_i \right)
= \lim_{n \to \infty} \mu\left( \bigcup_{i=1}^{n} A_i \right) = \lim_{n \to \infty} \mu\left( A_n \right).
\]

(vi): Define \( B_1 = A_1 \setminus A_1 = \emptyset, B_2 = A_1 \setminus A_2, B_3 = A_1 \setminus A_3 \), and so on. Then \( B_1, B_2, \ldots \in A \) and \( B_1 \subseteq B_2 \subseteq \ldots \). By (v),

\[
\mu\left( \bigcup_{i=1}^{\infty} B_i \right) = \lim_{n \to \infty} \mu(B_n).
\]

Since \( \bigcup_{i=1}^{\infty} B_i = A_1 \setminus \bigcap_{i=1}^{\infty} A_i \) and \( B_n = A_1 \setminus A_n \), we have

\[
\mu\left( A_1 \setminus \bigcap_{i=1}^{\infty} A_i \right) = \lim_{n \to \infty} \mu(A_1 \setminus A_n).
\]

By monotonicity, \( \mu\left( \bigcap_{i=1}^{\infty} A_i \right) \leq \mu(A_n) \leq \mu(A_1) < \infty \). Then (ii) gives

\[
\mu(A_1) - \mu\left( \bigcap_{i=1}^{\infty} A_i \right) = \lim_{n \to \infty} \left( \mu(A_1) - \mu(A_n) \right).
\]

Subtracting \( \mu(A_1) \) and then multiply by \(-1\) gives the desired result. \( \square \)

**Example 5.3.** Let \( X \) be any set.

(a) The **counting measure** on \( X \) is the measure \( \mu : \mathcal{P}(X) \to [0, \infty] \) defined by \( \mu(A) = \infty \) if \( A \) is an infinite subset of \( X \) and \( \mu(A) \) equals the number of elements in \( A \) if \( A \) is a finite subset of \( X \).

(b) Let \( x_0 \in X \). The **Dirac measure at** \( x_0 \) or **point mass at** \( x_0 \) is the measure \( \mu : \mathcal{P}(X) \to [0, \infty] \) defined by \( \mu(A) = 1 \) if \( x_0 \notin A \) and \( \mu(A) = 1 \) if \( x_0 \in A \).
(c) If $\mathcal{A}$ is any $\sigma$-algebra on $X$, the function $\mu$ defined by $\mu(A) = 0$ for every $A \in \mathcal{A}$ is a measure on $\mathcal{A}$. It is called the **zero measure** or **trivial measure** on $\mathcal{A}$.

(d) If $\mathcal{A}$ is any $\sigma$-algebra on $X$, the function $\mu$ defined by $\mu^*(\emptyset) = 0$ and $\mu(A) = \infty$ for every non-empty set $A \in \mathcal{A}$ is a measure on $\mathcal{A}$.

As an exercise, the reader should verify that these are indeed measures.

The most important measure is the Lebesgue measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Later, we will see how to define the Lebesgue measure. For now, we state without proof the following theorem that characterizes it.

**Theorem 5.4.** There is a unique measure $\lambda$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\lambda((a, b]) = b - a$$

for every $a, b \in \mathbb{R}, a \leq b$. It is called the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

We will also consider the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The following theorem characterizes it.

**Theorem 5.5.** There is a unique measure $\lambda_d$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$\lambda((a_1, b_1] \times \cdots \times (a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d)$$

for every $a_i, b_i \in \mathbb{R}, a_i \leq b_i$. It is called the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. 

6 Integral of Non-Negative Simple Functions

We define the integral in stages. We start by defining the integral for the class of simple functions and establishing its basic properties.

**Definition 6.1.** A function $s : X \to \mathbb{R}$ is called **simple** if its range $s(X) = \{s(x) : x \in X\}$ is a finite set.

**Lemma 6.2.** If $s : X \to \mathbb{R}$ is a simple function whose range consists of the distinct numbers $c_1, c_2, \ldots, c_n$, then

$$s = \sum_{i=1}^{n} c_i 1_{E_i} \quad \text{where } E_i = s^{-1}(\{c_i\}) = \{x \in X : s(x) = c_i\}. \quad (6.1)$$

We call (6.1) the **standard representation** of $s$. Note $E_1, \ldots, E_n$ are disjoint sets and $\bigcup_{i=1}^{n} E_i = X$. Note also that $s \geq 0$ iff $c_i \geq 0$ for all $1 \leq i \leq n$.

**Lemma 6.3.** Let $(X, \mathcal{A})$ be a measurable space. Let $s : X \to \mathbb{R}$ be a simple function whose range consists of the distinct numbers $c_1, c_2, \ldots, c_n$. Then $s$ is measurable iff each set $E_i = s^{-1}(\{c_i\})$ in its standard representation is measurable.

**Proof.** If each $E_i$ is measurable, then each $1_{E_i}$ is measurable, so $s = \sum_{i=1}^{n} c_i 1_{E_i}$ is measurable.

Conversely, if $s$ is measurable, then

$$E_i = \{x \in X : s(x) = c_i\} = \bigcap_{n=1}^{\infty} \{x \in X : c_i - 1/n < s(x) < c_i + 1/n\}$$

is measurable for each $i = 1, \ldots, n$. \qed

**Definition 6.4.** Let $(X, \mathcal{A}, \mu)$ be a measure space. If $s : X \to \mathbb{R}$ is a non-negative measurable simple function with standard representation

$$s = \sum_{i=1}^{n} c_i 1_{E_i},$$

then the **integral** of $s$ with respect to $\mu$ (or the $\mu$-integral of $s$) is

$$\int s d\mu = \sum_{i=1}^{n} c_i \mu(E_i)$$

For the sum, remember the convention $0 \cdot \infty = 0$. Note that $\int s d\mu$ may equal $\infty$. When there is no ambiguity, we write $\int s$ instead of $\int s d\mu$. Alternatively, it is sometimes convenient to write $\int s(x) d\mu(x)$ instead of $\int s d\mu$.

The next theorem says the integral is linear and monotone for non-negative measurable simple functions.
**Theorem 6.5.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(s\) and \(t\) be non-negative measurable simple functions.

(a) If \(c \in [0, \infty)\), then \(\int cs = c \int s\).

(b) \(\int (s + t) = \int s + \int t\)

(c) If \(s \leq t\), then \(\int s \leq \int t\).

**Proof.** Let \(s = \sum_{i=1}^{m} a_i 1_{A_i}\) and \(t = \sum_{j=1}^{n} b_j 1_{B_j}\) be the standard representations of \(s\) and \(t\). So \(A_i = \{x : s(x) = a_i\}\) and \(B_j = \{x : t(x) = b_i\}\). Note that \(A_i = \bigcup_{j=1}^{n} (A_i \cap B_j)\) for each fixed \(i = 1, \ldots, m\). Likewise \(B_j = \bigcup_{i=1}^{m} (A_i \cap B_j)\) for each fixed \(j = 1, \ldots, n\). Finally, note that \(A_i \cap B_j\) is disjoint from \(A_{i'} \cap B_{j'}\) whenever \((i, j) \neq (i', j')\).

(a): If \(c \in (0, \infty)\), then \(A_i = \{x : cs(x) = ca_i\}\), and so the standard representation of \(cs\) is

\[
\int cs = \sum_{i=1}^{m} ca_i 1_{A_i}.
\]

Therefore

\[
\int cs = \sum_{i=1}^{m} ca_i \mu(A_i) = c \sum_{i=1}^{m} a_i \mu(A_i) = c \int s.
\]

If \(c = 0\), then \(cs = 0\), and its standard representation is

\[
cs = 0 1_X.
\]

Thus

\[
\int cs = 0 \mu(X) = 0 = 0 \int s = c \int s.
\]

(b): Then, by the finite additivity of \(\mu\),

\[
\int s + \int t = \sum_{i=1}^{m} a_i \mu(A_i) + \sum_{j=1}^{n} b_j \mu(B_i)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i \mu(A_i \cap B_j) + \sum_{i=1}^{m} \sum_{j=1}^{n} b_j \mu(A_i \cap B_j)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{n} (a_i + b_j) \mu(A_i \cap B_j).
\]

Note \(s + t\) is a simple function because its range is the finite set

\[
\{a_i + b_j : 1 \leq i \leq m, 1 \leq j \leq n\}.
\]

Let \(c_1, \ldots, c_\ell\) be the distinct numbers in this set. For each \(k \in \{1, \ldots, \ell\}\), let

\[
E_k = (s + t)^{-1}(\{c_k\}) = \{x \in X : s(x) + t(x) = c_k\}
\]

Then

\[
E_k = \bigcup_{(i,j):a_i+b_j=c_k} (A_i \cap B_j)
\]

26
In other words, $E_k$ is the union of those sets $A_i \cap B_j$ such that $a_i + b_j = c_k$. So the standard representation of $s + t$ is

$$
s + t = \sum_{k=1}^{\ell} c_k 1_{E_k}.
$$

Then, by the finite additivity of $\mu$,

$$
\int (s + t) = \sum_{k=1}^{\ell} c_k \mu(E_k)
= \sum_{k=1}^{\ell} \sum_{i,j} c_k \mu(A_i \cap B_j)
= \sum_{i=1}^{m} \sum_{j=1}^{n} (a_i + b_j) \mu(A_i \cap B_j)
$$

Comparing the formulas above for $\int s + \int t$ and $\int (s + t)$ shows they are equal.

(c): If $s \leq t$, then for every $x \in A_i \cap B_j$ we have

$$
a_i = s(x) \leq t(x) \leq b_j
$$

Thus $a_i \leq b_j$ whenever $A_i \cap B_j$ is non-empty. Then, by the finite additivity of $\mu$,

$$
\int s = \sum_{i=1}^{m} a_i \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i \mu(A_i \cap B_j) \leq \sum_{i=1}^{m} \sum_{j=1}^{n} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{n} b_j \mu(B_j) = \int t
$$

□
7 Integral of Non-Negative Extended Real-Valued Functions

Now we extend the definition of the integral to the class of non-negative functions.

**Definition 7.1.** Let \((X, A, \mu)\) be a measure space. Let \(f : X \to [0, \infty]\) be a measurable function. The integral of \(f\) with respect to \(\mu\) (or the \(\mu\)-integral of \(f\)) is defined by

\[
\int f \, d\mu = \sup \left\{ \int s \, d\mu : s \text{ simple measurable, } 0 \leq s \leq f \right\}
\]  

(7.1)

Note that \(\int f \, d\mu\) may equal \(\infty\). When there is no ambiguity, we write \(\int f\) instead of \(\int f \, d\mu\).

Alternatively, it is sometimes convenient to write \(\int f(x) \, d\mu(x)\) instead of \(\int f \, d\mu\).

**Remark 7.2.** We must check that this definition of the integral agrees with the earlier one when \(f\) is simple. We temporarily use \(\oint\) to denote the integral defined in Definition 6.4. Then the formula for the integral in Definition 7.1 becomes

\[
\int f \, d\mu = \sup \left\{ \oint s \, d\mu : s \text{ simple measurable, } 0 \leq s \leq f \right\}
\]

If \(f : X \to [0, \infty]\) is a measurable simple function, then \(\oint f\) is an element of the set on the right. Moreover, for any simple measurable function \(s\) with \(0 \leq s \leq f\), Theorem 6.5 implies \(\oint s \leq \oint f\), and so \(\int f\) is the maximum of the set on the right. Since the supremum of a set equals the maximum whenever the maximum exists, we have

\[
\int f \, d\mu = \sup \left\{ \oint s \, d\mu : s \text{ simple measurable, } 0 \leq s \leq f \right\} = \oint f \, d\mu
\]

The next theorem establishes that the integral is monotone for non-negative functions.

**Theorem 7.3.** Let \((X, A, \mu)\) be a measure space. Let \(f, g : X \to [0, \infty]\) be measurable functions. If \(f \leq g\), then \(\int f \leq \int g\).

**Proof.** If \(s\) is a simple measurable function such that \(0 \leq s \leq f\), then \(0 \leq s \leq g\). Thus

\[
\left\{ \int s \, d\mu : s \text{ simple measurable, } 0 \leq s \leq f \right\} \subseteq \left\{ \int s \, d\mu : s \text{ simple measurable, } 0 \leq s \leq g \right\}
\]

Taking supremums gives the desired inequality. \(\square\)

The next theorem is the first of the fundamental convergence theorems in Lebesgue’s theory of integration.

**Theorem 7.4.** (Monotone Convergence Theorem) Let \((X, A, \mu)\) be a measure space. Let \(f_n : X \to [0, \infty]\) \((n = 1, 2, \ldots)\) be a sequence of measurable functions. If \(f_n\) increases to \(f\) pointwise (meaning that \(f_1(x) \leq f_2(x) \leq \ldots\) and \(f(x) = \lim_n f_n(x) = \sup_n f_n(x)\) for every \(x \in X\)), then \(f\) is measurable and

\[
\lim_n \int f_n = \int f.
\]
Proof. Since $f$ is the pointwise limit of a sequence of measurable functions, Corollary 4.18 implies $f$ is measurable. Since $f_1 \leq f_2 \leq \ldots \leq f$, we have $\int f_1 \leq \int f_2 \leq \ldots \leq \int f$, and so lim $\int f_n \leq \int f$. We must prove the reverse inequality. Let $0 < c < 1$ be given. Let $s$ be a simple measurable function such that $0 \leq s \leq f$. For every $n \in \mathbb{N}$, define

$$E_n = \{x \in X : cs(x) \leq f_n(x)\}.$$ 

For every $n \in \mathbb{N}$, we have $cs1_{E_n} \leq f_n$, and so

$$c \int s1_{E_n} \leq \int f_n. \quad (7.2)$$

We want to take the limit $n \to \infty$ in (7.2). Suppose the standard representation of $s$ is $s = \sum_{i=1}^{k} a_i 1_{A_i}$. Then $s1_{E_n}$ is a simple function and its standard representation is $s1_{E_n} = \sum_{i=1}^{k} a_i 1_{E_n \cap A_i}$. Therefore

$$\int s = \sum_{i=1}^{k} a_i \mu(A_i)$$

and

$$\int s1_{E_n} = \sum_{i=1}^{k} a_i \mu(E_n \cap A_i).$$

Note $E_1 \subseteq E_2 \subseteq \ldots$ and

$$X = \bigcup_{n=1}^{\infty} E_n.$$ (To see the last equality, consider an arbitrary $x \in X$. If $f(x) = 0$ or $s(x) = 0$, then $x \in E_1$. If $f(x) > 0$ and $s(x) > 0$, then $cs(x) < f(x)$, hence there exists $n \in \mathbb{N}$ such that $cs(x) < f_n(x) \leq f(x)$, and so $x \in E_n$. Let $i \in \{1, \ldots, k\}$ be arbitrary. We have $E_1 \cap A_i \subseteq E_2 \cap A_i \subseteq \ldots$ and

$$A_i = \bigcup_{n=1}^{\infty} (E_n \cap A_i).$$

By continuity from below,

$$\lim_{n \to \infty} \mu(E_n \cap A_i) = \mu(A_i).$$

Therefore

$$\lim_n \int s1_{E_n} = \lim_n \sum_{i=1}^{k} a_i \mu(E_n \cap A_i) = \sum_{i=1}^{k} a_i \mu(A_i) = \int s.$$ Thus taking $n \to \infty$ in (7.2) gives

$$c \int s \leq \lim_n \int f_n.$$ Since $0 < c < 1$ is arbitrary,

$$\int s \leq \lim_n \int f_n.$$ Since $s$ is an arbitrary simple measurable function with $0 \leq s \leq f$, we have

$$\int f \leq \lim_n \int f_n.$$
The next theorem concerns approximation of functions by simple functions. It is frequently used to produce a sequence of functions to which the monotone convergence theorem can be applied. Notice the theorem does not involve a measure \( \mu \).

**Theorem 7.5.** (Simple Approximation Theorem) Let \((X, \mathcal{A})\). Let \( f : X \to \mathbb{R} \). Then there exists a sequence \((s_n)\) of functions \( s_n : X \to \mathbb{R} \) such that

(i) \( s_1, s_2, \ldots \) are simple functions.

(ii) \(|s_1| \leq |s_2| \leq \ldots \leq |f|\).

(iii) \( \lim_{k \to \infty} s_k(x) = f(x) \) for each \( x \in X \).

(iv) If \( f \) is bounded, then \( s_k \to f \) uniformly on \( X \).

(v) If \( f \geq 0 \), then \( 0 \leq s_1 \leq s_2 \leq \ldots \leq f \).

(vi) If \((X, \mathcal{A})\) is a measurable space and \( f \) is measurable, then \( s_1, s_2, \ldots \) are measurable.

**Proof.** Let \( k \in \mathbb{N} \). Decompose \([-\infty, \infty]\) as the following union of subintervals:

\[-\infty, \infty] = [-\infty, -k] \cup (-k, 0] \cup [0, k) \cup [k, \infty]\]

Decompose \([-\infty, \infty]\) further by dividing \((-k, 0]\) and \([0, k)\) up into intervals of length \(1/2^k\). So we get

\[-\infty, \infty] = [-\infty, -k] \cup \bigcup_{m=1}^{k^2} \left(-\frac{m}{2^k}, -\frac{m-1}{2^k}\right] \cup \bigcup_{m=1}^{k^2} \left[\frac{m-1}{2^k}, \frac{m}{2^k}\right) \cup [k, \infty].\]

For each \( x \in X \), consider the subinterval to which \( f(x) \) belongs and define \( s_k(x) \) to the endpoint of that subinterval which is closest to 0. More precisely,

\[s_k(x) = \begin{cases} 
  k & \text{if } f(x) \in [k, \infty] \\
  \frac{m-1}{2^k} & \text{if } f(x) \in \left[\frac{m-1}{2^k}, \frac{m}{2^k}\right), m \in \{1, \ldots, k^2\} \\
  -\frac{m}{2^k} & \text{if } f(x) \in \left(-\frac{m}{2^k}, -\frac{m-1}{2^k}\right], m \in \{1, \ldots, k^2\} \\
  -k & \text{if } f(x) \in [-\infty, -k].
\end{cases}\]

Figure 3.1 illustrates the definition of \( s_k \) and \( s_{k+1} \).

(i): Since \( s_k \) takes only finitely many values, \( s_k \) is simple.

(ii): No matter which subinterval \( f(x) \) belongs, \( s_k(x) \) rounds \( f(x) \) to a number closer to zero than does \( s_{k+1}(x) \). See Figure 3.1.

(iii): The key observation is that if \( f(x) \in (-k, k) \), then

\[|f(x) - s_k(x)| \leq \frac{1}{2^k}.
\]

Here are the details. Let \( x \in X \). If \( f(x) = \infty \), then \( s_k(x) = k \) for every \( k \in \mathbb{N} \), and so \( \lim_{k \to \infty} s_k(x) = \infty = f(x) \). If \( f(x) = -\infty \), then \( s_k(x) = -k \) for every \( k \in \mathbb{N} \), and so \( \lim_{k \to \infty} s_k(x) = -\infty \).
Figure 3.1: Definition of $s_k$ and $s_{k+1}$

$-\infty = f(x)$. If $f(x) \in (-\infty, \infty)$, there exists $M > 0$ such that $f(x) \in (-M, M)$. Let $\epsilon > 0$. Choose $K \in \mathbb{N}$ such that $\frac{1}{2^K} < \epsilon$ and $K \geq M$. For every $k \in \mathbb{N}$, we have $f(x) \in (-k, k)$, and so

$$|f(x) - s_k(x)| \leq \frac{1}{2^k} \leq \frac{1}{2K} < \epsilon.$$

Therefore $\lim_{k \to \infty} s_k(x) = f(x)$.

(iv): If $f$ is bounded, then there exists $M > 0$ such that $f(x) \in (-M, M)$ for every $x \in X$. Let $\epsilon > 0$. Choose $K \in \mathbb{N}$ such that $\frac{1}{2^K} < \epsilon$ and $K \geq M$. For every $k \in \mathbb{N}$ and every $x \in X$, we have $f(x) \in (-k, k)$, and so

$$|f(x) - s_k(x)| \leq \frac{1}{2^k} \leq \frac{1}{2K} < \epsilon.$$

Therefore $s_k \to f$ uniformly on $X$.

(v): If $f(x) \geq 0$, then $s_k(x) \geq 0$ by definition. The rest follows from (ii).

(vi): For $k \in \mathbb{N}$ and $m \in \{1, \ldots, k2^k\}$, define

$$F_k = \{x : f(x) \in [k, \infty]\} = f^{-1}([k, \infty]),$$
$$F_{-k} = \{x : f(x) \in (-\infty, -k]\} = f^{-1}([\infty, -k]),$$
$$E_m = \{x : f(x) \in \left[\frac{m-1}{2^k}, \frac{m}{2^k}\right]\} = f^{-1}\left(\left[\frac{m-1}{2^k}, \frac{m}{2^k}\right]\right),$$
$$E_{-m} = \{x : f(x) \in \left(\frac{-m}{2^k}, \frac{-m-1}{2^k}\right]\} = f^{-1}\left(\left(\frac{-m}{2^k}, \frac{-m-1}{2^k}\right]\right).$$

Then

$$s_k = k1_{F_k} + (-k)1_{F_{-k}} + \sum_{m=1}^{k2^k} \frac{m-1}{2^k}1_{E_m} + \sum_{m=1}^{k2^k} \frac{-(m-1)}{2^k}1_{E_{-m}}.$$
If $f$ is measurable, then all the sets $F_k, F_{-k}, E_m, E_{-m}$ are measurable, and so $s_k$ is measurable.

Now we combine the monotone convergence theorem and the simple approximation theorem, to prove the linearity of the integral for non-negative functions.

**Theorem 7.6.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g : X \to [0, \infty]$ be measurable functions.

(a) If $c \in [0, \infty]$, then $\int cf = c \int f$.

(b) $\int (f + g) = \int f + \int g$

**Proof.** (a): Assume $c \in [0, \infty]$. Choose a sequence $(c_n)$ in $[0, \infty)$ such that $c_n$ increases to $c$. By Theorem 7.5, there is a sequence of non-negative measurable simple functions $(s_n)$ such that $s_n$ increases to $f$. Then $(c_n s_n)$ is a sequence of non-negative measurable simple functions that increases to $cf$. By the monotone convergence theorem and Theorem 6.5(a)

$$\int cf = \lim_n \int c_n s_n = \lim_n c_n \int s_n = \left( \lim_n c_n \right) \cdot \left( \lim_n \int s_n \right) = c \int f$$

(b): By Theorem 7.5, there are sequences of non-negative measurable simple functions $(s_n)$ and $(t_n)$ such that $s_n$ increases to $f$ and $t_n$ increases to $g$. Then $(s_n + t_n)$ is a sequence of non-negative measurable simple functions that increases to $f + g$. By the monotone convergence theorem and Theorem 6.5(b),

$$\int (f + g) = \lim_n \int (s_n + t_n) = \lim_n \int s_n + \lim_n \int t_n = \int f + \int g$$

**Remark 7.7.** In the exercises, you will be asked to prove the previous theorem without using the monotone convergence theorem or the simple approximation theorem.
8 Integral of Extended Real-Valued Functions

Finally, we extend the definition of the integral to $\mathbb{R}$-valued functions.

**Definition 8.1.** Let $f : X \to \mathbb{R}$. The **positive part** of $f$ is the function

$$f^+ = \max \{0, f\}.$$ 

The **negative part** of $f$ is the function

$$f^- = \max \{0, -f\}.$$ 

**Lemma 8.2.** Let $f : X \to \mathbb{R}$.

(a) $f^+ \geq 0$ and $f^- \geq 0$

(b) $f = f^+ - f^-$

(c) $|f| = f^+ + f^-$

(d) $f^+(x) = \frac{1}{2}(|f|(x) + f(x))$ whenever $f(x) \neq -\infty$

(e) $f^-(x) = \frac{1}{2}(|f|(x) - f(x))$ whenever $f(x) \neq \infty$

(f) For any measurable space $(X, \mathcal{A})$, $f$ is measurable iff $f^+$ and $f^-$ are measurable.

**Proof.** (a)-(e) are immediate from the definitions, and (f) follows from Theorem 4.14 and Theorem 4.16.

**Definition 8.3.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f : X \to \mathbb{R}$ be a measurable function. If both $\int f^+ d\mu$ and $\int f^- d\mu$ are finite, we say that $f$ is **integrable** (or integrable with respect to $\mu$ or $\mu$-integrable), and we define the **integral** of $f$ (or integral of $f$ with respect to $\mu$ or $\mu$-integral of $f$) to be

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

If exactly one of $\int f^+ d\mu$ and $\int f^- d\mu$ is finite, we still define the **integral** of $f$ with respect to $\mu$ (or $\mu$-integral of $f$) to be

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

but we do not say that $f$ is integrable.

If $\int f^+ d\mu = \int f^- d\mu = \infty$, the integral $\int f d\mu$ is not defined.

Note $f$ is integrable iff $\int f d\mu$ is finite.

**Theorem 8.4.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f : X \to \mathbb{R}$ be a measurable function.
(a) $f$ is integrable iff $|f|$ is integrable.

(b) If $\int f$ is defined, then $|\int f| \leq \int |f|$.

**Remark.** (b) is a triangle inequality for integrals.

**Proof.** (a): $f$ is integrable iff $\int f = \int f^+ - \int f^-$ is finite iff $\int f^+$ and $\int f^-$ are finite iff $\int |f| = \int (f^+ + f^-) = \int f^+ + \int f^-$ is finite iff $|f|$ is integrable.

(b):

$$\left| \int f \right| = \left| \int f^+ - \int f^- \right| \leq \left| \int f^+ \right| + \left| \int f^- \right| = \int f^+ + \int f^- = \int |f|$$

**Corollary 8.5.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{R}$ be measurable functions. If $|f| \leq |g|$ and $g$ is integrable, then $f$ is integrable.

**Proof.** Note $\int |f| \leq \int |g| < \infty$ and apply the previous theorem.

**Theorem 8.6.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{R}$ be integrable functions (which also means they are measurable).

(a) If $f \leq g$, then $\int f \leq \int g$.

(b) If $c \in \mathbb{R}$, then $\int cf = c \int f$.

(c) $\int (f + g) = \int f + \int g$

**Proof.** (a): Assume $f \leq g$. Then $f^+ \leq g^+$ and $f^- \geq g^-$. Therefore Theorem 7.3 gives

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \leq \int g^+ d\mu - \int g^- d\mu \leq \int g d\mu.$$ 

(b): Exercise.

(c): Note $(f + g)^+ \leq f^+ + g^+$. Then Theorem 7.3 and Theorem 7.6(b) imply

$$\int (f + g)^+ \leq \int (f^+ + g^+) = \int f^+ + \int g^+ < \infty.$$ 

Likewise $\int (f + g)^- < \infty$. Therefore $f + g$ is integrable. Note

$$(f + g)^+ - (f + g)^- = f + g = f^+ - f^- + g^+ - g^-.$$ 

Since $f^-, g^-$, and $(f + g)^-$ are finite, we can add them to both sides to get

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$
Then Theorem 7.6(b) gives
\[ \int (f + g)^+ + \int f^- + \int g^- = \int (f + g)^- + \int f^+ + \int g^+. \]

Since \( \int f^- , \int g^- , \) and \( \int (f + g)^- \) are finite, we can subtract them from both sides to get
\[ \int (f + g)^+ - \int (f + g)^- = \int f^+ - \int f^- + \int g^+ - \int g^- . \]

Therefore
\[ \int (f + g) = \int f + \int g . \]

We can partially extend the previous theorem to \( \mathbb{R} \)-valued functions.

**Theorem 8.7.** Let \((X, A, \mu)\) be a measure space. Let \( f, g : X \to \mathbb{R} \) be measurable functions. Assume \( \int f \) and \( \int g \) are defined.

(a) If \( f \leq g \), then \( \int f \leq \int g \).

(b) If \( c \in \mathbb{R} \), then \( \int cf = c \int f \).

(c) \( \int (f + g) = \int f + \int g \), provided that we do not have \( \int f = -\int g = \infty \) or \( \int f = -\int g = -\infty \) and provided that for all \( x \in X \) we do not have \( f(x) = -g(x) = \infty \) or \( f(x) = -g(x) = -\infty \).

**Proof.** Exercise. \( \square \)
9 Almost Everywhere

Definition 9.1. Let \((X, \mathcal{A}, \mu)\) be a measure space. We say a property \(P\) holds **almost everywhere** or for **almost every** \(x \in X\) if there exists a set \(A\) such that \(A \in \mathcal{A}, \mu(A^c) = 0\), and \(P\) holds for every \(x \in A\). Thus “almost everywhere” means “everywhere except on a set of measure zero.” For brevity, we often write “a.e.” instead of “almost everywhere” and “almost every.” The concept of “almost everywhere” depends on the measure \(\mu\). So, when clarity demands it, we write \(\mu\text{-a.e.}\) instead of a.e.

Definition 9.2. Let \((X, \mathcal{A}, \mu)\) be a measure space. We say a property \(P\) holds **almost everywhere** or for **almost every** \(x \in X\) if there exists a set \(A\) such that \(A \in \mathcal{A}, \mu(A^c) = 0\), and \(P\) holds for every \(x \in A\). Thus “almost everywhere” means “everywhere except on a set of measure zero.” For brevity, we often write “a.e.” instead of “almost everywhere” and “almost every.” The concept of “almost everywhere” depends on the measure \(\mu\). So, when clarity demands it, we write “\(\mu\text{-almost everywhere}\)” instead of “almost everywhere.”

Example 9.3. Let \((X, \mathcal{A}, \mu)\) be a measure space. If \(f, g: X \to \mathbb{R}\) and there exists \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \(f(x) = g(x)\) for every \(x \in A\), then we say that \(f = g\) a.e.

Example 9.4. Let \((X, \mathcal{A}, \mu)\) be a measure space. If \((f_n)\) is a sequence functions \(f_n: X \to \mathbb{R}\) and there exists \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \((f_n(x))\) converges for each \(x \in A\), then we say that \((f_n)\) converges a.e.

Observation. \(P\) a.e. if and only if there exists a set \(A\) such that \(A \in \mathcal{A}, \mu(A^c) = 0\), and \(A \subseteq \{x \in X : P \text{ holds at } x\}\) if and only if there exists a set \(N\) such that \(N \in \mathcal{A}, \mu(N) = 0\), and \(\{x \in X : P \text{ does not hold at } x\} \subseteq N\). For the second equivalence, take \(N = A^c\).

Observation. If the set \(\{x \in X : P \text{ holds at } x\}\) is in \(\mathcal{A}\), then:

(i) \(P\) a.e. is true iff \(\mu(\{x \in X : P \text{ does not hold at } x\}) = 0\).

(ii) \(P\) a.e. is false iff \(\mu(\{x \in X : P \text{ does not hold at } x\}) > 0\).

To see this take \(A = \{x \in X : P \text{ holds at } x\}\) in the definition.

Observation. If \(P_1\) and \(P_2\) are two properties, then \(P_1\) a.e. and \(P_2\) a.e. iff \(P_1\) and \(P_2\) a.e. If \(P_1, P_2, \ldots\) is a sequence of properties, then \(P_i\) holds a.e. for all \(i \in \mathbb{N}\) iff \(P_i\) for all \(i \in \mathbb{N}\) holds a.e. This is because a countable union of sets of measure 0 is a set of measure 0.

Remark. Almost everywhere convergence is not a topological convergence; there is no topology that can be placed on the set of measurable functions for which convergence in the topology is equivalent to convergence almost everywhere.

Theorem 9.5. Let \((X, \mathcal{A}, \mu)\) be a measure space. If \(f: X \to \mathbb{R}\) is an integrable function, then \(\mu(\{x \in X : |f(x)| = \infty\}) = 0\), hence \(f\) is finite a.e.

We give two proofs.
Proof 1. Let \( n \in \mathbb{N} \). Define \( E_n = \{ x \in X : n \leq |f(x)| \} \). Then \( \{ x \in X : |f(x)| = \infty \} \subseteq E_n \) and \( 1_{E_n} \leq n^{-1}|f| \). Therefore
\[
\mu ( \{ x \in X : |f(x)| = \infty \} ) \leq \mu ( E_n ) = \int 1_{E_n} \leq n^{-1} \int |f|.
\]
Since \( \int |f| \) is finite by assumption, letting \( n \to \infty \) gives \( \mu ( \{ x \in X : |f(x)| = \infty \} ) = 0 \).

Proof 2. Define \( A = \{ x \in X : |f(x)| < \infty \} \). Note \( A \in \mathcal{A} \) and \( |f| = |f|1_A + \infty 1_{A^c} \). Then
\[
\int |f| = \int |f|1_A + \int |f|1_{A^c} = \int |f|1_A + \infty \mu ( A^c ).
\]
If \( f \) is integrable, then \( \int |f| < \infty \), which forces \( \mu ( A^c ) = 0 \), hence \( f \) is finite a.e.

Theorem 9.6. Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \( f : X \to [0, \infty] \) be a measurable function. Then \( \int f = 0 \) iff \( f = 0 \) a.e.

We give two proofs.

Proof 1. Case 1: \( f \) is simple. Let \( f = \sum_{i=1}^{n} c_i 1_{E_i} \) be the standard representation of \( f \). So \( c_i \geq 0 \) and \( E_i = f^{-1}(\{c_i\}) \) for each \( 1 \leq i \leq n \). Then \( \int f = \sum_{i=1}^{n} c_i \mu (E_i) = 0 \) iff for every \( 1 \leq i \leq n \) either \( c_i = 0 \) or \( \mu (E_i) = 0 \) iff \( f = \sum_{i=1}^{n} c_i 1_{E_i} = 0 \) a.e.

Case 2: \( f \) not simple. Assume \( f = 0 \) a.e.. For every measurable simple function \( s \) with \( 0 \leq s \leq f \), we have \( s = 0 \) a.e., hence \( \int s = 0 \) by Case 1. Therefore
\[
\int f = \sup \left\{ \int s : s \text{ measurable simple, } 0 \leq s \leq f \right\} = \sup \{ 0 \} = 0.
\]
Conversely, assume \( \int f = 0 \). Note \( \{ x \in X : f(x) > 0 \} = \bigcup_{n=1}^{\infty} E_n \), where \( E_n = \{ x \in X : f(x) \geq n^{-1} \} \). For every \( n \in \mathbb{N} \), we have \( 1_{E_n} \leq nf \), and so
\[
\mu ( E_n ) = \int 1_{E_n} \leq n \int f = 0.
\]
Therefore
\[
\mu ( \{ x \in X : f(x) > 0 \} ) \leq \sum_{n=1}^{\infty} \mu ( E_n ) = 0
\]
Thus \( f = 0 \) a.e.

Proof 2. Assume \( f = 0 \) a.e. Then there exists \( A \in \mathcal{A} \) such that \( \mu ( A^c ) = 0 \) and \( f(x) = 0 \) for all \( x \in A \). Therefore
\[
f = f1_A + f1_{A^c} = f1_{A^c} \leq \infty 1_{A^c}.
\]
Thus
\[
\int f \leq \int \infty 1_{A^c} = \infty \mu ( A^c ) = 0.
\]
Conversely, assume \( \int f = 0 \). Note \( \{ x \in X : f(x) > 0 \} = \bigcup_{n=1}^{\infty} E_n \), where \( E_n = \{ x \in X : f(x) \geq n^{-1} \} \). For every \( n \in \mathbb{N} \), we have \( 1_{E_n} \leq nf \), and so

\[
\mu(E_n) = \int 1_{E_n} \leq n \int f = 0.
\]

Therefore

\[
\mu(\{ x \in X : f(x) > 0 \}) \leq \sum_{n=1}^{\infty} \mu(E_n) = 0
\]

Thus \( f = 0 \) a.e.

**Corollary 9.7.** Let \( (X, A, \mu) \) be a measure space. Let \( f, g : X \to \mathbb{R} \) be integrable functions. Then \( f = g \) a.e. iff \( \int |f - g| = 0 \).

**Proof.** \( f = g \) a.e. iff \( f - g = 0 \) a.e. iff \( |f - g| = 0 \) a.e. iff \( \int |f - g| = 0 \). The first equivalence uses that \( f \) and \( g \) are finite a.e., which follows from Theorem 9.5. The last equivalence comes from Theorem 9.6.

**Theorem 9.8.** Let \( (X, A, \mu) \) be a measure space. Let \( f, g : X \to \mathbb{R} \) be measurable functions.

Assume both \( \int f \) and \( \int g \) are defined.

(a) If \( f \leq g \) a.e., then \( \int f \leq \int g \).

(b) If \( f = g \) a.e., then \( \int f = \int g \).

**Proof.** (a): Case 1: \( f, g : X \to [0, \infty] \). Assume \( f \leq g \) a.e. Then there exists \( A \in A \) such that \( \mu(A^c) = 0 \) and \( f(x) \leq g(x) \) for all \( x \in A \). So \( f1_A = 0 \) a.e. and \( f1_A \leq g \) everywhere. Thus

\[
\int f = \int (f1_A + f1_{A^c}) = \int f1_A + \int f1_{A^c} = \int f1_A \leq \int g.
\]

Case 2: \( f, g : X \to \mathbb{R} \). Assume \( f \leq g \) a.e. Then \( f^+ \leq g^+ \) a.e and \( g^- \leq f^- \) a.e. Therefore, by Case 1, \( \int f^+ \leq \int g^+ \) and \( \int g^- \leq \int f^- \). Thus

\[
\int f = \int f^+ - \int f^- \leq \int g^+ - \int g^- = \int g.
\]

(b): Note that \( f = g \) a.e. iff \( f \leq g \) a.e. and \( g \leq f \) a.e. Then apply (a) twice.

**Corollary 9.9.** Let \( (X, A, \mu) \) be a measure space. Let \( f, g : X \to \mathbb{R} \) be measurable functions.

(a) If \( |f| \leq |g| \) a.e. and \( g \) is integrable, then \( f \) is integrable and \( \int |f| \leq \int |g| \).

(b) If \( f = g \) a.e. and \( g \) is integrable, then \( f \) is integrable and \( \int f = \int g \).

**Proof.** (a): By the previous theorem, we have \( \int |f| \leq \int |g| < \infty \). Thus \( f \) is integrable.

(b): If \( f = g \) a.e., then \( |f| = |g| \) a.e., and in particular \( |f| \leq |g| \) a.e. Then (a) implies \( f \) is integrable.

The previous theorem implies \( \int f = \int g \).
**Definition 9.10.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(f : X \to \mathbb{R}\). If \(f\) is measurable, \(E \in \mathcal{A}\), and \(\int f 1_E d\mu\) is defined, we define

\[
\int_E f d\mu = \int f 1_E d\mu.
\]

With this notation, \(\int f d\mu = \int_X f d\mu\).

**Theorem 9.11.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(f, g : X \to \mathbb{R}\) be measurable functions. Assume \(\int f\) and \(\int g\) are integrable. Then \(f = g\) a.e. iff \(\int_E f = \int_E g\) for every \(E \in \mathcal{A}\).

**Proof.** Exercise. \(\square\)
10 Fatou’s Lemma and Dominated Convergence Theorem

**Lemma 10.1** (Fatou’s Lemma). Let \((X, \mathcal{A}, \mu)\) be a measure space. If \(f_n : X \to [0, \infty) \ (n = 1, 2, \ldots)\) is a sequence of measurable functions, then

\[
\int \liminf f_n \leq \liminf \int f_n.
\]

*Proof.* Define \(g_k = \inf_{n \geq k} f_n\) for each \(k \in \mathbb{N}\). Then \((g_k)\) is an increasing sequence of measurable functions and \(\lim_k g_k = \liminf f_n\). By the monotone convergence theorem,

\[
\int \liminf f_n \leq \lim_k \int g_k.
\]

Note \(g_k \leq f_n\) whenever \(n \geq k\). So \(\int g_k \leq \int f_n\) whenever \(n \geq k\). Then, for every \(k \in \mathbb{N}\), \(\int g_k\) is a lower bound for the set \(\{\int f_n : n \geq k\}\). Thus \(\int g_k \leq \inf_{n \geq k} \int f_n\) for every \(k \in \mathbb{N}\). Therefore

\[
\int \liminf f_n = \lim_k \int g_k \leq \liminf \int_{n \geq k} f_n = \liminf \int f_n.
\]

\[\square\]

**Theorem 10.2.** (Dominated Convergence Theorem) Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \((f_n)\) be a sequence of measurable functions \(f_n : X \to \mathbb{R}\). Let \(f : X \to \mathbb{R}\) be a measurable function. Let \(g : X \to [0, \infty]\) be an integrable function. Suppose that the following properties hold a.e.: \(|f_n| \leq g\) for every \(n \in \mathbb{N}\) and \(f_n \to f\). Then \(f\) is integrable, \(f_n\) is integrable for each \(n\), and

\[
\lim_n \int f_n = \int f.
\]

*Proof.* We will need the following properties of lim inf that the reader should verify: If \(c \in \mathbb{R}\) and \((a_n)\) is a sequence in \(\mathbb{R}\), then

\[
\liminf(c + a_n) = c + \liminf a_n
\]

\[
\liminf(-a_n) = -\limsup a_n
\]

Let us first assume that the following properties hold everywhere:

\(g < \infty, \ |f_n| \leq g\) for every \(n \in \mathbb{N}\), and \(f_n \to f\).

Since \(f_n \to f\) pointwise and \(|f_n| \leq g\) for all \(n\), we have \(|f| \leq g\). Since \(g\) is integrable, Corollary 8.5 implies \(f\) is integrable and \(f_n\) is integrable for each \(n\).
Since $|f - f_n| \leq 2g$, we have $0 \leq 2g - |f - f_n|$. By Fatou’s lemma,

$$
\int 2g = \int \lim_n (2g - |f - f_n|)
= \int \liminf_n (2g - |f - f_n|)
\leq \liminf_n \int (2g - |f - f_n|)
= \liminf_n \left( \int 2g - \int |f - f_n| \right)
= \int 2g + \liminf_n \left( - \int |f - f_n| \right)
= \int 2g - \limsup_n \int |f - f_n|
$$

Since $\int 2g$ is finite, we can subtract it to obtain

$$
\limsup_n \int |f - f_n| \leq 0.
$$

Therefore we have

$$
0 \leq \liminf_n \int |f - f_n| \leq \limsup_n \int |f - f_n| \leq 0.
$$

Thus

$$
\lim_n \int |f - f_n| = \limsup_n \int |f - f_n| = \liminf_n \int |f - f_n| = 0.
$$

Furthermore,

$$
\lim_n \left| \int f_n - \int f \right| = \lim_n \left| \int (f_n - f) \right| \leq \lim_n \int |f - f_n| = 0,
$$

whence

$$
\lim_n \int f_n = \int f.
$$

Now let us only assume that the following properties hold almost everywhere:

$$
g < \infty, \ |f_n| \leq g \text{ for every } n \in \mathbb{N}, \text{ and } f_n \to f.
$$

(In fact, since $g$ is integrable, we already have $g < \infty$ a.e. because of Theorem 9.5, so we don’t need to assume it.) Therefore there exists a set $A \in \mathcal{A}$ with $\mu(A^c) = 0$ such that

$$
g(x) < \infty, \ |f_n(x)| \leq g(x) \text{ for every } n \in \mathbb{N}, \text{ and } f_n(x) \to f(x)
$$

for every $x \in A$. Then the following hold everywhere:

$$
g1_A < \infty, \ |f_n 1_A| \leq g1_A \text{ for every } n \in \mathbb{N}, \text{ and } f_n 1_A \to f1_A.
$$

The functions $g1_A, f1_A, f_n 1_A$ are measurable because $g, f, f_n, 1_A$ are all measurable. Since $g$ is integrable and $g1_A = g$ a.e., Corollary 9.9 implies $g1_A$ is integrable. Therefore, by the first part of the proof, $f1_A$ is integrable, $f_n 1_A$ is integrable for each $n$, and

$$
\lim_n \int f_n 1_A = \int f1_A
$$
Then since \( f1_A = f \) a.e. and \( f_n1_A = f_n \) a.e. for each \( n \), Corollary 9.9 implies \( f \) is integrable, \( f_n \) is integrable for each \( n \), and

\[
\lim_n \int f_n = \lim_n \int f_n1_A = \int f1_A = \int f.
\]

The next theorem is an application of the dominated convergence theorem. It concerns interchanging limits and derivatives with integrals.

**Theorem 10.3.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \( g : X \to [0, \infty] \) be an integrable function. Let \( f : X \times [a, b] \to \mathbb{R} \) such that \( f(\cdot, t) : X \to \mathbb{R} \) is integrable for each \( t \in [a, b] \). Define

\[
F(t) = \int f(x, t)d\mu(x)
\]

for each \( t \in [a, b] \). Let \( c \in [a, b] \).

(a) If \( \lim_{t \to c} f(x, t) = f(x, c) \) for all \( x \in X \) and \( |f(x, t)| \leq g(x) \) for all \( x \in X, t \in [a, b] \), then

\[
\lim_{t \to c} F(t) = F(c),
\]

i.e.,

\[
\lim_{t \to c} \int_X f(x, t)d\mu(x) = \int \lim_{t \to c} f(x, t)d\mu(x).
\]

(b) If \( \frac{\partial f}{\partial t}(x, t) \) exists for all \( x \in X, t \in [a, b] \) and \( |\frac{\partial f}{\partial t}(x, t)| \leq g(x) \) for all \( x \in X, t \in [a, b] \), then

\[
F'(c) = \int \frac{\partial f}{\partial t}(x, c)d\mu(x).
\]

i.e.,

\[
\left( \frac{d}{dt} \int f(x, t)d\mu(x) \right)(c) = \int \frac{\partial f}{\partial t}(x, c)d\mu(x).
\]

**Proof.** We need the following fact about limits that the reader should verify: \( \lim_{x \to c} h(x) = L \) iff \( \lim_{n \to \infty} h(x_n) = L \) for every sequence \((x_n)\) with \( x_n \neq c \) and \( x_n \to c \). This fact will allow us to apply the dominated convergence theorem.

(a): Exercise.

(b): Let \((t_n)\) be any sequence of numbers in \([a, b]\) such that \( t_n \neq c \) for all \( n \) and \( t_n \to c \). Define

\[
f_n(x) = \frac{f(x, t_n) - f(x, c)}{t_n - c}
\]

for each \( x \in X \). Then \( f_n \) is measurable and \( \lim_n f_n(x) = \frac{\partial f}{\partial t}(x, c) \). By the mean value theorem,

\[
|f_n(x)| \leq \sup_{t \in [a, b]} |\frac{\partial f}{\partial t}(x, t)| \leq g(x)
\]

42
for all $x \in X$. By the dominated convergence theorem,

$$\lim_{n} \frac{F(t_n) - F(c)}{t_n - c} = \lim_{n} \frac{1}{t_n - c} \left( \int f(x, t_n) d\mu(x) - \int f(x, c) d\mu(x) \right)$$

$$= \lim_{n} \int \frac{f(x, t_n) - f(x, c)}{t_n - c} d\mu(x)$$

$$= \lim_{n} \int f_n(x) d\mu(x)$$

$$= \int \frac{\partial f}{\partial t}(x, c) d\mu(x).$$

Since $(t_n)$ is arbitrary, the limit

$$F'(c) = \lim_{t \to c} \frac{F(t) - F(c)}{t - c}$$

exists, and

$$F'(c) = \int \frac{\partial f}{\partial t}(x, c) d\mu(x).$$

\[\square\]

**Exercise.** State and prove a stronger version of this theorem where the hypotheses are required to hold only almost everywhere.
Chapter 4

Spaces of Functions

11 Normed Spaces and Banach Spaces

We assume the reader is familiar with the definition of a vector space. Here are some examples.

Example 11.1. (i) $\mathbb{R}^d$ is a vector space with scalar field $\mathbb{R}$.

(ii) $\mathbb{C}^d$ is a vector space with scalar field $\mathbb{C}$.

(iii) If $X$ is a set and $V$ is a vector space with scalar field $K$, then the set $V^X$ of all functions from $X$ to $V$ is a vector space over $K$. Addition and scalar multiplication are defined pointwise: If $f, g : X \to V$ and $c \in K$, then $f + g$ and $cf$ are defined as usual by $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ for all $x \in X$.

(iv) Let $(X, \mathcal{A})$ be a measure space. The set $M(X)$ of measurable functions from $X$ to $\mathbb{R}$ is a vector space with scalar field $\mathbb{R}$. Indeed, Corollary 4.15 implies this set is closed under addition and scalar multiplication, so it is a subspace of the vector space of all functions from $X$ to $\mathbb{R}$.

Definition 11.2. A norm on $V$ is a function $\| \cdot \| : V \to [0, \infty)$ such that the following conditions are satisfied for all $x, y \in V$ and $c \in K$:

(i) $\|x\| = 0$ if and only if $x = 0$ (Definiteness)

(ii) $\|cx\| = |c|\|x\|$ (Homogeneity)

(iii) $\|x + y\| \leq \|x\| + \|y\|$ (Triangle Inequality)

The pair $(V, \| \cdot \|)$ is called a normed space. When the norm is clear from context, we write $V$ instead of $(V, \| \cdot \|)$.

Example 11.3. (i) For each $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, define

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_d|^2} = \left( \sum_{i=1}^d |x_i|^2 \right)^{1/2}.$$  

This is the Euclidean norm or 2-norm on $\mathbb{R}^d$.  

44
(ii) Let \(1 \leq p < \infty\). For each \(x = (x_1, \ldots, x_d) \in \mathbb{R}^d\), define
\[
\|x\|_p = \left( \sum_{i=1}^{d} |x_i|^p \right)^{1/p}.
\]
This is the \(p\)-norm on \(\mathbb{R}^d\).

(iii) For each \(x = (x_1, \ldots, x_d) \in \mathbb{R}^d\), define
\[
\|x\|_{\infty} = \max_{1 \leq i \leq n} |x_i|.
\]
This is called the max-norm or \(\infty\)-norm on \(\mathbb{R}^d\).

In later sections, we will see examples of norms on vector spaces of functions.

For the rest of this section, assume \(V\) is a normed space with scalar field \(K\), where \(K\) is either \(\mathbb{R}\) or \(\mathbb{C}\).

**Definition 11.4.** Let \((V, \| \cdot \|)\) be a normed space. A sequence \((x_n)\) in \(V\) is called **convergent** if there exists a point \(x \in V\) with the following property: For every \(\epsilon > 0\) there exists a positive integer \(N\) such that for every integer \(n \geq N\) we have \(\|x_n - x\| < \epsilon\). In such case, \(x\) is called the **limit** of \((x_n)\), we say that \((x_n)\) **converges** to \(x\), and we write \(\lim_n x_n = x\) or \(x_n \to x\). A sequence which is not convergent is called **divergent**.

**Theorem 11.5.** Let \((V, \| \cdot \|)\) be a normed space. Let \(x \in V\) and let \((x_n)\) be a sequence in \(V\). Then \(x_n \to x\) iff \(\|x_n - x\| \to 0\).

**Theorem 11.6.** Let \((V, \| \cdot \|)\) be a normed space. If \((x_n)\) is a convergent sequence in \(V\), then \((x_n)\) has a unique limit. In other words, if \((x_n)\) is a convergent sequence in \(V\) such that \(x_n \to x \in V\) and \(x_n \to y \in V\), then \(x = y\). (This justifies saying “the limit” rather than “a limit” in the definition.)

**Theorem 11.7.** Let \((V, \| \cdot \|)\) be a normed space. If \((x_n)\) and \((y_n)\) are convergent sequences in \(V\) and \(c \in K\), then

(a) \(\lim_n (x_n + y_n) = \lim_n x_n + \lim y_n\).

(b) \(\lim_n cx_n = c \lim x_n\).

(c) \(\lim_n \|x_n\| = \|\lim_n x_n\|\)

**Definition 11.8.** Let \((V, \| \cdot \|)\) be a normed space. A sequence \((x_n)\) in \(V\) is called **Cauchy** if for every \(\epsilon > 0\), there exists a positive integer \(N\) such that for all integers \(m, n \geq N\) we have \(\|x_m - x_n\| < \epsilon\).

**Theorem 11.9.** In a normed space \(V\), every convergent sequence is Cauchy.

**Definition 11.10.** Let \((V, \| \cdot \|)\) be a normed space. A set \(E \subseteq V\) is called **complete** if every Cauchy sequence in \(E\) converges and its limit is in \(E\). If \(V\) is complete, \(V\) is called a **Banach space**.

**Example 11.11.** \(\mathbb{C}^n\) and \(\mathbb{R}^n\) are complete, but \(\mathbb{Q}^n\) is not.
12 \( L^p \): Definitions and Basic Properties

**Definition 12.1.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(1 \leq p < \infty\). For each measurable function \(f : X \to \mathbb{R}\), define the \(L^p\)-norm (or \(p\)-norm) of \(f\) to be
\[
\|f\|_p = \left( \int_X |f(x)|^p \, d\mu(x) \right)^{1/p}
\]
with the convention that \((\infty)^{1/p} = \infty\) for \(1 \leq p < \infty\).

Define the \(L^p\) space to be
\[
L^p = \{f : X \to \mathbb{R}, \ f \text{ measurable, } \|f\|_p < \infty\}
\]
with the convention that functions in \(L^p\) are viewed as equal iff they are equal a.e. That is, \(f = g\) in \(L^p\) iff \(f = g\) a.e. (To be precise, \(L^p\) is the set of equivalence classes of measurable functions \(f : X \to \mathbb{R}\) such that \(\|f\|_p < \infty\), where the equivalence relation is “equal almost everywhere.”)

When necessary for clarity, we may use the following more descriptive notations:
\[
L^p = L^p(\mu) = L^p(X) = L^p(X, \mathcal{A}, \mu) = L^p(X, \mathcal{A}, \mu, \mathbb{R})
\]
\[
\| \cdot \|_p = \| \cdot \|_{L^p(\mu)} = \| \cdot \|_{L^p(X, \mathcal{A}, \mu)} = \| \cdot \|_{L^p(X, \mathcal{A}, \mathbb{R})}
\]

**Theorem 12.2.** Let \((X, \mathcal{A}, \mu)\) be a measure space. For each \(1 \leq p < \infty\), \(L^p\) is a vector space and \(\| \cdot \|_p\) is a norm on \(L^p\).

The rest of this section is devoted to the proof of Theorem 12.2. We will show that \(L^p\) is closed under addition and scalar multiplication (which implies that \(L^p\) is a vector subspace of the space of all measurable functions), and that \(\| \cdot \|_p\) satisfies the three properties in the definition of a norm (Definition 11.2): definiteness, homogeneity, and the triangle inequality.

**Remark.** We always assume \(L^p\) is equipped with the norm \(\| \cdot \|_p\). Whenever we talk about convergence in \(L^p\), we are talking about convergence with respect to this norm.

In the following theorem, we prove the definiteness and homogeneity of the \(L^p\) norm, and we show that \(L^p\) is closed under scalar multiplication.

**Theorem 12.3.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(1 \leq p < \infty\). Let \(f : X \to \mathbb{R}\) be a measurable function.

(a) \(\|f\|_p = 0\) iff \(f = 0\) a.e. iff \(f = 0\) in \(L^p\)

(b) If \(c \in \mathbb{R}\), then \(\|cf\|_p = |c|\|f\|_p\).

(c) If \(c \in \mathbb{R}\) and \(f \in L^p\), then \(cf \in L^p\).

**Proof.** (a): By Theorem, \(\|f\|_p = 0\) iff \(\int |f|^p = 0\) iff \(|f| = 0\) a.e. iff \(f = 0\) a.e. The second equivalence is the convention from the definition of \(L^p\).
(b): \[ \| cf \|_p^p = \int |cf|^p = |c|^p \int |f|^p = |c|^p \| f \|_p^p. \]

(c): Immediate from (b).

To prove the triangle inequality for the $L^p$ norm and that $L^p$ is closed under addition, we first establish:

**Theorem 12.4.** (Holder’s Inequality). Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{R}$ be measurable functions. Let $1 < p < \infty$ and let $1 < q < \infty$ be such that $p^{-1} + q^{-1} = 1$ (equivalently, $q = p/(p-1)$). We call $q$ the conjugate exponent of $p$. Then

\[ \| fg \|_1 \leq \| f \|_p \| g \|_q \]

The proof of Holder’s inequality will use the following lemma.

**Lemma 12.5.** (Young’s Inequality) If $a \geq 0$, $b \geq 0$, $p > 0$, $q > 0$, and $p^{-1} + q^{-1} = 1$, then

\[ ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \]

**Proof.** Fix $b, p, q$ and consider the function $h(a) = p^{-1}a^p + q^{-1}b^q - ab$ for $a \in [0, \infty)$. The desired inequality is equivalent to $h(a) \geq 0$. So it will suffice to show that the minimum value of $h(a)$ is 0. Note $h'(a) = a^{p-1} - b$. So $h'(a) > 0$ if $a > b^{1/(p-1)}$ and $h'(a) < 0$ if $a < b^{1/(p-1)}$. Thus the minimum value of $h(a)$ occurs at $a = b^{1/(p-1)}$ and is equal to

\[ h(b^{1/(p-1)}) = p^{-1}(b^{1/(p-1)})^p + q^{-1}b^q - (b^{1/(p-1)})b = (p^{-1} + q^{-1})b^q - b^q = 0. \]

For the second equality we used that $p/(p-1) = 1 + 1/(p-1) = q$, which is equivalent to $p^{-1} + q^{-1} = 1$. \hfill \Box

**Proof of Theorem 12.4 (Holder’s Inequality).** We consider three cases.

Case 1: $\| f \|_p \| g \|_q = \infty$. The desired inequality is clear.

Case 2: $\| f \|_p \| g \|_q = 0$. Then $\| f \|_p = 0$ or $\| g \|_q = 0$. So $f = 0$ a.e. or $g = 0$ a.e. Either way, we have $fg = 0$ a.e. Hence $\| fg \|_1 = 0$ and the desired inequality holds.

Case 3: $0 < \| f \|_p \| g \|_q < \infty$. Set $a = |f(x)|/\| f \|_p$ and $b = |g(x)|/\| g \|_q$, and apply the lemma to get

\[ \frac{|f(x)g(x)|}{\| f \|_p \| g \|_q} \leq \frac{|f(x)|^p}{p \| f \|_p^p} + \frac{|g(x)|^q}{q \| g \|_q^q}. \]

Integrating gives

\[ \frac{\| fg \|_1}{\| f \|_p \| g \|_q} \leq \frac{1}{p} + \frac{1}{q} = 1. \]

\hfill \Box
The next theorem is the triangle inequality for $L^p$ norms when $1 \leq p < \infty$. It also establishes that $L^p$ is closed under addition.

**Theorem 12.6.** (Minkowski’s Inequality). Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{R}$ be measurable functions. Let $1 \leq p < \infty$. Then

$$
\|f + g\|_p \leq \|f\|_p + \|g\|_p
$$

Consequently, if $f, g \in L^p$, then $f + g \in L^p$.

**Proof.** If $p = 1$, then, since $|f + g| \leq |f| + |g|$, we have

$$
\|f + g\|_1 = \int |f + g| \leq \int |f| + \int |g| = \|f\|_1 + \|g\|_1.
$$

Assume $1 < p < \infty$. We consider three cases.

Case 1: $\|f + g\|_p = 0$. The desired inequality is clear.

Case 2: $\|f + g\|_p = \infty$. Note

$$
|f + g|^p \leq (|f| + |g|)^p \leq (2 \max\{|f|, |g|\})^p = 2^p \max\{|f|^p, |g|^p\} \leq 2^p(|f|^p + |g|^p).
$$

Integrating gives

$$
\|f + g\|^p_p \leq 2^p(\|f\|^p_p + \|g\|^p_p).
$$

Therefore, if $\|f + g\|_p = \infty$, we must have $\|f\|_p = \infty$ or $\|g\|_p = \infty$, so either way the desired inequality holds.

Case 3: $0 < \|f + g\|_p < \infty$. Note

$$
|f + g|^p = |f + g||f + g|^{p-1} \leq |f||f + g|^{p-1} + |g||f + g|^{p-1}
$$

Integrating gives

$$
\|f + g\|^p_p \leq \|f\|_p \|f + g\|^{p-1}_1 + \|g\|_p \|f + g\|^{p-1}_1.
$$

Applying Holder’s inequality separately to the two terms on the right implies

$$
\|f + g\|^p_p \leq \|f\|_p \|(f + g)^{p-1}\|_q + \|g\|_p \|(f + g)^{p-1}\|_q
$$

where $p^{-1} + q^{-1} = 1$. But

$$
\|(f + g)^{p-1}\|_q = \left(\int |f + g|^{(p-1)q}\right)^{1/q} = \left(\int |f + g|^p\right)^{(p-1)/p} = \|f + g\|^{p-1}_p.
$$

Therefore the inequality above is

$$
\|f + g\|^p_p \leq (\|f\|_p + \|g\|_p)\|f + g\|^{p-1}_p.
$$

Dividing by $\|f + g\|^{p-1}_p$ gives the desired result.
13 \( L^p \): Completeness

**Theorem 13.1.** Let \((X, A, \mu)\) be a measure space. For every \(1 \leq p < \infty\), \(L^p\) is complete.

**Proof.** Let \((f_n)\) be a Cauchy sequence in \(L^p\). Inductively choose positive integers \(n_1 < n_2 < \ldots\) such that
\[
\|f_{n_{j+1}} - f_{n_j}\|_p < 2^{-j}
\]  
(13.1)
for all \(j\). Then \((f_{n_k})_{k=1}^\infty\) is a subsequence of \((f_n)\). For every \(x \in X\) and \(k \in \mathbb{N}\),
\[
f_{n_k}(x) = f_{n_1}(x) + \sum_{j=1}^{k-1} (f_{n_{j+1}}(x) - f_{n_j}(x)),
\]  
(13.2)
so the terms of the sequence \((f_{n_k}(x))\) are the partial sums of the series
\[
f_{n_1}(x) + \sum_{j=1}^\infty (f_{n_{j+1}}(x) - f_{n_j}(x)).
\]  
(13.3)

We will show that this series converges a.e. This will show that the sequence \((f_{n_k})\) converges a.e. Let us consider the series of absolute values. Define
\[
g_k = |f_{n_1}| + \sum_{j=1}^{k-1} |f_{n_{j+1}} - f_{n_j}|,
\]  
\[
g = |f_{n_1}| + \sum_{j=1}^\infty |f_{n_{j+1}} - f_{n_j}|.
\]

By (13.1), we have
\[
\|g_k\|_p \leq \|f_{n_1}\|_p + \sum_{j=1}^{k} \|f_{n_{j+1}} - f_{n_j}\|_p \leq \|f_{n_1}\|_p + \sum_{j=1}^{k} 2^{-j} < \|f_{n_1}\|_p + 1 < \infty.
\]

Since \(g_k\) increases to \(g\), it follows that \(g_k^p\) increases to \(g^p\), and so the monotone convergence theorem implies
\[
\int g^p = \lim_k \int g_k^p = \lim_k \|g_k\|_p^p \leq (\|f_{n_1}\|_p + 1)^p < \infty.
\]
So \(g^p\) is integrable. By Theorem 9.5, \(g^p\) is finite a.e. Therefore \(g\) is finite a.e. Thus the series
\[
|f_{n_1}(x)| + \sum_{j=1}^\infty |f_{n_{j+1}}(x) - f_{n_j}(x)|
\]
converges in \(\mathbb{R}\) for a.e. \(x \in X\). This means that the series (13.3) converges absolutely in \(\mathbb{R}\) for a.e. \(x \in X\). Since every absolutely convergent series of real numbers is convergent, the series (13.3) converges in \(\mathbb{R}\) for a.e. \(x \in X\). Therefore, since \((f_{n_k}(x))\) is the sequence of partial sums of the
series (13.3), the sequence \((f_{n_k}(x))\) converges in \(\mathbb{R}\) for a.e. \(x \in X\). So there exists \(A \in \mathcal{A}\) such that 
\(\mu(A^c) = 0\) and \((f_{n_k}(x))\) converges in \(\mathbb{R}\) for every \(x \in A\). Define \(f : X \to \mathbb{R}\) by

\[
  f(x) = \begin{cases} 
    \lim_k f_{n_k}(x) & \text{if } x \in A \\
    0 & \text{if } x \notin A 
  \end{cases}
\]

Then \(f = \lim_k f_{n_k}\) a.e. Moreover, \(f\) is measurable because \(f = \lim_k f_{n_k} \mathbf{1}_A\) and each function \(f_{n_k} \mathbf{1}_A\) is measurable.

Now we show that \(f_{n_k} \to f\) in \(L^p\). Recall that \(g^p\) is integrable. By (13.2), we have

\[
  |f_{n_k}| \leq g_k \leq g
\]

and so

\[
  |f| = \lim_k |f_{n_k}| \leq \lim_k g_k = g \quad \text{a.e.}
\]

Thus

\[
  \|f_{n_k}\|_p^p = \int |f_{n_k}|^p \leq \int g^p < \infty
\]

\[
  \|f\|_p^p = \int |f|^p \leq \int g^p < \infty
\]

So \(f_{n_k}, f \in L^p\). Moreover,

\[
  |f_{n_k} - f|^p \leq (|f_{n_k}| + |f|)^p \leq (2g)^p \quad \text{a.e.}
\]

Since \(g^p\) is integrable, \((2g)^p\) is integrable. Since \(f_{n_k} \to f\) a.e., we have

\[
  |f_{n_k} - f|^p \to 0 \quad \text{a.e.}
\]

Now the dominated convergence theorem implies

\[
  \lim_k \|f_{n_k} - f\|_p^p = \lim_k \int |f_{n_k} - f|^p = 0.
\]

Thus \(f_{n_k} \to f\) in \(L^p\).

Finally, we show that \(f_n \to f\) in \(L^p\). Let \(\epsilon > 0\) be given. We must show that there exists \(N \in \mathbb{N}\) such that for all \(n \geq N\)

\[
  \|f_n - f\|_p < \epsilon.
\]

Since \(f_{n_k} \to f\) in \(L^p\), there exists \(K \in \mathbb{N}\) such that for all \(k \geq K\)

\[
  \|f_{n_k} - f\|_p < \epsilon/2.
\]

Since \((f_n)\) is Cauchy in \(L^p\), there exists \(M \in \mathbb{N}\) such that for all \(m, n \geq M\)

\[
  \|f_m - f_n\|_p < \epsilon/2.
\]

Set \(N = \max \{M, K\}\). Let \(n \geq N\) be arbitrary. Choose \(k \geq N\). Then \(k \geq N \geq K\) and \(n_k \geq k \geq N \geq M\) and \(n \geq N \geq M\). In summary, \(k \geq K\) and \(n_k, n \geq M\). So

\[
  \|f_n - f\|_p \leq \|f_{n_k} - f\|_p + \|f_{n_k} - f_n\|_p < \epsilon/2 + \epsilon/2 = \epsilon.
\]

\(\Box\)
14  \( B(X) \) and \( C(X) \)

**Definition 14.1.** Let \( X \) be any set. For every function \( f : X \to \mathbb{R} \), define the **uniform norm** (or **supremum norm**) of \( f \) to be

\[
\|f\|_u = \sup\{|f(x)| : x \in X\}
\]

Since \( \|f\|_u \) is the least upper bound for the set \( \{|f(x)| : x \in X\} \), we have the alternative formulas

\[
\|f\|_u = \min\{M \in [0, \infty] : |f(x)| \leq M \text{ for every } x \in X\}
\]

\[
= \inf\{M \in [0, \infty) : |f(x)| \leq M \text{ for every } x \in X\}
\]

Given a function \( f : X \to \mathbb{R} \), a number \( M \in [0, \infty] \) is called an **upper bound** for \( |f| \) if \( |f(x)| \leq M \) for all \( x \in X \). We say \( f \) is **bounded** if \( |f| \) has a finite upper bound, i.e., if there is a number \( M \in [0, \infty) \) such that \( |f(x)| \leq M \) for all \( x \in X \). Note that \( f \) is bounded iff \( \|f\|_u < \infty \).

The set of all bounded functions \( f : X \to \mathbb{R} \) is denoted by \( B(X) \) (or \( B(X, \mathbb{R}) \)). In other words,

\[
B(X) = \{f : f : X \to \mathbb{R}, \|f\|_u < \infty\}
\]

**Theorem 14.2.** Let \( X \) be a set. Let \( f, g : X \to \mathbb{R} \).

(a) \( f = 0 \) iff \( \|f\|_u = 0 \)

(b) If \( c \in \mathbb{R} \), then \( \|cf\|_u = |c|\|f\|_u \)

(c) \( \|f + g\|_u \leq \|f\|_u + \|g\|_u \)

(d) \( B(X) \) is a vector space and \( \|\cdot\|_u \) is norm on \( B(X) \).

**Proof.**

(a) If \( f = 0 \), then \( \|f\|_u = \sup\{|f(x)| : x \in X\} = \sup\{0\} = 0 \). Conversely, if \( \|f\|_u = 0 \), then \( |f| \leq \|f\|_u = 0 \), so \( f = 0 \).

(b) We have

\[
\|cf\|_u = \sup\{|cf(x)| : x \in X\} = \sup\{|c||f(x)| : x \in X\} = |c|\sup\{|f(x)| : x \in X\} = |c|\|f\|_u
\]

The third inequality is the following general property of suprema: If \( A \subseteq \mathbb{R} \) and \( c \in [0, \infty] \), then \( \sup cA = c\sup A \), where \( cA = \{ca : a \in A\} \).

(c) Since \( |f + g| \leq |f| + |g| \leq \|f\|_u + \|g\|_u \), we have \( \|f + g\|_u \leq \|f\|_u + \|g\|_u \).

(d) Parts (b) and (c) imply that \( B(X) \) is closed scalar and multiplication. So \( B(X) \) is a vector subspace of the vector space of all functions \( f : X \to \mathbb{R} \). Parts (a), (b), and (c) imply \( \|\cdot\|_u \) is a norm on \( B(X) \). 

\( \square \)
**Remark.** We always assume $B(X)$ is equipped with the uniform norm. Whenever we talk about convergence in $B(X)$, we are talking about convergence with respect to this norm.

The next theorem says that convergence with respect to the uniform norm is equivalent to uniform convergence. This explain why the uniform norm is called the *uniform* norm.

**Theorem 14.3.** Let $X$ be a set. Let $(f_n)$ be a sequence of functions $f_n : X \to \mathbb{R}$. Let $f : X \to \mathbb{R}$. Then $\|f_n - f\|_u \to 0$ iff $f_n \to f$ uniformly on $X$.

**Proof.** The statement $\|f_n - f\|_u \to 0$ means that for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for every $n \geq N$ we have

$$\sup \{|f_n(x) - f(x)| : x \in X\} \leq \epsilon. \quad (14.1)$$

On the other hand, the statement $f_n \to f$ uniformly on $X$ means that for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| \leq \epsilon \quad \text{for all } x \in X. \quad (14.2)$$

Since (14.1) holds iff (14.2) holds, the two statements in question are equivalent. \qed

**Theorem 14.4.** Let $X$ be a set. Then $B(X)$ is complete.

**Proof.** Let $(f_n)$ be a Cauchy sequence in $B(X)$. For every $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that

$$\|f_m - f_n\|_u < \epsilon \quad \text{for all } m, n \geq N_\epsilon. \quad (14.3)$$

Since $|f_m - f_n| \leq \|f_m - f_n\|_u$, it follows that

$$|f_m(x) - f_n(x)| < \epsilon \quad \text{for all } x \in X \text{ and } m, n \geq N_\epsilon. \quad (14.4)$$

Thus $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$ for each fixed $x \in X$. By the completeness of $\mathbb{R}$, $(f_n(x))$ converges in $\mathbb{R}$ for each fixed $x \in X$. Define $f : X \to \mathbb{R}$ by $f(x) = \lim_n f_n(x)$ for each $x \in X$.

Now we show $f_n \to f$ in $B(X)$. Let $\epsilon > 0$. Choose $N_\epsilon$ so that (14.4) holds. Let $x \in X$ and $n \geq N_\epsilon$ be arbitrary. Since $f(x) = \lim_m f(x)$, there exists an integer $m(x) \geq N_\epsilon$ such that

$$|f_{m(x)}(x) - f(x)| < \epsilon.$$

Combining this inequality and (14.4) gives

$$|f_n(x) - f(x)| \leq |f_n(x) - f_{m(x)}(x)| + |f_{m(x)}(x) - f(x)| < \epsilon + \epsilon = 2\epsilon.$$

This shows that

$$|f_n(x) - f(x)| < 2\epsilon \quad \text{for all } x \in X \text{ and } n \geq N_\epsilon.$$

It follows that

$$\|f_n - f\|_u \leq 2\epsilon \quad \text{for all } n \geq N_\epsilon.$$

Thus $\|f_n - f\|_{\infty} \to 0$. Moreover, since $f_n \in B(X)$ for all $n$, we have

$$\|f\|_u \leq \|f - f_n\|_u + \|f_n\|_u < 2\epsilon + \infty = \infty$$

for all $n \geq N_\epsilon$. So $f \in B(X)$. Therefore $f_n \to f$ in $B(X)$.

\qed
**Definition 14.5.** Let $X$ be a subset of $\mathbb{R}^d$. Define $C(X)$ to be the set of continuous functions $f : X \to \mathbb{R}$.

**Theorem 14.6.** Let $X$ be a compact subset of $\mathbb{R}^d$ (for example, $X = [a, b]^d$). Then $C(X)$ is a vector subspace of $B(X)$, $\| \cdot \|_u$ is a norm on $C(X)$, and $C(X)$ is complete with respect to this norm. Moreover, for $f \in C(X)$,

$$
\|f\|_u = \max \{ |f(x)| : x \in X \}
$$

**Proof.** As any continuous function on a compact set is bounded, $C(X) \subseteq B(X)$. As the sum of two continuous functions is continuous and a constant multiple of a continuous functions is continuous, $C(X)$ is a vector subspace of $B(X)$. As $\| \cdot \|_u$ satisfies the properties of a norm on for the elements $B(X)$, it also satisfies those properties for the elements of the subset $C(X) \subseteq B(X)$. Thus $\| \cdot \|_u$ is a norm on $C(X)$. To see that, $\|f\|_u = \max \{ |f(x)| : x \in X \}$ for every $f \in C(X)$ we argue as follows. From undergraduate analysis, we know that each continuous function on compact set achieves their maximum on the set. In particular, $|f|$ achieves its maximum on $X$. Thus the supremum of $|f|$ on $X$ is actually the maximum of $f$ on $X$; i.e.,

$$
\|f\|_u = \sup \{ |f(x)| : x \in X \} = \max \{ |f(x)| : x \in X \}.
$$

To show that $C(X)$ is complete, we first show:

**Claim.** For every sequence $(f_n)$ of functions in $C(X)$, if $(f_n)$ converges to a function $f \in B(X)$, then $f \in C(X)$.

(In the language of topology, the claims says that $C(X)$ is a closed subset of $B(X)$). The proof of the claim is an an example of an $\epsilon/3$-proof.

Proof of Claim. Let $(f_n)$ be a sequence of functions in $C(X)$ such that $(f_n)$ converges to a function $f \in B(X)$. We need to show that $f$ is continuous on $X$. Let $x_0 \in X$ be given. Let $\epsilon > 0$ be given. Then for every $\epsilon > 0$ there exists an $N_\epsilon \in \mathbb{N}$ such that

$$
\|f_n - f\|_u \leq \epsilon/3 \quad \text{for all } n \geq N,
$$

which means

$$
|f_n(x) - f(x)| \leq \epsilon/3 \quad \text{for all } x \in X \text{ and } n \geq N. \tag{14.5}
$$

Since $f_N$ is continuous, it is continuous at $x_0$. So there exists a $\delta_{x_0} > 0$ such that

$$
|f_N(x) - f_N(x_0)| \leq \epsilon/3 \quad \text{for all } x \in X \text{ with } |x - x_0| < \delta_{x_0}. \tag{14.6}
$$

Now, for every $x \in X$ with $|x - x_0| < \delta_{x_0}$, we can apply (14.6) and apply (14.6) (at $x = x$ and $x = x_0$ with $n = N$) to get

$$
|f(x) - f(x_0)| = |f(x) - f_N(x) + f_N(x) - f_N(x_0) + f_N(x_0) - f(x_0)| \tag{14.7}
\leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \tag{14.8}
\leq \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \tag{14.9}
$$
This shows \( f \) is continuous at \( x_0 \). Since \( x_0 \) is arbitrary point in \( X \), this shows \( f \in C(X) \). (Notice that the boundedness of \( f \) was not used. Only the fact that \( f_n \to f \) in the norm \( \| \cdot \|_u \) was needed.)

Now we are ready to show that \( C(X) \) is complete. Let \( (f_n) \) be a Cauchy sequence in \( C(X) \). Since \( C(X) \) is a subset of \( B(X) \), \( (f_n) \) is a Cauchy sequence in \( B(X) \). Since \( B(X) \) is complete, \( (f_n) \) converges in uniform norm to some \( f \in B(X) \). Now the Claim implies that \( f \in C(X) \). So \( (f_n) \) converges in uniform norm to some \( f \in C(X) \). This proves \( C(X) \) is complete.

(The argument above can be generalized to show that closed subsets of complete subsets of normed spaces are complete).
**Definition 15.1.** Let \((X, \mathcal{A}, \mu)\) be a measure space. For each measurable function \(f : X \to \mathbb{R}\), define the \(L^\infty\) norm (or \(\infty\)-norm) of \(f\) to be

\[
\|f\|_\infty = \inf \{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.
\]

Define

\[
L^\infty = \{f : f : X \to \mathbb{R}, f \text{ measurable, } \|f\|_\infty < \infty\}
\]

with the convention that two functions in \(L^\infty\) are equal as elements in \(L^\infty\) iff they are equal a.e. That is, \(f = g\) in \(L^\infty\) iff \(f = g\) a.e. (To be precise, \(L^\infty\) is the set of equivalence classes of measurable functions \(f : X \to \mathbb{R}\) such that \(\|f\|_\infty < \infty\), where the equivalence relation is “equal almost everywhere.”)

If necessary for clarity, we may use the following more descriptive notations:

\[
L^\infty = L^\infty(\mu) = L^\infty(X) = L^\infty(X, \mathcal{A}, \mu) = L^\infty(X, \mathcal{A}, \mu, \mathbb{R})
\]

\[
\| \cdot \|_\infty = \| \cdot \|_{L^\infty(\mu)} = \| \cdot \|_{L^\infty(X)} = \| \cdot \|_{L^\infty(X, \mathcal{A}, \mu)} = \| \cdot \|_{L^\infty(X, \mathcal{A}, \mu, \mathbb{R})}
\]

Given a measurable function \(f : X \to \mathbb{R}\), a number \(M \in [0, \infty]\) is called an **essential upper bound** for \(|f|\) if \(|f(x)| \leq M\) for a.e. \(x \in X\). We say \(f\) is **essentially bounded** if \(|f|\) has a finite essential upper bound, i.e., if there is a number \(M \in [0, \infty)\) such that \(|f(x)| \leq M\) for a.e. \(x \in X\). Thus \(f\) is essentially bounded iff \(\|f\|_\infty < \infty\), and \(L^\infty\) is the set of all essentially bounded measurable functions from \(X\) to \(\mathbb{R}\) (with functions that are equal a.e. identified).

**Remark.** The \(L^\infty\) norm is closely related to the uniform norm. Indeed, the uniform norm of \(f\) is the least upper bound \(|f|\) and the \(L^\infty\) norm is the least essential upper bound of \(|f|\).

**Lemma 15.2.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \(f : X \to \mathbb{R}\) be a measurable function. Then \(|f| \leq \|f\|_\infty\) a.e. and

\[
\|f\|_\infty = \min \{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.
\]

**Proof.** First we show that \(|f| \leq \|f\|_\infty\). Since the set \(\{x : |f(x)| \leq M\}\) belongs to \(\mathcal{A}\), \(|f| \leq M\) a.e. holds iff \(\mu(\{x : |f(x)| > M\}) = 0\). (To see this, take \(A = \{x : |f(x)| \leq M\}\) in the definition of \(|f| \leq M\) a.e.). Thus

\[
\|f\|_\infty = \inf \{M \in [0, \infty) : \mu(\{x : |f(x)| > M\}) = 0\}
\]

If \(\|f\|_\infty = \infty\), inequality is clear. Assume \(\|f\|_\infty < \infty\). For every \(n \in \mathbb{N}\), since \(\|f\|_\infty < \|f\|_\infty + 1/n\), the definition of the infimum implies there is a number \(M_n\) such that \(M_n < \|f\|_\infty + 1/n\) and \(\mu(\{x : |f(x)| > M_n\}) = 0\). Then

\[
\{x : |f(x)| > \|f\|_\infty\} = \bigcup_{n=1}^{\infty} \{x : |f(x)| > \|f\|_\infty + 1/n\} \subseteq \bigcup_{n=1}^{\infty} \{x : |f(x)| > M_n\}
\]

55
and so

$$\mu\left( \{ x : |f(x)| > \|f\|_\infty \} \right) \leq \sum_{n=1}^{\infty} \mu\left( \{ x : |f(x)| > M_n \} \right) = 0.$$  

Thus $|f| \leq \|f\|_\infty$ a.e.

By definition, $\|f\|_\infty$ is a lower bound for

$$\{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.$$  

So $\|f\|_\infty$ is also a lower bound for

$$\{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.$$  

Since $|f| \leq \|f\|_\infty$ a.e.,

$$\|f\|_\infty \in \{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.$$  

Therefore

$$\|f\|_\infty = \min \{M \in [0, \infty) : |f(x)| \leq M \text{ for almost every } x \in X\}.$$  

Theorem 15.3. Let $(X, A, \mu)$ be a measure space. Let $f, g : X \to \mathbb{R}$ be measurable functions.

(a) $\|f\|_\infty = 0$ iff $f = 0$ a.e. iff $f = 0$ in $L^\infty$

(b) If $c \in \mathbb{R}$, then $\|cf\|_\infty = |c|\|f\|_\infty$

(c) $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$ (Minkowski Inequality)

(d) $\|fg\|_1 \leq \|f\|_\infty \|g\|_1$ (Holder Inequality)

(e) $L^\infty$ is a vector space and $\| \cdot \|_\infty$ is norm on $L^\infty$.

Proof.

(a) The second equivalence is the convention from the definition of $L^\infty$. Now we prove the first equivalence. If $f = 0$ a.e., then $|f| = 0$ a.e., so $\|f\|_\infty = 0$ by definition. Conversely, if $\|f\|_\infty = 0$, then $|f| \leq \|f\|_\infty = 0$ a.e., so $f = 0$ a.e.

(b) If $c = 0$, then $\|cf\|_\infty = \|0f\|_\infty = 0 = \|0\|_\infty = c\|f\|_\infty$ If $c \neq 0$, then

$$\|cf\|_\infty = \inf \\{M \in [0, \infty) : |cf(x)| \leq M \text{ for almost every } x \in X\}$$

$$= \inf \{M \in [0, \infty) : |f(x)| \leq M/|c| \text{ for almost every } x \in X\}$$

$$= \inf \{|c|M' : M' \in [0, \infty), \ |f(x)| \leq M' \text{ for almost every } x \in X\}$$

$$= |c| \inf \{M' : M' \in [0, \infty), \ |f(x)| \leq M' \text{ for almost every } x \in X\}$$

$$= |c|\|f\|_\infty$$

The third equality holds because the sets involved are equal. The fourth inequality is the following general property of infima: If $A \subseteq \mathbb{R}^+$ and $c \in [0, \infty]$, then $\inf cA = c\inf A$, where $cA = \{ca : a \in A\}$.
(c) Since \(|f + g| \leq |f| + |g| \leq \|f\|_{\infty} + \|g\|_{\infty}\) a.e., we have \(\|f + g\|_{\infty} \leq \|f\|_{\infty} + \|g\|_{\infty}\).

(d) Since \(|f| \leq \|f\|_{\infty}\) a.e., we have \(|fg| \leq \|f\|_{\infty}|g|\) a.e., hence \(\int |fg| \leq \|f\|_{\infty} \int |g|\).

(e) Parts (b) and (c) imply that \(L^\infty\) is closed scalar and multiplication. So \(L^\infty\) is a vector subspace of the space of all measurable functions. Parts (a), (b), and (c) imply \(\|\cdot\|_{\infty}\) is a norm on \(L^\infty\).

\[\square\]

**Remark.** We always assume \(L^\infty\) is equipped with the norm \(\| \cdot \|_p\). Whenever we talk about convergence in \(L^\infty\), we are talking about convergence with respect to this norm.

The next theorem says that convergence in the \(L^\infty\) norm is equivalent to uniform convergence almost everywhere.

**Theorem 15.4.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Let \((f_n)\) be a sequence of measurable functions \(f_n : X \to \mathbb{R}\). Let \(f : X \to \mathbb{R}\) be a measurable function. Then \(\|f_n - f\|_{\infty} \to 0\) iff \(f_n \to f\) uniformly a.e. on \(X\). In other words, \(\|f_n - f\|_{\infty} \to 0\) iff there exists \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \(f_n \to f\) uniformly on \(A\).

**Proof.** \(\Rightarrow\): Assume \(\|f_n - f\|_{\infty} \to 0\). For each \(n \in \mathbb{N}\), we have \(|f_n - f| \leq \|f_n - f\|_{\infty}\) a.e., hence there exists \(A_n \in \mathcal{A}\) such that \(\mu(A_n^c) = 0\) and

\[|f_n(x) - f(x)| \leq \|f_n - f\|_{\infty}\quad \text{for all } x \in A_n.\]

Define \(A = \cap_{n=1}^{\infty} A_n\). Then \(\mu(A^c) = 0\) and

\[|f_n(x) - f(x)| \leq \|f_n - f\|_{\infty}\quad \text{for all } x \in A\text{ and all } n \in \mathbb{N}.\]

(15.1)

Now we will use the assumption that \(\|f_n - f\|_{\infty} \to 0\). Let \(\epsilon > 0\). There exists an \(N_\epsilon \in \mathbb{N}\) such that

\[\|f_n - f\|_{\infty} < \epsilon\quad \text{for all } n \geq N_\epsilon.\]

(15.2)

Combining (15.1) and (15.2) gives

\[|f_n(x) - f(x)| < \epsilon\quad \text{for all } x \in A\text{ and } n \geq N_\epsilon.\]

Thus \(f_n \to f\) uniformly on \(A\) and (as noted above) \(\mu(A^c) = 0\). So \(f_n \to f\) uniformly a.e.

\(\Leftarrow\): Assume \(f_n \to f\) uniformly a.e. This means there exists \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \(f_n \to f\) uniformly on \(A\). Let \(\epsilon > 0\) be given. Choose \(N \in \mathbb{N}\) such that \(|f_n(x) - f(x)| < \epsilon\) for every \(x \in A\) and every \(n \geq N\). Since \(A \in \mathcal{A}\) and \(\mu(A^c) = 0\), this means that \(|f_n - f| \leq \epsilon\) a.e. for every \(n \geq N\). Therefore \(\|f_n - f\|_{\infty} < \epsilon\) for every \(n \geq N\). So \(\|f_n - f\|_{\infty} \to 0\).

\(\square\)

**Theorem 15.5.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Then \(L^\infty\) is complete.

**Proof.** Let \((f_n)\) be a Cauchy sequence in \(L^\infty\). For all \(m, n \in \mathbb{N}\), define

\[A_{m,n} = \{x : |f_m(x) - f_n(x)| \leq \|f_m - f_n\|_{\infty}\}.\]
These sets are measurable because the functions \( f_n \) are measurable. Since \( |f_m - f_n| \leq \|f_m - f_n\|_\infty \) a.e. we have \( \mu(A_{m,n}^c) = 0 \). Define \( A = \bigcap_{m,n \in \mathbb{N}} A_{m,n} \). Then \( \mu(A^c) = 0 \) and
\[
|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_\infty \quad \text{for all } x \in A \text{ and } m, n \in \mathbb{N} \tag{15.3}
\]
Now we use the that assumption \((f_n)\) is Cauchy in \( L^\infty \). For every \( \epsilon > 0 \), there exists \( N_\epsilon \in \mathbb{N} \) such that
\[
\|f_m - f_n\|_\infty < \epsilon \quad \text{for all } m, n \geq N_\epsilon. \tag{15.4}
\]
Combining (15.3) and (15.4) gives: For every \( \epsilon > 0 \), there exists \( N_\epsilon \in \mathbb{N} \) such that
\[
|f_m(x) - f_n(x)| < \epsilon \quad \text{for all } x \in A \text{ and } m, n \geq N_\epsilon \tag{15.5}
\]
Thus \((f_n(x))\) is a Cauchy sequence in \( \mathbb{R} \) for each fixed \( x \in A \). By the completeness of \( \mathbb{R} \), \((f_n(x))\) converges in \( \mathbb{R} \) for each fixed \( x \in A \). Define \( f : X \to \mathbb{R} \) by
\[
f(x) = \begin{cases} 
\lim_n f_n(x) & \text{if } x \in A \\
0 & \text{if } x \in A^c
\end{cases}
\]
Note \( f \) is measurable because \( f = \lim_n f_n \mathbb{1}_A \) and the functions \( f_n \mathbb{1}_A \) are measurable.

Now we show \( f_n \to f \) in \( L^\infty \). Let \( \epsilon > 0 \). Choose \( N_\epsilon \) so that (15.4) holds. Let \( x \in A \) and \( n \geq N_\epsilon \) be arbitrary. Since \( x \in A \), we have \( f(x) = \lim_m f_m(x) \), so there exists an integer \( m(x) \geq N_\epsilon \) such that
\[
|f_m(x) - f(x)| < \epsilon.
\]
Combining this inequality and (15.5) gives
\[
|f_n(x) - f(x)| \leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \epsilon + \epsilon = 2\epsilon.
\]
This shows that
\[
|f_n(x) - f(x)| < 2\epsilon \quad \text{for all } x \in A \text{ and } n \geq N_\epsilon.
\]
Since \( \mu(A^c) = 0 \), it follows that
\[
\|f_n - f\|_\infty \leq 2\epsilon \quad \text{for all } n \geq N_\epsilon.
\]
Thus \( \|f_n - f\|_\infty \to 0 \). Moreover, since \( f_n \in L^\infty \) for all \( n \), we have
\[
\|f\|_\infty \leq \|f - f_n\|_\infty + \|f_n\|_\infty < \epsilon + \infty = \infty
\]
for all \( n \geq N_\epsilon \). So \( f \in L^\infty \). Therefore \( f_n \to f \) in \( L^\infty \).  

\( \Box \)
Chapter 5

Some Generalizations

16 Complex-Value Measurable Functions

Let \((X, \mathcal{A}, \mu)\) be a measure space.

**Definition 16.1.** We identify the set of complex numbers \(\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}\) with \(\mathbb{R}^2 = \{(a, b) : a, b \in \mathbb{R}\}\). Then the open sets in \(\mathbb{C}\) and exactly the open sets in \(\mathbb{R}^2\), and \(\mathcal{B}(\mathbb{C}) = \mathcal{B}(\mathbb{R}^2)\).

If \(z = a + ib \in \mathbb{C}\) with \(a, b \in \mathbb{R}\), the **real part** of \(z\) is \(\text{Re} z = a\) and the **imaginary part** of \(z\) is \(\text{Im} z = b\).

If \(f : X \to \mathbb{C}\), the **real part** of \(f\) and the **imaginary part** of \(f\) are the functions \(\text{Re} f : X \to \mathbb{R}\) and \(\text{Im} f : X \to \mathbb{R}\) defined by \(\text{Re}(f)(x) = \text{Re}(f(x))\) and \(\text{Im}(f)(x) = \text{Im}(f(x))\) for every \(x \in X\). Note that \(f = \text{Re} f + \text{Im} f\).

As a special case of Definition 4.3, A function \(f : X \to \mathbb{C}\) is called **\((\mathcal{A}, \mathcal{B}(\mathbb{C}))\)-measurable** (or simply **\(\mathcal{A}\)-measurable** or **measurable**) if
\[
f^{-1}(B) \in \mathcal{A} \text{ whenever } B \in \mathcal{B}(\mathbb{C}).
\]

The following characterization of measurable complex-valued functions is often useful.

**Theorem 16.2.** A function \(f : X \to \mathbb{C}\) is measurable iff \(\text{Re} f : X \to \mathbb{R}\) and \(\text{Im} f : X \to \mathbb{R}\) are measurable.

The reader is asked to prove this theorem in an exercise. Let us outline the proof here. First note \(\mathcal{B}(\mathbb{C}) = \mathcal{B}(\mathbb{R}^2)\) is generated by the collection of open rectangles \(\mathcal{R} = \{(a, b) \times (c, d) : a, b, c, d \in \mathbb{R}\}\).

To see this, in the proof of Theorem 3.9, replace the open balls \(B'_x\) with rational centers and radii by open rectangles \(R'_x\) with rational endpoints. Then use Theorem 4.4 and the formulas
\[
\begin{align*}
f^{-1}((a,b) \times (c,d)) &= (\text{Re} f)^{-1}(a,b) \cap (\text{Im} f)^{-1}(c,d) \\
f^{-1}((a,b) \times \mathbb{R}) &= (\text{Re} f)^{-1}(a,b) \\
f^{-1}(\mathbb{R} \times (c,d)) &= (\text{Im} f)^{-1}(c,d)
\end{align*}
\]
Using Theorem 16.2, the reader can show that sums, products, and limits of complex-valued measurable functions are measurable. One simply applies the analogous results for real-valued functions to the real and imaginary parts of these functions separately. The reader is asked to supply the details in an exercise.

Definition 16.3. Let $f : X \to \mathbb{C}$ be a measurable function. If $\text{Re} f$ and $\text{Im} f$ are both integrable, we say that $f$ is integrable and we define the integral of $f$ to be

$$\int f \, d\mu = \int \text{Re} f \, d\mu + i \int \text{Im} f \, d\mu$$

The reader can check that if $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ are integrable and $c \in \mathbb{C}$, then

(a) $cf$ and $f + g$ are integrable
(b) $\int cf = c \int f$ and $\int (f + g) = \int f + \int g$

To prove it, we apply the analogous results for real-valued functions to the real and imaginary parts of these functions separately. The reader is asked to supply the details in an exercise.

Theorem 16.4. Let $f : X \to \mathbb{C}$ be a measurable function. Then:

(a) $|f|$ is measurable.

(b) $f$ is integrable iff $|f|$ is integrable
(c) If $f$ is integrable, then $\left| \int f \right| \leq \int |f|$.

Proof. (a): Since $f$ is measurable, $\text{Re} f$ and $\text{Im} f$ are measurable, and so $(\text{Re} f)^2 + (\text{Im} f)^2$ is measurable. For each $a \in \mathbb{R}$, if $a \leq 0$ we have \{ $x \in X : |f(x)| \geq a$ \} $= X \in \mathcal{A}$, and if $a > 0$ we have \{ $x \in X : |f(x)| \geq a$ \} $= \{ x \in X : (\text{Re}(f(x))^2 + (\text{Im}(f(x))^2 \geq a^2 \} \in \mathcal{A}$.

(b): $f$ is integrable iff $\int f = \int \text{Re} f + i \int \text{Im} f$ is finite iff $\int \text{Re} f$ and $\int \text{Im} f$ are finite iff $\int |\text{Re} f|$ and $\int |\text{Im} f|$ are finite. But $|f| = \sqrt{|\text{Re} f|^2 + |\text{Im} f|^2} \leq |\text{Re} f| + |\text{Im} f| \leq 2|f|$, so that

$$\int |f| \leq \int |\text{Re} f| + \int |\text{Im} f| \leq 2 \int |f|.$$ 

Thus $\int |\text{Re} f|$ and $\int |\text{Im} f|$ are finite iff $\int |f|$ is finite iff $|f|$ is integrable.

(c): The inequality is trivial if $\int f = 0$. Assume $\int f \neq 0$. Set $\alpha = |\int f| (\int f)^{-1}$. Then

$$\left| \int f \right| = \alpha \int f = \int \alpha f$$

In particular, $\int \alpha f$ is real. Therefore

$$\left| \int f \right| = \alpha \int f = \int \alpha f = \text{Re} \int \alpha f = \int \text{Re}(\alpha f) \leq \int |\alpha f| = \int |f|.$$
The dominated convergence theorem is valid if the functions $f$ and $f_n$ appearing in it are complex-valued. To prove it, we apply the dominated convergence to the real and imaginary parts of these functions separately.

We can also consider $L^p$ spaces for complex-valued functions. All the definitions are analogous, all the theorems are valid, and all the proofs are identical. We use the notations $L^p(X, \mathcal{A}, \mu, \mathbb{R})$ and $L^p(X, \mathcal{A}, \mu, \mathbb{C})$ if we need distinguish the $L^p$ spaces with real-valued functions and complex-valued functions.
Chapter 6

Construction of Measures

17 Outline

In the next few sections, we will study a method for constructing measures due to Caratheodory. After developing the abstract theory, we will use the method to construct Lebesgue measure on $\mathbb{R}$, Lebesgue measure on $\mathbb{R}^d$, and product measures.

We now outline Caratheodory’s method for constructing a measure on a set $X$. We start by defining the measure only a collection $\mathcal{E}$ of “elementary” subsets of $X$. The collection $\mathcal{E}$ need not be a $\sigma$-algebra. Denote this “primitive measure” defined on $\mathcal{E}$ by $\mu_0$. Next, we use $\mu_0$ to construct a function $\mu^*$ defined on the collection $\mathcal{P}(X)$ of all subsets of $X$. Typically, $\mu^*$ will not be countably additive, hence it will not be a measure. Instead, $\mu^*$ will only be an outer measure (the definition will come later). To remedy this, we identify a collection $M(\mu^*)$ of subsets of $X$ such that $M(\mu^*)$ is a $\sigma$-algebra on $X$ and the restriction of $\mu^*$ to $M(\mu^*)$ is a measure. We denote this measure by $\mu$. If the “primitive measure” $\mu_0$ and the collection of “elementary” subsets $\mathcal{E}$ obey certain natural conditions, then the $\sigma$-algebra $M(\mu^*)$ will contain the collection $\mathcal{E}$, and $\mu^*$ and (hence) $\mu$ will agree with $\mu_0$ on $\mathcal{E}$.

Let us consider this outline for the specific example of Lebesgue measure on $\mathbb{R}$. The collection $\mathcal{E}$ of “elementary” subsets is a collection of intervals, and the “primitive measure” $\lambda_0$ gives the length of an interval. The outer measure $\lambda^*$ is called the Lebesgue outer measure. The $\sigma$-algebra $M(\lambda^*)$ is denoted by $\mathcal{L}(\mathbb{R})$; it is called the Lebesgue $\sigma$-algebra, and its elements are called Lebesgue measurable sets. The restriction of $\lambda^*$ to $M(\lambda^*) = \mathcal{L}(\mathbb{R})$ is the Lebesgue measure $\mathcal{L}(\mathbb{R})$ and is denoted by $\lambda$. The Lebesgue measure assigns every interval its length (in particular, it agrees with $\lambda_0$). The Lebesgue $\sigma$-algebra $\mathcal{L}(\mathbb{R})$ contains the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$, so that restricting $\lambda^*$ even further gives a measure on $\mathcal{B}(\mathbb{R})$, which we call Lebesgue measure on $\mathcal{B}(\mathbb{R})$ and which we also denote by $\lambda$.

18 Set Functions

Let $X$ be a set.
Definition 18.1. Let $\mathcal{A}$ be a collection of subsets of $X$. That is, $\mathcal{A} \subseteq \mathcal{P}(X)$. Let $\phi$ be a function $\phi : \mathcal{A} \to [0, \infty]$.

(i) $\phi$ is additive if $\phi(A \cup B) = \phi(A) + \phi(B)$ whenever $A, B$ are disjoint sets in $\mathcal{A}$ and $A \cup B \in \mathcal{A}$

(ii) $\phi$ is finitely additive if $\phi(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \phi(A_i)$ whenever $A_1, \ldots, A_n$ are disjoint sets in $\mathcal{A}$ and $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$

(iii) $\phi$ is countably additive if $\phi(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \phi(A_i)$ whenever $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{A}$ and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

(iv) $\phi$ is subadditive if $\phi(A \cup B) \leq \phi(A) + \phi(B)$ whenever $A, B \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$

(v) $\phi$ is finitely subadditive if $\phi(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \phi(A_i)$ whenever $A_1, \ldots, A_n \in \mathcal{A}$ and $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$

(vi) $\phi$ is countably subadditive if $\phi(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \phi(A_i)$ whenever $A_1, A_2, \ldots \in \mathcal{A}$ and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

(vii) $\phi$ is monotone if $\phi(A) \leq \phi(B)$ whenever $A, B \in \mathcal{A}$ and $A \subseteq B$.

(viii) $\phi$ is finitely monotone if $\phi(A) \leq \sum_{i=1}^{n} \phi(A_i)$ whenever $A, A_1, \ldots, A_n \in \mathcal{A}$ and $A \subseteq \bigcup_{i=1}^{n} A_i$.

(ix) $\phi$ is countably monotone if $\phi(A) \leq \sum_{i=1}^{\infty} \phi(A_i)$ whenever $A, A_1, A_2, \ldots \in \mathcal{A}$ and $A \subseteq \bigcup_{i=1}^{\infty} A_i$.

Remark. In the definition of additivity, we must assume explicitly that $A \cup B \in \mathcal{A}$ because $\mathcal{A}$ is not necessarily closed under finite unions. Similar remarks apply to the other properties.

Remark. Assume $\phi(\emptyset) = 0$. Countable additivity implies finite additivity (take $A_i = \emptyset$ for $i > n$), and finite additivity implies additivity (take $n = 2$). In general, additivity does not imply finite additivity. But, if $\mathcal{A}$ is closed under finite unions, then additivity does imply finite additivity by induction. Regardless of whether $\mathcal{A}$ is closed under countable unions, finite additivity does not imply countable additivity. Similar relationships hold amongst the three flavours of subadditivity and amongst the three flavours of monotonicity.

Recall the definition of a measure.

Definition 18.2. Let $\mathcal{A}$ be a $\sigma$-algebra on $X$. A measure on $\mathcal{A}$ is a function $\mu : \mathcal{A} \to [0, \infty]$ that is countably additive and satisfies $\mu(\emptyset) = 0$.

The concept of an outer measure will be very important in the following sections.

Definition 18.3. An outer measure on $X$ is a function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ that is countably subadditive, monotone, and satisfies $\mu^*(\emptyset) = 0$.

Remark. Since $\mu^*$ is defined on $\mathcal{P}(X)$ and since $\mu^*(\emptyset) = 0$, countably subadditivity and monotonicity together are equivalent to countable monotonicity. Thus, in the definition of outer measure, we can replace countably subadditive and monotone by countable monotonicity.

Any measure defined on $\mathcal{P}(X)$ is an outer measure on $X$. See Example 5.3 for some examples.
19 Caratheodory’s Theorem: Outer Measures to Measures

Let $X$ be a set. In this section, we tackle the following problem. Given an outer measure $\mu^*$ on $X$, we want to find a $\sigma$-algebra on $X$ such that the restriction $\mu^*$ to the $\sigma$-algebra is a measure. In other words, we want to find a $\sigma$-algebra on which $\mu^*$ is countably additive. Naturally, we want this $\sigma$-algebra to be as large as possible, so that we have many measure sets and (consequently) many measurable functions to work with.

The next lemma gives necessary and sufficient conditions for an outer measure to be countably additive on a $\sigma$-algebra.

**Lemma 19.1.** Let $\mu^*$ be an outer measure on $X$ and let $\mathcal{A}$ be a $\sigma$-algebra on $X$. Then the following are equivalent:

(a) $\mu^*$ is countably additive on $\mathcal{A}$, i.e., $\mu^*(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu^*(A_i)$ for all sequences of disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$.

(b) $\mu^*$ is additive on $\mathcal{A}$, i.e., $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ for all disjoint sets $A, B \in \mathcal{A}$.

(c) Every $F \in \mathcal{A}$ satisfies $\mu^*(E) = \mu^*(E \cap F) + \mu^*(E \cap F^c)$ for all $E \in \mathcal{A}$

**Proof.** (a) $\Rightarrow$ (b): Let $A$ and $B$ be disjoint sets in $\mathcal{A}$. Define $A_1 = A$, $A_2 = B$, and $A_i = \emptyset$ for $i > 2$. Since $\mu^*(\emptyset) = 0$ and since $\mu^*$ is assumed to be countably additive, we have $\mu^*(A \cup B) = \mu^*(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{\infty} \mu(A_i) = \mu^*(A) + \mu^*(B)$.

(b) $\Rightarrow$ (a): Let $A_1, A_2, \ldots$ be any sequence of disjoint sets in $\mathcal{A}$. By the countably subadditivity of $\mu^*$, we have $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i) = \mu^*(A) + \mu^*(B)$. We must prove the reverse inequality. Since $\mu^*$ is additive on $\mathcal{A}$ and since $\mathcal{A}$ is closed under finite unions, an induction gives that $\mu^*$ is finitely additive on $\mathcal{A}$. By the finitely additivity and monotonicity of $\mu^*$ on $\mathcal{A}$, we have

$$\sum_{i=1}^{\infty} \mu^*(A_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \mu^*(A_i) = \lim_{n \to \infty} \mu^*\left(\bigcup_{i=1}^{n} A_i\right) \leq \lim_{n \to \infty} \mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu^*(\bigcup_{i=1}^{\infty} A_i)$$

(b) $\Rightarrow$ (c): Take $A = E \cup F$ and $B = E \cap F^c$.

(c) $\Rightarrow$ (b): Take $E = A \cup B$ and $F = A$.

Condition (c) in Lemma 19.1 inspires the following definition.

**Definition 19.2.** Let $\mu^* : \mathcal{P}(X) \to [0, \infty]$. A set $F \in \mathcal{P}(X)$ is called $\mu^*$-measurable if

$$\mu^*(E) = \mu(E \cap F) + \mu^*(E \cap F^c) \text{ for all } E \in \mathcal{P}(X)$$

The collection of all $\mu^*$-measurable subsets of $X$ is denoted by $M(\mu^*)$. 

64
Remark. A $\mu^*$-measurable set $F$ can be thought of as a sharp knife; it can cut up any set $E$ into two pieces $E \cap F$ and $E \cap F^c$ so cleanly that the sum of the outer measure of the pieces equals the outer measure of the whole.

Observation. Assume $\mu^*$ is an outer measure on $X$. Let $F \in \mathcal{P}(X)$. By the countable subadditivity of $\mu^*$, the $\leq$ inequality in (19.1) holds for all $E \in \mathcal{P}(X)$. Thus, to prove $F$ is $\mu^*$-measurable, it suffices to prove the $\geq$ inequality. Moreover, the $\geq$ inequality is trivial if $\mu^*(E) = \infty$. Therefore $F$ is $\mu^*$-measurable if

$$\mu^*(E) \geq \mu(E \cap F) + \mu^*(E \cap F^c) \text{ for all } E \in \mathcal{P}(X) \text{ with } \mu^*(E) < \infty$$

(19.2)

Now we come to the main theorem of this section, which provides a solution to the problem posed above. It is due to Caratheodory.

**Theorem 19.3.** (Caratheodory’s Restriction Theorem) Let $\mu^*$ be an outer measure on $X$. Then $M(\mu^*)$ is a $\sigma$-algebra on $X$ and $\mu^*|_{M(\mu^*)}$ (the restriction of $\mu^*$ to $M(\mu^*)$) is a measure on $M(\mu^*)$.

**Proof.** To show that $M(\mu^*)$ is a $\sigma$-algebra, we first show that $\emptyset \in M(\mu^*)$ and that $M(\mu^*)$ is closed under complements. For every $E \in \mathcal{P}(X)$,

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

so $\emptyset \in M(\mu^*)$. If $A \in M(\mu^*)$, then, for every $E \in \mathcal{P}(X)$,

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

so $A^c \in M(\mu^*)$.

It remains to show that $M(\mu^*)$ is closed under countable unions. Let $A_1, A_2, \ldots \in M(\mu^*)$. We must show $\bigcup_{i=1}^{\infty} A_i \in M(\mu^*)$. Let $E \in \mathcal{P}(X)$. Define $E_0 = \bigcup_{i=1}^{\infty} A_i$ for $n \geq 0$, with the understanding that $E_0 = \bigcup_{i=1}^{0} A_i = \emptyset$. Since $A_1 \in M(\mu^*)$, we have

$$\mu^*(E) = \mu^*(E \cap A_1) + \mu^*(E \cap A_1^c) = \mu^*(E \cap E_0^c \cap A_1) + \mu^*(E \cap E_1^c).$$

Since $A_2 \in M(\mu^*)$, the last term on the right above is

$$\mu^*(E \cap E_1^c) = \mu^*(E \cap E_1^c \cap A_2) + \mu^*(E \cap E_1^c \cap A_2^c) = \mu^*(E \cap E_1^c \cap A_2) + \mu^*(E \cap E_2^c).$$

Iterating (or induction) shows that

$$\mu^*(E) = \sum_{i=1}^{n} \mu^*(E \cap E_{i-1}^c \cap A_i) + \mu^*(E \cap E_n^c).$$

Since $E_n^c \supseteq (\bigcup_{i=1}^{\infty} A_i)^c$, the monotonicity of $\mu^*$ implies

$$\mu^*(E) \geq \sum_{i=1}^{n} \mu^*(E \cap E_{i-1}^c \cap A_i) + \mu^*\left(E \cap \left(\bigcup_{i=1}^{\infty} A_i\right)^c\right).$$

65
Now we let \( n \to \infty \), use \( \bigcup_{i=1}^{\infty} (E \cap E_{i-1}^c \cap A_i) = E \cap (\bigcup_{i=1}^{\infty} A_i) \), and apply the countable subadditivity of \( \mu^* \) (twice) to get

\[
\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(E \cap E_{i-1}^c \cap A_i) + \mu^*(E \cap \left( \bigcup_{i=1}^{\infty} A_i \right)^c) \\
\geq \mu^*(E \cap \left( \bigcup_{i=1}^{\infty} A_i \right)) + \mu^*(E \cap \left( \bigcup_{i=1}^{\infty} A_i \right)^c) \\
\geq \mu^*(E).
\]

Thus all the inequalities in the last calculation are inequalities. This proves \( \bigcup_{i=1}^{\infty} A_i \in M(\mu^*) \). Therefore \( M(\mu^*) \) is a \( \sigma \)-algebra.

Furthermore, if \( E = \bigcup_{i=1}^{\infty} A_i \) and the sets \( A_1, A_2, \ldots \) are disjoint, then \( E \cap E_{i-1}^c \cap A_i = A_i \) and \( E \cap (\bigcup_{i=1}^{\infty} A_i)^c = \emptyset \), hence the calculation above reduces to the statement

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

So \( \mu^* \) is countably additive on \( M(\mu^*) \). Thus the restriction of \( \mu^* \) to \( M(\mu^*) \) is a measure on \( M(\mu^*) \).

\( \square \)
20 Construction of Outer Measures

Let $X$ be a set. In this section, we see how to construct an outer measure $\mu^*$ on $X$ by starting with a function $\mu_0$ defined on a collection $\mathcal{E}$ of subsets of $X$. Moreover, we will give sufficient conditions for the outer measure $\mu^*$ to be an extension of $\mu_0$ and for $M(\mu^*)$ to contain $\mathcal{E}$.

**Definition 20.1.** A demi-ring on $X$ is a collection $\mathcal{E} \subseteq \mathcal{P}(X)$ that satisfies the following properties.

(i) $\emptyset \in \mathcal{E}$

(ii) If $A, B \in \mathcal{E}$, then there exists a finite collection of disjoint sets $C_1, \ldots, C_m \in \mathcal{E}$ such that

$$A \setminus B = \bigcup_{i=1}^m C_i.$$ 

Every $\sigma$-algebra is a demi-ring, but not conversely.

**Lemma 20.2.** Let $\mathcal{E}$ be a demi-ring on $X$.

(a) If $A, A_1, \ldots, A_n \in \mathcal{E}$, then there exists a finite collection of disjoint sets $C_1, \ldots, C_m \in \mathcal{E}$ such that

$$A \setminus \bigcup_{i=1}^n A_i = \bigcup_{j=1}^m C_j.$$ 

(b) If $A, B \in \mathcal{E}$, then there exists a finite collection of disjoint sets $C_1, \ldots, C_m \in \mathcal{E}$ such that

$$A \cap B = \bigcup_{j=1}^m C_j.$$ 

**Proof.**

(a): We prove this by induction on $n$. If $n = 1$, this is immediate from the definition of $\mathcal{E}$ as a demi-ring. Assume the statement holds for some $n \in \mathbb{N}$. So there exists disjoint sets $D_1, \ldots, D_k \in \mathcal{A}$ such that $A \setminus \bigcup_{i=1}^n A_i = \bigcup_{i=1}^k D_i$. Then

$$A \setminus \bigcup_{i=1}^{n+1} A_i = \left( A \setminus \bigcup_{i=1}^n A_i \right) \setminus A_{n+1} = \left( \bigcup_{i=1}^k D_i \right) \setminus A_{n+1} = \bigcup_{i=1}^k (D_i \setminus A_{n+1}).$$ 

By the definition of $\mathcal{E}$ as a demi-ring, for each $i \in \{1, \ldots, k\}$, there exist disjoint sets $E_{i,1}, \ldots, E_{i,\ell_i} \in \mathcal{E}$ such that $D_i \setminus A_{n+1} = \bigcup_{j=1}^{\ell_i} E_{i,j}$. Then

$$A \setminus \bigcup_{i=1}^{n+1} A_i = \bigcup_{i=1}^k \bigcup_{j=1}^{\ell_i} E_{i,j}.$$ 

By construction, for each fixed $i \in \{1, \ldots, k\}$, the sets $E_{i,1}, \ldots, E_{i,\ell_i} \in \mathcal{E}$ are disjoint. If $i, i' \in \{1, \ldots, k\}$ with $i \neq i'$, then for every $j \in \{1, \ldots, \ell_i\}$ and every $j' \in \{1, \ldots, \ell_{i'}\}$ we have $E_{i,j} \subseteq D_i$ and $E_{i',j'} \subseteq D_{i'}$, so $E_{i,j}$ and $E_{i',j'}$ are disjoint because $D_i$ and $D_{i'}$ are disjoint. Thus the sets $E_{i,j}$ are disjoint for all $i, j$. By relabeling the $E_{i,j}$’s as $F_1, \ldots, F_m$ where $m = \ell_1 + \cdots + \ell_k$, we have

$$A \setminus \bigcup_{i=1}^{n+1} A_i = \bigcup_{i=1}^m F_i$$ 

and $F_1, \ldots, F_m$ are disjoint sets in $\mathcal{E}$. This completes the proof by induction.
Let’s show that

\( \mu \)

\( \mu \)

(a): For every \( E \in \mathcal{E} \), hence

\[ A \cap B = A \setminus (A \setminus B) = A \setminus \left( \bigcup_{i=1}^{m} C_i \right) \]

The desired result now follows from (a).

\[ \square \]

**Theorem 20.3.** (Caratheodory Construction of Outer Measures) Let \( \mathcal{E} \subseteq \mathcal{P}(X) \) and let \( \mu_0: \mathcal{E} \to [0, \infty] \). Assume \( \emptyset \in \mathcal{E} \) and \( \mu_0(\emptyset) = 0 \). Define \( \mu^*: \mathcal{P}(X) \to [0, \infty] \) by

\[
\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu_0(E_i) : E_1, E_2, \ldots \in \mathcal{E}, A \subseteq \bigcup_{i=1}^{\infty} E_i \right\}
\]

for every \( A \in \mathcal{P}(X) \), with the convention that \( \inf \emptyset = \infty \).

(a) \( \mu^* \) is an outer measure on \( X \). It is called the outer measure on \( X \) generated by \( \mu_0 \).

(b) If \( \mu_0 \) is countably monotone, then \( \mu^*(A) = \mu_0(A) \) for all \( A \in \mathcal{E} \).

(c) If \( \mathcal{E} \) is a demi-ring and \( \mu_0 \) is finitely additive, then \( \mathcal{E} \subseteq \mathcal{M}(\mu^*) \).

**Remark.** The idea behind the definition of \( \mu^* \) is as follows. Given a set \( A \subseteq X \), we wish to measure it from the outside. So we cover \( A \) by a countable unions of sets in \( \mathcal{E} \). The measure of \( A \) should be no larger than the sum of the measures of the sets in the cover. We define the outer measure of \( A \) to be the smallest possible sum that can be obtained in this way.

**Remark.** It is easy to check that the following statements also hold,

(b’) \( \mu_0 \) is countably monotone if and only if \( \mu^*(A) = \mu_0(A) \) for all \( A \in \mathcal{E} \).

(c’) If \( \mathcal{E} \subseteq \mathcal{M}(\mu^*) \), then \( \mu_0 \) is finitely additive.

**Proof.**

(a): For every \( E_1, E_2, \ldots \in \mathcal{E} \), we have \( \sum_{i=1}^{\infty} \mu_0(E_i) \in [0, \infty] \). So \( \mu^*(A) \in [0, \infty] \) for every \( A \in \mathcal{P}(X) \).

Let’s show that \( \mu^*(\emptyset) = 0 \). If \( E_i = \emptyset \) for \( i = 1, 2, \ldots \), then \( E_1, E_2, \ldots \in \mathcal{E} \) and \( \emptyset \subseteq \bigcup_{i=1}^{\infty} E_i \), hence \( \mu^*(\emptyset) \leq \sum_{i=1}^{\infty} \mu_0(E_i) = 0 \). Therefore \( \mu^*(\emptyset) = 0 \).

Now let’s show that \( \mu^* \) is monotone. Let \( A, B \in \mathcal{P}(X) \) with \( A \subseteq B \). For any sequence \( E_1, E_2, \ldots \in \mathcal{E} \), if \( B \subseteq \bigcup_{i=1}^{\infty} E_i \), then \( A \subseteq \bigcup_{i=1}^{\infty} E_i \). Thus

\[
\left\{ \sum_{i=1}^{\infty} \mu_0(E_i) : E_1, E_2, \ldots \in \mathcal{E}, B \subseteq \bigcup_{i=1}^{\infty} E_i \right\} \subseteq \left\{ \sum_{i=1}^{\infty} \mu_0(E_i) : E_1, E_2, \ldots \in \mathcal{E}, A \subseteq \bigcup_{i=1}^{\infty} E_i \right\}
\]

Taking infimums gives \( \mu^*(A) \leq \mu^*(B) \). Therefore \( \mu^* \) is monotone.

Finally, let’s show that \( \mu^* \) is countably subadditive. Let \( A_1, A_2, \ldots \in \mathcal{P}(X) \). Let \( \epsilon > 0 \). For each \( n \in \mathbb{N} \), by the definition of \( \mu^*(A_n) \) as an infimum, there exists \( E_{1,n}, E_{2,n}, \ldots \in \mathcal{E} \) such that
\( A_n \subseteq \bigcup_{m=1}^{\infty} E_{m,n} \) and
\[
\sum_{m=1}^{\infty} \mu_0(E_{m,n}) \leq \mu^*(A_n) + \frac{\epsilon}{2^n}.
\]

Then \( \bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} E_{m,n} \) and
\[
\mu^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mu_0(E_{m,n}) \leq \sum_{n=1}^{\infty} \left( \mu^*(A_n) + \frac{\epsilon}{2^n} \right) = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon.
\]

Since \( \epsilon > 0 \) is arbitrary, we have
\[
\mu^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).
\]

Therefore \( \mu^* \) is countably subadditive. This completes the proof that \( \mu^* \) is an outer measure.

(b): Assume \( \mu_0 \) is countably monotone.

First we show that \( \mu^*(A) \leq \mu_0(A) \) for every \( A \in \mathcal{E} \). Let \( A \in \mathcal{E} \). If \( E_1 = A \) and \( E_i = \emptyset \) for \( i > 1 \), then \( E_1, E_2, \ldots \in \mathcal{E} \) and \( A \subseteq \bigcup_{i=1}^{\infty} E_i \), hence
\[
\mu^*(A) \leq \sum_{i=1}^{\infty} \mu_0(E_i) = \mu_0(A).
\]

Now we show that \( \mu_0(A) \leq \mu^*(A) \) for every \( A \in \mathcal{E} \). Let \( A \in \mathcal{E} \). If \( E_1, E_2, \ldots \in \mathcal{E} \) with \( A \subseteq \bigcup_{i=1}^{\infty} E_i \), then the countable monotonicity of \( \mu_0 \) implies that \( \mu_0(A) \leq \sum_{i=1}^{\infty} \mu_0(E_i) \). Therefore \( \mu_0(A) \) is a lower bound for the set
\[
\left\{ \sum_{i=1}^{\infty} \mu_0(E_i) : E_1, E_2, \ldots \in \mathcal{E}, A \subseteq \bigcup_{i=1}^{\infty} E_i \right\}.
\]

Since \( \mu^*(A) \) is the infimum of this set, we have \( \mu_0(A) \leq \mu^*(A) \).

(c): Assume \( \mathcal{E} \) is a demi-ring and \( \mu_0 \) is finitely additive.

To show that \( \mathcal{E} \subseteq M(\mu^*) \), we must show that every \( A \in \mathcal{E} \) satisfies
\[
\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \quad \text{for every } E \in \mathcal{P}(X).
\]

Let \( A \in \mathcal{E} \). Let \( E \in \mathcal{P}(X) \). Let \( \epsilon > 0 \). By the definition of \( \mu^*(E) \) as an infimum, there exists \( E_1, E_2, \ldots \in \mathcal{E} \) such that \( E \subseteq \bigcup_{n=1}^{\infty} E_n \) and
\[
\sum_{n=1}^{\infty} \mu_0(E_n) \leq \mu^*(E) + \epsilon. \quad (20.1)
\]

By Part (b) of Lemma 20.2, for each \( n \in \mathbb{N} \), there exist disjoint \( C_{n,1}, \ldots, C_{n,K_n} \in \mathcal{A} \) such that
\[
E_n \cap A = \bigcup_{k=1}^{K_n} C_{n,k}.
\]
By the definition of a demi-ring, for each \( n \in \mathbb{N} \), there exist disjoint \( D_{n,1}, \ldots, D_{n,L_n} \in A \) such that

\[
E_n \cap A^c = E_n \setminus A = \bigcup_{\ell=1}^{L_n} D_{n,\ell}.
\]

Since \( E_n \cap A \) and \( E_n \setminus A \) are disjoint, the sets \( C_{n,1}, \ldots, C_{n,K_n}, D_{n,1}, \ldots, D_{n,L_n} \) are all disjoint. Note that

\[
E_n = (E_n \cap A) \cup (E_n \cap A^c) = \left( \bigcup_{k=1}^{K_n} C_{n,k} \right) \cup \left( \bigcup_{\ell=1}^{L_n} D_{n,\ell} \right).
\]

By the finite additivity of \( \mu_0 \),

\[
\mu_0(E_n) = \sum_{k=1}^{K_n} \mu_0(C_{n,k}) + \sum_{\ell=1}^{L_n} \mu_0(D_{n,\ell}).
\]

Combining this with (20.1) gives

\[
\sum_{n=1}^{\infty} \sum_{k=1}^{K_n} \mu_0(C_{n,k}) + \sum_{n=1}^{\infty} \sum_{\ell=1}^{L_n} \mu_0(D_{n,\ell}) \leq \mu^*(E) + \epsilon. \tag{20.2}
\]

Since \( E \subseteq \bigcup_{n=1}^{\infty} E_n \), we have

\[
E \cap A \subseteq \bigcup_{n=1}^{\infty} (E_n \cap A) \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{K_n} C_{n,k}
\]

and

\[
E \cap A^c \subseteq \bigcup_{n=1}^{\infty} (E_n \cap A^c) \subseteq \bigcup_{n=1}^{\infty} \bigcup_{\ell=1}^{L_n} D_{n,\ell}
\]

Recall that \( C_{n,k} \in A \) for all \( n, k \) and \( D_{n,\ell} \in A \) for all \( n, \ell \). Then the definition of \( \mu^* \) implies

\[
\mu^*(E \cap A) \leq \sum_{n=1}^{\infty} \sum_{k=1}^{K_n} \mu^*(C_{n,k})
\]

and

\[
\mu^*(E \cap A^c) \leq \sum_{n=1}^{\infty} \sum_{\ell=1}^{L_n} \mu^*(D_{n,\ell}).
\]

Combining these inequalities with (20.2) gives

\[
\mu(E \cap A) + \mu(E \cap A^c) \leq \mu^*(E) + \epsilon.
\]

Since \( \epsilon > 0 \) is arbitrary, we have

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

The reverse inequality holds by the subadditivity of the outer measure \( \mu^* \). This completes the proof that \( \mathcal{E} \subseteq M(\mu^*) \). \( \square \)
Theorem 20.4. Let $X$ be a set. Let $\mathcal{E} \subseteq \mathcal{P}(X)$ and let $\mu_0 : \mathcal{E} \to [0, \infty]$. Assume $\emptyset \in \mathcal{E}$ and $\mu_0(\emptyset) = 0$. Let $\mu^*$ be the outer measure on $X$ generated by $\mu_0$. Let $A \in \mathcal{P}(X)$.

(a) Assume there exist $E_1, E_2, \ldots \in \mathcal{E}$ such that $A \subseteq \bigcup_{i=1}^{\infty} E_i$. For every $\epsilon > 0$ there exists $G_\epsilon \in \mathcal{E}_\sigma$ such that $A \subseteq G_\epsilon$ and $\mu^*(A) \leq \mu^*(G_\epsilon) \leq \mu^*(A) + \epsilon$. Moreover, there exists $H \in \mathcal{E}_\sigma\delta$ such that $A \subseteq H$ and $\mu^*(A) = \mu^*(H)$.

(b) Assume $\mathcal{E} \subseteq M(\mu^*)$. If $\mu^*(A) < \infty$, then the following are equivalent:

(i) $A \in M(\mu^*)$

(ii) For every $\epsilon > 0$ there exists $G_\epsilon \in \mathcal{E}_\sigma$ such that $A \subseteq G_\epsilon$ and $\mu^*(G_\epsilon \setminus A) \leq \epsilon$.

(iii) There exists $H \in \mathcal{E}_\sigma\delta$ such that $A \subseteq H$ and $\mu^*(H \setminus A) = 0$.

(c) Assume $\mathcal{E} \subseteq M(\mu^*)$. If $X$ is $\sigma$-finite with respect to $\mu_0$, then the following are equivalent:

(i) $A \in M(\mu^*)$

(ii) For every $\epsilon > 0$ there exists $G_\epsilon \in \mathcal{E}_\sigma$ such that $A \subseteq G_\epsilon$ and $\mu^*(G_\epsilon \setminus A) \leq \epsilon$.

(iii) There exists $H \in \mathcal{E}_\sigma\delta$ such that $A \subseteq H$ and $\mu^*(H \setminus A) = 0$.

Proof. Exercise. \qed
Let $X$ be a set. In this section, we prove the existence and uniqueness of a measure $\mu$ that extends a given primitive function $\mu_0$ defined on a collection of elementary subsets of $X$.

**Theorem 21.1.** Let $\mathcal{E}$ be a demi-ring on $X$. Suppose $\mu_0 : \mathcal{E} \to [0, \infty]$ is countably monotone, finitely additive, and satisfies $\mu_0(\emptyset) = 0$. Let $\mu^*$ be the outer measure on $X$ generated by $\mu_0$, as in Theorem 20.3(a). Let $\mathcal{A}$ be a $\sigma$-algebra on $X$ such that $\mathcal{E} \subseteq \mathcal{A} \subseteq M(\mu^*)$. Let $\mu = \mu^*|_{\mathcal{A}}$ (i.e., $\mu$ is the restriction of $\mu^*$ to $\mathcal{A}$).

(a) $\mu$ is a measure on $\mathcal{A}$ and $\mu = \mu_0$ on $\mathcal{E}$.

(b) If $\nu$ is a measure on $\mathcal{A}$ such that $\nu = \mu_0$ on $\mathcal{E}$, then $\nu \leq \mu$ on $\mathcal{A}$.

(c) If $\nu$ is a measure on $\mathcal{A}$ such that $\nu = \mu_0$ on $\mathcal{E}$ and if $X$ is $\sigma$-finite with respect to $\mu_0$ (i.e., $X = \bigcup_{i=1}^{\infty} E_i$ for some sets $E_i \in \mathcal{A}$ with $\mu_0(E_i) < \infty$), then $\nu = \mu$.

**Proof.**

(a): Theorem 20.3(a) implies $\mu^*$ is an outer measure on $X$. Theorem 19.3 implies that $M(\mu^*)$ is a $\sigma$-algebra and $\mu^*|_{M(\mu^*)}$ is a measure on $M(\mu^*)$. Since $\mathcal{A} \subseteq M(\mu^*)$, $\mu$ is a measure on $\mathcal{A}$. Theorem 20.3(b) implies that $\mu^* = \mu_0$ on $\mathcal{E}$. Since $\mathcal{E} \subseteq \mathcal{A}$ and $\mu = \mu^*$ on $\mathcal{A}$, $\mu = \mu_0$ on $\mathcal{E}$.

(b): Let $A \in \mathcal{A}$. If $(E_i) \subseteq \mathcal{E}$ and $A \subseteq \bigcup_{i=1}^{\infty} E_i$, then $\nu(A) \leq \sum_{i=1}^{\infty} \nu(E_i) = \sum_{i=1}^{\infty} \mu_0(E_i)$. The definition of $\mu^*(A)$ as an infimum implies $\nu(A) \leq \mu^*(A)$. Since $\mu = \mu^*$ on $\mathcal{A}$, we get $\nu(A) \leq \mu(A)$.

(c): Claim: If $A \in \mathcal{A}$, $E \in \mathcal{E}$, $A \subseteq E$, and $\mu_0(E) < \infty$, then $\mu(A) = \nu(A)$.

Proof of Claim: By (b), $\nu(A) \leq \mu(A) \leq \mu(E) \leq \mu_0(E) < \infty$ and $\nu(E \setminus A) \leq \mu(E \setminus A)$. Therefore

$$\mu_0(E) - \nu(A) = \nu(E) - \nu(A) \leq \nu(E \setminus A) \leq \mu(E \setminus A) = \mu(E) - \mu(A) \leq \mu_0(E) - \mu(A),$$

hence $\mu(A) \leq \nu(A)$.

Now we use the claim to prove (c). Let $A \in \mathcal{A}$. Since $X$ is $\sigma$-finite with respect to $\mu_0$, there exist sets $E_1, E_2, \ldots \in \mathcal{E}$ such that $X = \bigcup_{i=1}^{\infty} E_i$ and $\mu_0(E_i) < \infty$ for all $i$. Define $B_i = E_i \setminus \bigcup_{j=1}^{i-1} E_j$ for $i \geq 1$ with the understanding that $\bigcup_{j=1}^{0} E_j = \emptyset$. Then $B_1, B_2, \ldots$ are disjoint sets in $\mathcal{A}$ such that $X = \bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} B_i$ and $B_i \subseteq E_i$ for all $i$. Then $A \cap B_1, A \cap B_2, \ldots$ are disjoint sets in $\mathcal{A}$ such that $A = \bigcup_{i=1}^{\infty} (A \cap B_i)$ and $\mu(A \cap B_i) \leq \mu(B_i) \leq \mu(E_i) = \mu_0(E_i) < \infty$ for all $i$. The claim implies that $\mu(A \cap B_i) = \nu(A \cap B_i)$ for all $i$. So by the countable additivity of $\mu$ and $\nu$ we have

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A \cap B_i) = \sum_{i=1}^{\infty} \nu(A \cap B_i) = \nu(A).$$

**Remark.** (b) and (c) do not require that $\mathcal{E}$ is a demi-ring. The fact that $\nu = \mu_0$ on $\mathcal{E}$ implies that $\mu_0$ is finitely additive and countably monotone, so those assumptions can also be dropped.
Chapter 7

Lebesgue Measure on $\mathbb{R}$

22 Lebesgue Measure on $\mathbb{R}$

In this section, we construct the Lebesgue measure on $\mathbb{R}$.

**Theorem 22.1.** Let $\mathcal{E}$ be the collection of all intervals of the form $(a, b]$ where $a, b \in \mathbb{R}$ and $a \leq b$. Define $\lambda_0 : \mathcal{E} \to [0, \infty]$ by

$$\lambda_0((a, b]) = b - a$$

for each $(a, b] \in \mathcal{E}$. Notice that $\emptyset = (a, a] \in \mathcal{E}$ and $\lambda_0(\emptyset) = \lambda((a, a]) = 0$. Let $\lambda^*$ be the outer measure generated by $\lambda_0$, as in Theorem 20.3(a). It is called the **Lebesgue outer measure** on $\mathbb{R}$.

(a) $\mathcal{E}$ is a demi-ring on $\mathbb{R}$.

(b) $\lambda_0$ is finitely additive.

(c) $\lambda_0$ is countably monotone.

(d) $\mathbb{R}$ is $\sigma$-finite with respect to $\mu_0$.

(e) $\lambda^* = \lambda_0$ on $\mathcal{E}$. Moreover, $\lambda^*$ assigns each interval its length.

(f) Define $\mathcal{L}(\mathbb{R}) = M(\lambda^*)$. We call $\mathcal{L}(\mathbb{R})$ the **Lebesgue $\sigma$-algebra on $\mathbb{R}$**. It satisfies $\mathcal{E} \subseteq \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}) \subseteq \mathcal{L}(\mathbb{R})$.

(g) $\lambda^*|_{\mathcal{L}(\mathbb{R})}$ is a measure on $\mathcal{L}(\mathbb{R})$. It is the unique measure on $\mathcal{L}(\mathbb{R})$ that agrees with $\lambda_0$ on $\mathcal{E}$. It is called **Lebesgue measure on $\mathcal{L}(\mathbb{R})$**. It is denoted by $\lambda$.

(h) $\lambda^*|_{\mathcal{B}(\mathbb{R})}$ is a measure on $\mathcal{B}(\mathbb{R})$. It is the unique measure on $\mathcal{B}(\mathbb{R})$ that agrees with $\lambda_0$ on $\mathcal{E}$. It is called **Lebesgue measure on $\mathcal{B}(\mathbb{R})$**. It is also denoted by $\lambda$.

**Proof.**

(a) Let $(a, b], (c, d] \in \mathcal{E}$.

Case 1: If $a \leq b \leq c \leq d$ or $c \leq d \leq a \leq b$, then $(a, b] \setminus (c, d] = (a, b]$. 

73
Case 2: If \( a \leq c \leq b \leq d \), then \( (a,b) \setminus (c,d) = (a,c) \).

Case 3: If \( c \leq a \leq d \leq b \), then \( (a,b) \setminus (c,d) = (d,b) \).

Case 4: If \( a \leq c \leq d \leq b \), then \( (a,b) \setminus (c,d) = (a,c] \cup (d,b) \).

Case 5: If \( c \leq a \leq b \leq d \), then \( (a,b) \setminus (c,d) = \emptyset \).

In all cases, \( (a,b) \setminus (c,d) \) is a finite union of disjoint members of \( \mathcal{E} \). Thus \( \mathcal{E} \) is a demi-ring.

(b) Assume \( (a,b),(c,d) \in \mathcal{E} \) are disjoint and \( (a,b) \cup (c,d) \in \mathcal{E} \).

Case 1: \( (a,b) \neq \emptyset \) and \( (c,d) \neq \emptyset \). Then \( a \leq b = c \leq d \) or \( c \leq d = a \leq b \). Without loss of generality, assume \( a \leq b = c \leq d \). Then \( (a,b) \cup (c,d) = (a,d) \) and

\[
\lambda_0((a,b) \cup (c,d)) = \lambda_0((a,d)) = d - a = d - a + b - c = b - a + d - c = \lambda_0((a,b)) + \lambda_0((c,d))
\]

Case 2: \( (a,b) = \emptyset \) or \( (c,d) = \emptyset \). Without loss of generality, assume \( (a,b) = \emptyset \). Then \( a = b \), \( (a,b) \cup (c,d) = (c,d) \), and

\[
\lambda_0((a,b) \cup (c,d)) = \lambda_0((c,d)) = d - c = b - a + d - c = \lambda_0((a,b)) + \lambda_0((c,d)).
\]

(c) Let \( (a,b),(a_1,b_1),(a_2,b_2),\ldots, \in \mathcal{E} \) such that \( (a,b) \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n) \). If \( a = b \), then \( \lambda_0((a,b)) = 0 \leq \sum_{n=1}^{\infty} \lambda_0((a_n, b_n)) \), and we are done. Assume \( a < b \). Let \( \epsilon > 0 \) such that \( a + \epsilon < b \). Define \( c = a + \epsilon \), \( d = b \), \( c_n = a_n \), and \( d_n = b_n + \epsilon 2^{-n} \). Then

\[
[c,d] \subseteq (a,b) \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n) \subseteq \bigcup_{n=1}^{\infty} (c_n, d_n).
\]

Since \( [c,d] \) is compact and the collection \( \{(c_n, d_n) : n \in \mathbb{N}\} \) is an open cover of \( [c,d] \), there is a finite subcover \( \{(c_n, d_n) : n = 1, 2, \ldots, N\} \). So

\[
[c,d] \subseteq \bigcup_{n=1}^{N} (c_n, d_n).
\]

Take the collection \( \{(c_n, d_n) : n = 1, 2, \ldots, N\} \), discard any \( (c_n, d_n) \) that is a subset of another \( (c_m, d_m) \), and relabel so that \( c \in (c_1, d_1) \), \( d_1 \in (c_2, d_2) \), \( d_2 \in (c_3, d_3) \), \ldots, \( d \in (a_N, d_N) \). Then \( c_1 < c < d < d_N \) and \( c_n < d_{n-1} \) for \( n = 2, \ldots, N \). Therefore

\[
\sum_{n=1}^{N} (d_n - c_n) = d_N - c_N + d_{N-1} - c_{N-1} + \cdots + d_2 - c_2 + d_1 - c_1
\]

\[
= d_N + (d_{N-1} - c_N) + (d_{N-2} - c_{N-1}) + \cdots + (d_1 - c_2) - c_1
\]

\[
> d_N - c_1 > d - c = b - a - \epsilon.
\]

On the other hand,

\[
\sum_{n=1}^{N} (d_n - c_n) \leq \sum_{n=1}^{\infty} (d_n - c_n) = \sum_{n=1}^{\infty} (b_n + \epsilon 2^{-n} - a_n) = \epsilon + \sum_{n=1}^{\infty} (b_n - a_n).
\]

Thus

\[
b - a - \epsilon < \epsilon + \sum_{n=1}^{\infty} (b_n - a_n).
\]
Letting $\epsilon \to 0$ gives

$$b - a \leq \sum_{n=1}^{\infty} (b_n - a_n),$$

which is equivalent to

$$\lambda_0((a, b]) \leq \sum_{n=1}^{\infty} \lambda_0((a_n, b_n)).$$

(d) $\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n]$ and $\mu_0((-n, n]) = 2n < \infty$ for all $n \in \mathbb{N}$.

(e) By (c) and Theorem 20.3(b), $\lambda^* = \lambda_0$ on $\mathcal{E}$. Let us check that, for each interval $I$, $\lambda^*(I)$ is equal to the length of the interval.

Let $a, b \in \mathbb{R}$ with $a \leq b$. Since $\lambda^* = \lambda_0$ on $\mathcal{E}$, we have $\lambda^*([a, b]) = \lambda_0((a, b]) = b - a$.

Assume $a < b$. For every $0 < \epsilon < b - a$,

$$b - a - \epsilon = \lambda^*([a, b - \epsilon]) \leq \lambda^*([a, b]) \leq \lambda^*([a, b]) \leq \lambda^*([a - \epsilon, b]) = b - a + \epsilon.$$

Letting $\epsilon \to 0$ gives $\lambda^*([a, b]) = \lambda^*([a, b]) = \lambda^*([a, b]) = b - a$.

Now assume $a = b$. For every $\epsilon > 0$,

$$0 \leq \lambda^*([a, b]) \leq \lambda^*([a, b]) \leq \lambda^*([a, b]) \leq \lambda^*((a - \epsilon, b]) = b - a + \epsilon = \epsilon.$$

Letting $\epsilon \to 0$ gives $\lambda^*([a, b]) = \lambda^*([a, b]) = \lambda^*([a, b]) = 0$.

Let $I$ be an interval with infinite length. In other words, $I$ is an interval of the form $(a, \infty), (-\infty, b), [a, \infty), (-\infty, b), (-\infty, \infty)$ where $a, b \in \mathbb{R}$. For every $M > 0$ there are $a, b \in \mathbb{R}$ with $a \leq b$ such that $(a, b) \subseteq I$ and $b - a > M$, hence $\lambda^*(I) \geq \lambda^*([a, b]) = b - a > M$. Thus $\lambda^*(I) = \infty$.

(f) Theorem 20.3(c) implies $\mathcal{E} \subseteq \mathcal{L}(\mathbb{R})$. However, Theorem 3.10 says that $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-algebra on $\mathbb{R}$ that contains $\mathcal{E}$. Thus $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{L}(\mathbb{R})$.

(g) Theorem 19.3 implies $\lambda^*|_{\mathcal{L}(\mathbb{R})}$ is a measure on $\mathcal{L}(\mathbb{R})$. By (a),(b),(c),(d),(e),(f) and Theorem 21.1(c), $\lambda^*|_{\mathcal{L}(\mathbb{R})}$ is the unique measure on $\mathcal{L}(\mathbb{R})$ that equals $\lambda_0$ on $\mathcal{E}$.

(h) Since $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{L}(\mathbb{R})$ and since $\lambda^*|_{\mathcal{L}(\mathbb{R})}$ is a measure on $\mathcal{L}(\mathbb{R})$, $\lambda^*|_{\mathcal{B}(\mathbb{R})}$ is a measure on $\mathcal{B}(\mathbb{R})$.

By (a),(b),(c),(d),(e),(f) and Theorem 21.1(c), $\lambda^*|_{\mathcal{B}(\mathbb{R})}$ is the unique measure on $\mathcal{B}(\mathbb{R})$ that equals $\lambda_0$ on $\mathcal{E}$.

$\square$
23 Complete Measures

Definition 23.1. Let \((X, \mathcal{A}, \mu)\) be a measure space. We say that \((X, \mathcal{A}, \mu)\) is complete (or \(\mu\) is complete) if the following property holds: If \(N \in \mathcal{A}, \mu(N) = 0\), and \(A \subseteq N\), then \(A \in \mathcal{A}\). In other words, \(\mu\) is complete if all subsets of measure zero sets are measurable.

Theorem 23.2. Let \((X, \mathcal{A}, \mu)\) be a measure space. The following are equivalent.

(a) \((X, \mathcal{A}, \mu)\) is complete.

(b) For all \(f, g : X \to \mathbb{R}\), if \(f\) is measurable and \(f = g\) a.e., then \(g\) is measurable.

(c) For all \(f, f_1, f_2, \ldots : X \to \mathbb{R}\), if \(f_n\) is measurable for each \(n\) and \(f_n \to f\) a.e., then \(f\) is measurable.

Proof. (a) \(\Rightarrow\) (b): Assume \(f\) is measurable and \(f = g\) a.e. Then there exists a set \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \(f(x) = g(x)\) for all \(x \in A\). Let \(B \in \mathcal{B}(\mathbb{R})\) be given. Write

\[
g^{-1}(B) = (g^{-1}(B) \cap A) \cup (g^{-1}(B) \cap A^c).
\]

Note \(g^{-1}(B) \cap A = f^{-1}(B) \cap A \in \mathcal{A}\). Note \(g^{-1}(B) \cap A^c \subseteq A^c, A^c \in \mathcal{A}\), and \(\mu(A^c) = 0\). Since \(\mu\) is complete, \(g^{-1}(B) \cap A^c \in \mathcal{A}\). Therefore \(g^{-1}(B) \in \mathcal{A}\). This proves \(g\) is measurable.

(b) \(\Rightarrow\) (c): Assume \(f_n\) is measurable for each \(n\) and \(f_n \to f\) a.e. Then there exists a set \(A \in \mathcal{A}\) such that \(\mu(A^c) = 0\) and \(f_n(x) \to f(x)\) for all \(x \in A\). Define \(g_n = f_n 1_A\) and \(g = f 1_A\). Then \(g_n\) is measurable for each \(n\) and \(g_n \to g\) pointwise. So \(g\) is measurable. Moreover, \(g = f\) a.e. Then (b) implies \(f\) is measurable.

(c) \(\Rightarrow\) (b): Assume \(f\) is measurable and \(f = g\) a.e. Define \(h_n = f\) for all \(n\). Then \(h_n\) is measurable for all \(n\) and \(h_n \to g\) a.e.. Then (c) implies \(g\) is measurable.

(b) \(\Rightarrow\) (a): We prove the contrapositive. Assume \(\mu\) is not complete. So there exist sets \(A, N \subseteq X\) such that \(N \in \mathcal{A}\), \(\mu(N) = 0\), \(A \subseteq N\), and \(A \notin \mathcal{A}\). Define \(f = 1_N\) and \(g = 1_{A^c}\). Then \(f = g\) a.e., \(f\) is measurable, but \(g\) is not measurable.

The next theorem says that the Carathéodory restriction theorem produces a complete measure.

Theorem 23.3. Let \(\mu^*\) be an outer measure on a set \(X\). Let \(M(\mu^*)\) be the \(\sigma\)-algebra of \(\mu^*\)-measurable sets and let \(\mu\) be the measure which is the restriction of \(\mu^*\) to \(M(\mu^*)\). Then \((X, M(\mu^*), \mu)\) is complete.

Proof. Assume \(N \in M(\mu^*), \mu(N) = 0\), and \(A \subseteq N\). Then \(\mu^*(A) \leq \mu^*(N) = \mu(N) = 0\), hence \(\mu^*(A) = 0\). Therefore, for any \(E \in \mathcal{P}(X)\),

\[
\mu^*(E \cap A) + \mu^*(E \cap A^c) \leq \mu^*(A) + \mu^*(E) = \mu^*(E).
\]

So \(A \in M(\mu^*)\).

Corollary 23.4. \((\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)\) is complete.
24 Completion of Measures (Optional)

Definition 24.1. Let \((X, \mathcal{A}, \mu)\) be a measure space. A null set (or \(\mu\)-null set) is a set \(N \in \mathcal{A}\) such that \(\mu(N) = 0\). The completion of \(\mathcal{A}\) with respect to \(\mu\) is the collection \(\overline{\mathcal{A}}\) consisting of all sets of the form \(A \cup B\), where \(A \in \mathcal{A}\) and \(B\) is a subset of a null set.

Observation. According to Definition 23.1, \((X, \mathcal{A}, \mu)\) is complete iff every subset of a null set belongs to \(\mathcal{A}\).

Theorem 24.2. Let \((X, \mathcal{A}, \mu)\) be a measure space.

(a) \(\overline{\mathcal{A}}\) is a \(\sigma\)-algebra on \(X\).

(b) There is a unique extension of \(\mu\) to \(\overline{\mathcal{A}}\) (i.e., there is a one and only one measure on \(\overline{\mathcal{A}}\) that agrees with \(\mu\) on \(\mathcal{A}\)). This extension is called the completion of \(\mu\) and is denoted by \(\overline{\mu}\). It satisfies \(\overline{\mu}(A \cup B) = \mu(A)\) whenever \(A \in \mathcal{A}\) and \(B\) is a subset of a null set.

(c) The measure space \((X, \overline{\mathcal{A}}, \overline{\mu})\) is complete. It is called the completion of \((X, \mathcal{A}, \mu)\).

Theorem 24.3. Let \((X, \mathcal{A}, \mu)\) be a measure space.

(a) \(\overline{\mathcal{A}}\) is the smallest \(\sigma\)-algebra on \(X\) that contains both \(\mathcal{A}\) and all subsets of null sets.

(b) \(\overline{\mathcal{A}}\) is the smallest \(\sigma\)-algebra on \(X\) for which some extension of \(\mu\) is complete (i.e., for which there exists a measure \(\nu\) on \(\overline{\mathcal{A}}\) such that \((X, \overline{\mathcal{A}}, \nu)\) is complete and \(\mu = \nu\) on \(\mathcal{A}\)).

Theorem 24.4. Let \((X, \mathcal{A}, \mu)\) be a \(\sigma\)-finite measure space. Let \(\mu^*\) be the outer measure generated by \(\mu\) (as in Theorem 20.3 with \(\mu_0\) replaced by \(\mu\)). Then \(M(\mu^*) = \overline{\mathcal{A}}, \mu^*|_{M(\mu^*)} = \overline{\mu}\), and \((X, M(\mu^*), \mu^*|_{M(\mu^*)})\) is the completion of \((X, \mathcal{A}, \mu)\).

Remark: Exploring what happens in the non-\(\sigma\)-finite case is an exercise left to the reader.

Corollary 24.5. \((\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)\) is the completion of \((\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)\).

Corollary 24.6. Let \((X, \mathcal{S}, \mu)\) and \((Y, \mathcal{T}, \nu)\) be \(\sigma\)-finite measure spaces. Let \(\pi^*\) be the outer measure on \(X \times Y\) generated by \(\mu\) and \(\nu\) (as in Theorem 28.3). Then \((X, M(\pi^*), \mu \otimes \nu)\) is the completion of \((X, \mathcal{S} \otimes \mathcal{T}, \mu \otimes \nu)\).

Theorem 24.7. Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((X, \overline{\mathcal{A}}, \overline{\mu})\) be its completion.

(a) If \(f : X \to \mathbb{R}\) is \(\mathcal{A}\)-measurable, then \(f\) is \(\overline{\mathcal{A}}\)-measurable.

(b) If \(f : X \to \mathbb{R}\) is \(\overline{\mathcal{A}}\)-measurable, then there is a \(\mathcal{A}\)-measurable function \(g : X \to \mathbb{R}\) such that \(f = g\) \(\overline{\mu}\)-a.e.

Proof Outline. (a): Use \(\mathcal{A} \subseteq \overline{\mathcal{A}}\) and the definition of measurability.

(b): Prove it first for \(f\) simple. For general \(f\), consider a sequence of simple functions converging pointwise to \(f\). 

77
25 Riemann Integral

Definition 25.1. Let \( f : [a, b] \to \mathbb{R} \) be a bounded function. A partition of \([a, b]\) is a finite set \( P = \{x_0, x_1, \ldots, x_n\} \) such that \( a = x_0 < x_1 < \cdots < x_n = b \). For every partition \( P = \{x_0, \ldots, x_n\} \) of \([a, b]\), the lower Riemann sum for \( f \) and \( P \) is

\[
L(f, P) = \sum_{i=1}^{n} (x_i - x_{i-1}) \inf \{ f(x) : x \in [x_{i-1}, x_i] \}
\]

and the upper Riemann sum for \( f \) and \( P \) is

\[
U(f, P) = \sum_{i=1}^{n} (x_i - x_{i-1}) \sup \{ f(x) : x \in [x_{i-1}, x_i] \}.
\]

The lower Riemann integral of \( f \) on \([a, b]\) is

\[
\int_{a}^{b} f = \sup \{ L(f, P) : P \text{ is a partition of } [a, b] \}
\]

and the upper Riemann integral of \( f \) on \([a, b]\) is

\[
\overline{\int}_{a}^{b} f = \sup \{ U(f, P) : P \text{ is a partition of } [a, b] \}.
\]

If \( \int_{a}^{b} f = \overline{\int}_{a}^{b} f \), we define the Riemann integral of \( f \) on \([a, b]\) to be \( \int_{a}^{b} f = \int_{a}^{b} f = \overline{\int}_{a}^{b} f \), and we say \( f \) is Riemann integrable on \([a, b]\).

Lemma 25.2. Let \( f : [a, b] \to \mathbb{R} \) be a bounded function. If \( P_1 \) and \( P_2 \) are partitions of \([a, b]\) such that \( P_1 \subseteq P_2 \), then

\[
L(f, P_1) \leq L(f, P_2) \leq U(f, P_2) \leq U(f, P_1).
\]

Proof. First inequality: Consider \( P_2 \) with just one more point that \( P_1 \) and compare sums; repeat for larger \( P_2 \). Third inequality: Similar to first. Middle inequality: The infimum of a set is less than or equal to the supremum of the set. \( \square \)
26 Comparison of Lebesgue Integral and Riemann Integral

In this section, \((X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)\). If \(f : [a, b] \to \mathbb{R}\) is measurable, then the Lebesgue integral of \(f\) on \([a, b]\) is \(\int_{[a, b]} f \, d\lambda = \int_{[a, b]} \hat{f} \, d\lambda\) where \(\hat{f} : \mathbb{R} \to \mathbb{R}\) is defined by \(\hat{f} = f\) on \([a, b]\) and \(\hat{f} = 0\) on \(\mathbb{R} \setminus [a, b]\). It is easy to check \(f\) is measurable iff \(\hat{f}\) is measurable.

**Theorem 26.1.** Suppose \(f : [a, b] \to \mathbb{R}\) is bounded on \([a, b]\).

(a) If \(f\) is Riemann integrable on \([a, b]\), then \(f\) is Lebesgue measurable, \(f\) is Lebesgue integrable, and the Riemann and Lebesgue integrals of \(f\) on \([a, b]\) are equal.

(b) \(f\) is Riemann integrable on \([a, b]\) iff \(f\) is continuous at almost every point of \([a, b]\).

**Proof.** We use the notation \(\underline{\int} f\) for the lower Riemann integral of \(f\) on \([a, b]\), \(\overline{\int} f\) for the upper Riemann integral of \(f\) on \([a, b]\), and \(\int f \, d\lambda\) for the Lebesgue integral on \([a, b]\). By the definitions of \(\underline{\int} f\) and \(\overline{\int} f\), we can choose sequences \((P_n')\), \((P_n'')\) of partitions of \([a, b]\) such that

\[
\lim_{n} L(f, P_n') = \underline{\int} f \quad \text{and} \quad \lim_{n} U(f, P_n'') = \overline{\int} f.
\]

Define \(P_n''' = \{a + k(b - a)/n : k = 0, 1, \ldots, n\}\). Define \(P_n = \bigcup_{m=1}^{n} (P_n' \cup P_n'' \cup P_m''')\). Note that \(P_n' \subseteq P_n\), \(P_n'' \subseteq P_n\), and \(P_n \subseteq P_{n+1}\). Note also that, if \(P_n = \{x_0, x_1, \ldots, x_k\}\), then \(0 \leq x_i - x_{i-1} \leq 1/n\) for all \(i = 0, 1, \ldots, k\). Lemma 25.2 implies

\[
L(f, P_n') \leq L(f, P_n) \leq L(f, P_{n+1}) \leq \underline{\int} f \quad \text{and} \quad U(f, P_n') \geq U(f, P_n) \geq U(f, P_{n+1}) \geq \overline{\int} f.
\]

It follows that

\[
\lim_{n} L(f, P_n) = \underline{\int} f \quad \text{and} \quad \lim_{n} U(f, P_n) = \overline{\int} f.
\]

For each \(n\), write \(P_n = \{x_0, \ldots, x_k\}\), and define

\[
s_n = f(x_0)1_{\{x_0\}} + \sum_{i=1}^{k} \inf \{f(x) : x \in [x_{i-1}, x_i]\} 1_{(x_{i-1}, x_i]},
\]

\[
t_n = f(x_0)1_{\{x_0\}} + \sum_{i=1}^{k} \sup \{f(x) : x \in [x_{i-1}, x_i]\} 1_{(x_{i-1}, x_i]}.
\]

Then \(s_n\) and \(t_n\) are measurable, and

\[
\int s_n \, d\lambda = L(f, P_n) \quad \text{and} \quad \int t_n \, d\lambda = U(f, P_n).
\]

Therefore

\[
\lim_{n} \int s_n \, d\lambda = \underline{\int} f \quad \text{and} \quad \lim_{n} \int t_n \, d\lambda = \overline{\int} f.
\]

Note

\[
s_n \leq s_{n+1} \leq f \leq t_{n+1} \leq t_n.
\]

79
Define \( s = \sup_n s_n \) and \( t = \inf_n t_n \). Then \( s \) and \( t \) are measurable and

\[
s_1 \leq s \leq f \leq t \leq t_1. \tag{26.1}
\]

Since \( s_1 \) and \( t_1 \) are finite everywhere, \( s \) and \( t \) are finite everywhere. Moreover, since \( \int s_1 \) and \( \int t_1 \) are finite, \( \int sd\lambda \) and \( \int td\lambda \) are finite, in other words, \( s \) and \( t \) are Lebesgue integrable. Since \( 0 \leq s_n - s_1 \uparrow s - s_1 \), the monotone convergence theorem implies

\[
\int (s - s_1)d\lambda = \lim_n \int (s_n - s_1)d\lambda.
\]

Adding \( \int s_1 \) gives

\[
\int sd\lambda = \lim_n \int s_n d\lambda = \int f. \tag{26.2}
\]

A similar argument gives

\[
\int td\lambda = \lim_n \int t_n d\lambda = \bar{\int} f. \tag{26.3}
\]

Since \( s \leq t \), we see from (26.2) and (26.3) that \( f \) is Riemann integrable on \([a, b]\) iff \( \int f = \bar{\int} f \) iff \( \int (t - s)d\lambda = 0 \) iff \( s = t \) a.e.

Now we prove (a). Assume \( f \) is Riemann integrable on \([a, b]\). By the observation above, \( s = t \) a.e. By (26.1), we have \( s \leq f \leq t \). Therefore \( s = f = t \) a.e. Since \( s \) and \( t \) are measurable and since the Lebesgue measure is complete, Theorem 23.2 implies \( f \) is measurable. Moreover,

\[
\int fd\lambda = \int sd\lambda = \int td\lambda = \int f = \bar{\int} f.
\]

So \( f \) is Lebesgue integrable and the Riemann and Lebesgue integrals are equal.

Now we prove (b). As we observed above, \( f \) is Riemann integrable on \([a, b]\) iff \( s(x) = t(x) \) for a.e. \( x \in [a, b] \). Since \( \bigcup_{n=1}^{\infty} P_n \) is a countable subset of \([a, b]\), it has Lebesgue measure 0. Thus \( f \) is Riemann integrable on \([a, b]\) iff \( s(x) = t(x) \) for a.e. \( x \in [a, b] \setminus \bigcup_{n=1}^{\infty} P_n \). Therefore, it will suffice to show that, for every \( x \in [a, b] \setminus \bigcup_{n=1}^{\infty} P_n \), \( s(x) = t(x) \) iff \( f \) is continuous at \( x \). Let \( x \in [a, b] \setminus \bigcup_{n=1}^{\infty} P_n \) be given.

Assume \( s(x) = t(x) \). Then \( \lim_n (t_n(x) - s_n(x)) = 0 \). Let \( \epsilon > 0 \). Choose \( n \) such that \( t_n(x) - s_n(x) < \epsilon \). Write \( P_n = \{x_0, x_1, \ldots, x_k\} \). Then \( x \in (x_{j-1}, x_j) \) for some \( j \). Choose \( \delta > 0 \) such that \( (x - \delta, x + \delta) \subseteq (x_{j-1}, x_j) \). Let \( y \in [a, b] \) with \( |x - y| < \delta \). Then \( x, y \in (x_{j-1}, x_j) \). Therefore

\[
|f(x) - f(y)| \leq \sup \{f(x) : x \in [x_{i-1}, x_i]\} - \inf \{f(x) : x \in [x_{i-1}, x_i]\}
= \sup \{f(x) : x \in [x_{i-1}, x_i]\} - \inf \{f(x) : x \in [x_{i-1}, x_i]\} 1_{(x_{j-1}, x_j)}(x)
= \sum_{i=1}^{k} (\sup \{f(x) : x \in [x_{i-1}, x_i]\} - \inf \{f(x) : x \in [x_{i-1}, x_i]\}) 1_{(x_{j-1}, x_j)}(x)
= t_n(x) - s_n(x) < \epsilon.
\]

Therefore \( f \) is continuous at \( x \).
Conversely, now we assume \( f \) is continuous at \( x \) and deduce that \( s(x) = t(x) \). Let \( \epsilon > 0 \). Choose \( \delta > 0 \) such that, for all \( y \in [a, b] \), if \( |x - y| < \delta \), then \( |f(x) - f(y)| < \epsilon \). Choose \( N \in \mathbb{N} \) such that \( 1/N < \delta \). Let \( n \geq N \) be arbitrary. Write \( P_n = \{x_1, x_0, \ldots, x_k\} \). Then \( x \in (x_{j-1}, x_j) \) for some \( 1 \leq j \leq k \). Recall, from the definition of \( P_n \), that \( 0 \leq x_i - x_{i-1} \leq 1/n \) for all \( 1 \leq i \leq k \). So, for every \( y \in [x_{j-1}, x_j] \), we have \( 0 \leq |y - x| \leq x_j - x_{j-1} \leq 1/n \leq 1/N < \delta \), hence \( |f(y) - f(x)| < \epsilon \).

Thus, for every \( y_1, y_2 \in [x_{j-1}, x_j] \) we have

\[
|f(y_1) - f(y_2)| \leq |f(y_1) - f(x)| + |f(y_2) - f(x)| \leq 2\epsilon.
\]

It follows that

\[
t_n(x) - s_n(x) = \sum_{i=1}^{k} (\sup \{f(x) : x \in [x_{i-1}, x_i]\} - \inf \{f(x) : x \in [x_{i-1}, x_i]\}) 1_{(x_{j-1}, x_j)}(x)
\]

\[
= \sup \{f(x) : x \in [x_{i-1}, x_i]\} - \inf \{f(x) : x \in [x_{i-1}, x_i]\} \leq 2\epsilon.
\]

Thus

\[
t(x) - s(x) = \lim_{n} (t_n(x) - s_n(x)) \leq 2\epsilon.
\]

Letting \( \epsilon \to 0 \) gives \( s(x) = t(x) \).
Chapter 8

Product Measures and Fubini’s Theorem

The goal of this chapter is to find sufficient conditions for changing the order of iterated integrals. Given measure spaces \((X, \mathcal{S}, \mu)\) and \((Y, \mathcal{T}, \nu)\) and a function \(f : X \times Y \rightarrow \mathbb{R}\), we will derive conditions under which the following formula is valid:

\[
\int \left( \int f(x, y) \, d\mu(x) \right) \, d\nu(y) = \int \left( \int f(x, y) \, d\nu(y) \right) \, d\nu(x)
\]

27 Measurable Rectangles and Product \(\sigma\)-Algebras

Let \(X\) and \(Y\) be sets.

**Notation.** If \(S \subseteq \mathcal{P}(X)\) and \(T \subseteq \mathcal{P}(Y)\), we define

\[
S \times T = \{A \times B : A \in S, B \in T\}
\]

Note that this is not the usual Cartesian product of \(S\) and \(T\).

**Remark.** If \(S\) and \(T\) are \(\sigma\)-algebras, then \(S \times T\) need not be a \(\sigma\)-algebra (see Exercise N).

**Definition 27.1.** Let \(S\) be a \(\sigma\)-algebra on \(X\) and \(T\) be a \(\sigma\)-algebra on \(Y\).

(a) The sets in \(S \times T\) are called **measurable rectangles**.

(b) The \(\sigma\)-algebra on \(X \times Y\) generated by \(S \times T\) is called the **product \(\sigma\)-algebra** of \(S\) and \(T\). It is denoted by \(S \otimes T\). In other words, \(S \otimes T = \sigma(S \times T)\).

**Definition 27.2.** Let \(E \subseteq X \times Y\).

(a) For each \(x_0 \in X\), the **\(x_0\)-section** of \(E\) is the set \(E_{x_0} = \{y \in Y : (x_0, y) \in E\} \subseteq Y\).

(b) For each \(y_0 \in Y\), the **\(y_0\)-section** of \(E\) is the set \(E^{y_0} = \{x \in X : (x, y_0) \in E\} \subseteq X\).
Example 27.3.

(a) If \( E = [1, 2] \times [3, 4] \subseteq \mathbb{R} \times \mathbb{R} \), then \( E_x = [3, 4] \) for every \( x \in [1, 2] \) and \( E_x = \emptyset \) otherwise.

(b) If \( E = \{(x, y) : x^2 + y^2 = 1\} \subseteq \mathbb{R} \times \mathbb{R} \), then \( E_0 = \{1, -1\} \) and \( E_1 = \{0\} \) and \( E_2 = \emptyset \).

Definition 27.4. Let \( f : X \times Y \to \mathbb{R} \).

(a) For each \( x_0 \in X \), the \( x_0 \)-section of \( f \) is the function \( f_{x_0} : Y \to \mathbb{R} \) defined by \( f_{x_0}(y) = f(x_0, y) \) for every \( x_0 \in X \) and \( y_0 \in Y \).

(b) For each \( y_0 \in Y \), the \( y_0 \)-section of \( f \) is the function \( f_{y_0} : X \to \mathbb{R} \) defined by \( f_{y_0}(x) = f(x, y_0) \) for every \( x_0 \in X \) and \( y_0 \in Y \).

Example 27.5.

(a) If \( f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \) is defined by \( f(x, y) = 5x^2 + y^3 \), then \( f_{2}(y) = f(2, y) = 20 + y^3 \).

(b) For any \( E \subseteq X \times Y \) and \( x_0 \in X \) and \( y_0 \in Y \), we have \((1_E)_{x_0} = 1_{E_{x_0}}\) and \((1_E)^{y_0} = 1_{E_{y_0}}\).

Theorem 27.6. Let \( S \) be a \( \sigma \)-algebra on \( X \) and let \( T \) be a \( \sigma \)-algebra on \( Y \).

(a) If \( E \in S \otimes T \), then \( E_{x_0} \in T \) for each \( x_0 \in X \) and \( E_{y_0} \in S \) for each \( y_0 \in Y \).

(b) If \( f : X \times Y \to \mathbb{R} \) is \( S \otimes T \)-measurable, then \( f_{x_0} \) is \( T \)-measurable for each \( x_0 \in X \) and \( f_{y_0} \) is \( S \)-measurable for each \( y_0 \in Y \).

Proof.

(a) We must show

\[ S \otimes T \subseteq \{E \subseteq X \times Y : E_{x_0} \in T \text{ for all } x_0 \in X \text{ and } E_{y_0} \in S \text{ for all } y_0 \in Y \} \cdot \]

It is easy to check that the right-hand side is a \( \sigma \)-algebra containing the collection of measurable rectangles. As \( S \otimes T \) is the smallest \( \sigma \)-algebra containing the collection of measurable rectangles, the desired containment follows.

(b) Let \( x_0 \in X \). For every Borel set \( B \subseteq \mathbb{R} \), we have \( f^{-1}(B) \in S \otimes T \), and so

\[ (f_{x_0})^{-1}(B) = \left[ f^{-1}(B) \right]_{x_0} \in T. \]

This proves that \( f_{x_0} \) is \( T \)-measurable for each \( x_0 \in X \). The other conclusion is proved similarly.
28 Product Measures

If $S$ and $T$ are $\sigma$-algebras, then $S \times T$ need not be a $\sigma$-algebra (see Exercise N). However, we do have the following lemma.

**Lemma 28.1.** If $E$ is a demi-ring on a set $X$ and $F$ is a demi-ring on a set $Y$, then $E \times F$ is a demi-ring on $X \times Y$.

**Proof.** First note that $\emptyset = \emptyset \times \emptyset \in E \times F$. Now let $A_1 \times A_2$ and $B_1 \times B_2$ be elements of $E \times F$. Then $A_1 \setminus B_1 = \bigcup_{i=1}^{n_1} C_{1,i}$ for some disjoint sets $C_{1,1}, \ldots, C_{1,n_1} \in E$. Also $A_2 \setminus B_2 = \bigcup_{i=1}^{n_2} C_{2,i}$ for some disjoint sets $C_{2,1}, \ldots, C_{2,n_2} \in E$. Define $C_{1,0} = B_1$ and $C_{2,0} = B_2$. Then $A_1 = \bigcup_{i=0}^{n_1} C_{1,i}$ and $C_{1,0}, \ldots, C_{1,n_1}$ are disjoint. Also $A_2 = \bigcup_{i=0}^{n_2} C_{2,i}$ and $C_{2,0}, \ldots, C_{2,n_2}$ are disjoint. Therefore

$$A_1 \times A_2 = \bigcup_{i_1=0}^{n_1} \bigcup_{i_2=0}^{n_2} (C_{1,i_1} \times C_{2,i_2}).$$

Moreover, the sets $C_{1,i_1} \times C_{2,i_2}$ are disjoint sets in $E \times F$ for all pairs $(i_1, i_2)$. Indeed, if $(i_1, i_2) \neq (j_1, j_2)$, then $i_1 \neq j_1$ or $i_2 \neq j_2$, so $C_{1,i_1} \cap C_{1,j_1} = \emptyset$ or $C_{2,i_2} \cap C_{2,j_2} = \emptyset$, and hence

$$(C_{1,i_1} \times C_{2,i_2}) \cap (C_{1,j_1} \times C_{2,j_2}) = (C_{1,i_1} \cap C_{1,j_1}) \times (C_{2,i_2} \cap C_{2,j_2}) = \emptyset.$$

Therefore

$$(A_1 \times A_2) \setminus (B_1 \times B_2) = \bigcup_{i_1=0}^{n_1} \bigcup_{i_2=0}^{n_2} ((C_{1,i_1} \times C_{2,i_2}) \setminus (C_{1,0} \times C_{2,0})) = \bigcup_{i_1=1}^{n_1} \bigcup_{i_2=1}^{n_2} (C_{1,i_1} \times C_{2,i_2}).$$

\[\square\]

**Corollary 28.2.** If $S$ is a $\sigma$-algebra on a set $X$ and $T$ is a $\sigma$-algebra on a set $Y$, then $S \times T$ is a demi-ring on $X \times Y$.

**Theorem 28.3.** Let $(X, S, \mu)$ and $(Y, T, \nu)$ be measure spaces. Define $\pi_0 : S \times T \to [0, \infty]$ by

$$\pi_0(A \times B) = \mu(A)\nu(B)$$

for each $A \times B \in S \times T$. Let $\pi^*$ be the outer measure generated by $\pi_0$.

(a) $\pi_0$ is finitely additive and countably monotone.

(b) $S \times T \subseteq M(\pi^*)$ and $\pi^* = \pi_0$ on $S \times T$.

(c) Let $A$ be a $\sigma$-algebra on $X \times Y$ such that $S \times T \subseteq A \subseteq M(\pi^*)$. Define $\mu \otimes \nu = \pi^*|_A$ (i.e., $\mu \otimes \nu$ is the restriction of $\pi^*$ to $A$). Then $\mu \otimes \nu$ is a measure on $A$ and

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B) \quad \text{for all } A \times B \in S \times T. \quad (28.1)$$

Moreover, if $A = M(\pi^*)$, then $\mu \otimes \nu$ is complete. We call $\mu \otimes \nu$ the **product measure** of $\mu$ and $\nu$ on $A$.

(d) Under the assumptions of (c), if $X$ is $\sigma$-finite with respect to $\mu$ and $Y$ is $\sigma$-finite with respect to $\nu$, then $X \times Y$ is $\sigma$-finite with respect to $\pi_0$ and $\mu \otimes \nu$ is the unique measure on $A$ that satisfies (28.1).
Proof.

(a): We first prove that $\pi_0$ is countably additive. Let $A_1 \times B_1, A_2 \times B_2, \ldots \in S \times T$ be disjoint sets. Let $A \times B \in S \times T$ such that $A \times B = \bigcup_{i=1}^\infty (A_i \times B_i)$. For every $x \in X$ and $y \in Y$,

$$1_A(x)1_B(y) = 1_{A \times B}(x,y) = \sum_{i=1}^\infty 1_{A_i \times B_i}(x,y) = \sum_{i=1}^\infty 1_{A_i}(x)1_{B_i}(y).$$

For fixed $y \in Y$, integrating in $x$ gives

$$\mu(A)1_B(y) = \int 1_A(x)1_B(y)\,d\mu(x) = \int \sum_{i=1}^\infty 1_{A_i}(x)1_{B_i}(y)\,d\mu(x).$$

By applying the monotone convergence theorem to the partial sums of the last series, we obtain

$$\mu(A)1_B(y) = \sum_{i=1}^\infty \int 1_{A_i}(x)1_{B_i}(y)\,d\mu(x) = \sum_{i=1}^\infty \mu(A_i)1_{B_i}(y).$$

Now integrating in $y$ gives

$$\mu(A)\nu(B) = \int \mu(A)1_B(y)\,d\nu(y) = \int \sum_{i=1}^\infty \mu(A_i)1_{B_i}(y)\,d\nu(y).$$

By applying the monotone convergence theorem to the partial sums of the last series, we obtain

$$\mu(A)\nu(B) = \sum_{i=1}^\infty \int \mu(A_i)1_{B_i}(y)\,d\nu(y) = \sum_{i=1}^\infty \mu(A_i)\nu(B_i).$$

Hence $\pi_0(A \times B) = \sum_{i=1}^\infty \pi_0(A_i \times B_i)$. Thus $\pi_0$ is countably additive. It follows immediately that $\pi_0$ is finitely additive. To see that $\pi_0$ is countably monotone, we can appeal to HW8 Exercise 4(c). Alternatively, we can modify the argument above as follows. We drop the condition that the sets $A_i \times B_i$ are disjoint and require only that $A \times B \subseteq \bigcup_{i=1}^\infty (A_i \times B_i)$. Then the second $=$ in the first displayed equation above becomes $\leq$, and in all subsequent equations the first $=$ becomes $\leq$.

(b): Combine (a), Lemma 28.1, and Theorem 20.3.

(c): Combine (a), (b), Theorem 21.1(a), and Theorem 23.3.

(d): Since $X$ is $\sigma$-finite with respect to $\mu$, there exists $A_1, A_2, \ldots \in S$ such that $X = \bigcup_{i=1}^\infty A_i$ and $\mu(A_i) < \infty$ for all $i$. Since $Y$ is $\sigma$-finite with respect to $\nu$, there exists $B_1, B_2, \ldots \in T$ such that $Y = \bigcup_{j=1}^\infty B_j$ and $\mu(B_j) < \infty$ for all $j$. Then

$$X \times Y = \left( \bigcup_{i=1}^\infty A_i \right) \times \left( \bigcup_{j=1}^\infty B_j \right) = \bigcup_{i=1}^\infty \bigcup_{j=1}^\infty (A_i \times B_j),$$

$A_i \times B_j \in S \times T$ and $\pi_0(A_i \times B_j) = \mu(A_i)\nu(B_j) < \infty$ for all $i,j$. So $X \times Y$ is $\sigma$-finite with respect to $\pi_0$. Now Theorem 21.1(c) implies the result. \qed

85
29  Fubini’s Theorem

Let \((X, \mathcal{S}, \mu)\) and \((Y, \mathcal{T}, \nu)\) be complete measure spaces. The reader should not overlook the completeness assumptions. Let \(\mathcal{A}\) be a \(\sigma\)-algebra on \(X \times Y\) such that \(S \times T \subseteq \mathcal{A} \subseteq M(\pi^*)\). Let \(\mu \otimes \nu\) be the product of \(\mu\) and \(\nu\)
on \(\mathcal{A}\).

**Lemma 29.1.** Let \(E \in (S \times T)_{\sigma\delta}\) with \(\mu \otimes \nu(E) < \infty\).

(a) The function \(g(y) = \mu(E_y)\) is \(\mathcal{T}\)-measurable

(b) The function \(h(x) = \mu(E_x)\) is \(\mathcal{S}\)-measurable.

(c) \(\int \mu(E_y) d\nu(y) = \mu \otimes \nu(E) = \int \nu(E_x) d\mu(x)\)

**Proof Outline.** Verify the theorem for measurable rectangles directly. Then use countable additivity, the monotone convergence theorem and the fact that limits of measurable functions are measurable to extend the theorem to countable unions of measurable rectangles. Finally, use continuity from above and the dominated convergence theorem to extend the theorem to countable intersections of countable unions of measurable rectangles.

**Proof.** The proof that \(g(y) = \mu(E_y)\) is \(\mathcal{T}\)-measurable is symmetric with the proof that \(h(x) = \mu(E_x)\) is \(\mathcal{S}\)-measurable. Thus we omit the latter. Similarly, the proofs of the first equality and second equality in (c) are virtually identical, so we omit the latter.

First note that Theorem 27.6 implies \(E_y \in \mathcal{S}\) for every \(y \in Y\). We consider three cases. We only use the completeness of \((X, \mathcal{S}, \mu)\) and \((Y, \mathcal{T}, \nu)\) in the third case.

**Case 1:** \(E = A \times B \in \mathcal{S} \times \mathcal{T}\).

For each \(y \in Y\), we have \(E_y = A\) if \(y \in B\) and \(E_y = \emptyset\) if \(y \notin B\), hence \(\mu(E_y) = \mu(A)1_B(y)\). Therefore, since \(B \in \mathcal{T}\), the function \(g(y) = \mu(E_y)\) is \(\mathcal{T}\)-measurable. Moreover,

\[
\int \mu(E_y) d\nu(y) = \int \mu(A)1_B(y) d\nu(y) = \mu(A)\nu(B) = \mu \otimes \nu(A \times B) = \mu \otimes \nu(E)
\]

**Case 2:** \(E \in (\mathcal{S} \times \mathcal{T})_\sigma\).

Then \(E = \bigcup_{i=1}^\infty E_i\) for some sets \(E_i = A_i \times B_i \in \mathcal{S} \times \mathcal{T}\). Since \(\mathcal{S} \times \mathcal{T}\) is a demi-ring, we can use Lemma 20.2(a) and relabel to ensure that the sets \(E_i\) are disjoint (as in HW8 Exercise 4(b)). Then, for each \(y \in Y\), the sets \(E_i^y\) are disjoint, \(E^y = \bigcup_{i=1}^\infty E_i^y\), and

\[
\mu(E^y) = \sum_{i=1}^\infty \mu(E_i^y) = \lim_n \sum_{i=1}^n \mu(E_i^y).
\]

By Case 1, \(g_i(y) = \mu(E_i^y)\) is \(\mathcal{T}\)-measurable for each \(i\). Therefore \(g(y) = \mu(E_y)\) is \(\mathcal{T}\)-measurable because \(g = \sum_{i=1}^\infty g_i = \lim_n \sum_{i=1}^n g_i\). Moreover,

\[
\int \mu(E^y) d\nu(y) = \int \sum_{i=1}^\infty \mu(E_i^y) d\nu(y)
\]
By applying the monotone convergence theorem to the partial sums of the series, we obtain
\[ \int \mu(E^y)d\nu(y) = \sum_{i=1}^{\infty} \int \mu(E_i^y)d\nu(y). \]

Now using Case 1, we get
\[ \int \mu(E^y)d\nu(y) = \sum_{i=1}^{\infty} \mu \otimes \nu(E_i) \]

Since \( E = \bigcup_{i=1}^{\infty} E_i \) and the sets \( E_i \) are disjoint, we have
\[ \int \mu(E^y)d\nu(y) = \mu \otimes \nu(E). \]

Case 3: \( E \in (S \times T)_{\sigma \delta}. \)

Then \( E = \bigcap_{n=1}^{\infty} E_n \) for some sets \( E_n \in (S \times T)_{\sigma}. \)

Claim: There exists a decreasing sequence \( (G_n) \) of sets in \( (S \times T)_{\sigma} \) such that \( E = \bigcap_{n=1}^{\infty} G_n \) and \( \mu \otimes \nu(G_n) < \infty \) for every \( n \).

Proof of Claim: By Theorem 20.4, we can find a set \( F \in (S \times T)_{\sigma} \) such that \( E \subseteq F \) and \( \mu \otimes \nu(F) \leq \mu \otimes \nu(E) + 1 \). Define \( F_n = E_n \cap F \) for each \( n \). Then \( E = \bigcap_{n=1}^{\infty} F_n \). Since \( \mu \otimes \nu(F) < \infty \) and \( F_n \subseteq F \), we have \( \mu \otimes \nu(F_n) < \infty \) for each \( n \). By Lemma 20.2(b), \( F_n \in (S \times T)_{\sigma} \) for each \( n \).

Define \( G_n = \bigcap_{k=1}^{n} F_k \) for each \( n \). Then \( (G_n) \) is a decreasing sequence of sets, \( E = \bigcap_{n=1}^{\infty} G_n \), and \( \mu \otimes \nu(G_n) \leq \mu \otimes \nu(F_n) < \infty \) for each \( n \). Using Lemma 20.2(b) again, we can show that \( G_n \in (S \times T)_{\sigma} \) for each \( n \). This completes the proof of the claim.

By Case 2, for every \( n \), the function \( g_n(y) = \mu(G_n^y) \) is \( T \)-measurable and \( \int \mu(G_n^y)d\nu(y) = \mu \otimes \nu(G_n) \).

So, by continuity from above (Theorem 5.2), we have
\[ \mu \otimes \nu(E) = \lim_{n} \mu \otimes \nu(G_n) = \lim_{n} \int \mu(G_n^y)d\nu(y) = \lim_{n} \int g_n(y)d\nu(y). \tag{29.1} \]

On the other hand, we have
\[ \int g_1(y)d\nu(y) = \mu \otimes \nu(G_1) < \infty. \]

Thus \( g_1(y) = \mu(G_1^y) \) is finite for \( \nu \)-a.e. \( y \in Y \). For each \( y \in Y \), \( E^y = \bigcap_{n=1}^{\infty} G_n^y \) and \( (G_n^y) \) is a decreasing sequence of sets in \( S \). For each \( y \in Y \) such that \( g_1(y) = \mu(G_1^y) \) is finite, continuity from above (Theorem 5.2) implies \( \lim_{n} g_n(y) = \lim_{n} \mu(G_n^y) = \lim_{n} \mu(E^y) = g(y) \). Therefore \( g_n \to g \) for \( \nu \)-a.e. \( y \in Y \). Since \( (Y, T, \nu) \) is complete and each \( g_n \) is \( T \)-measurable, Theorem 23.2 implies \( g \) is \( T \)-measurable. In other words, the function \( g(y) = \mu(E^y) \) is \( T \)-measurable. We have observed that \( g_n \to g \) for \( \nu \)-a.e. and \( g_1 \) is \( \nu \)-integrable. We also have \( 0 \leq g_n \leq g_1 \) for all \( n \). Thus the dominated convergence theorem gives
\[ \int \mu(E^y)d\nu(y) = \int g(y)d\nu(y) = \lim_{n} \int g(y)d\nu(y) \]

Combining this with (29.1) gives
\[ \int \mu(E^y)d\nu(y) = \mu \otimes \nu(E). \]
This completes the proof.

Note: Everywhere that we used continuity from above, we could have instead used the dominated convergence theorem applied to indicator functions. Moreover, if we used the dominated convergence theorem, we would only need that $G_n \subseteq G_1$, instead of that $(G_n)$ is decreasing. 

**Lemma 29.2.** Let $E \in \mathcal{A}$ with $\mu \otimes \nu(E) < \infty$.

(a) $E^y \in \mathcal{S}$ for $\nu$-a.e. $y \in Y$. In other words, there is a set $B \in \mathcal{T}$ such that $\nu(B^c) = 0$ and $E^y \in \mathcal{S}$ for every $y \in B$.

(b) $E_x \in \mathcal{T}$ for $\mu$-a.e. $x \in X$. In other words, there is a set $A \in \mathcal{S}$ such that $\mu(A^c) = 0$ and $E_x \in \mathcal{T}$ for every $x \in A$.

(c) There exists a set $B$ as in (a) such that the function $g(y) = \begin{cases} \mu(E^y) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{T}$-measurable.

(d) There exists a set $A$ as in (b) such that the function $h(x) = \begin{cases} \nu(E_x) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{S}$-measurable.

(e) $\int g(y)d\nu(y) = \mu \otimes \nu(E) = \int h(x)d\mu(x)$.

**Proof Outline.** Use Theorem 20.4 to approximate an arbitrary set $E \in \mathcal{A}$ by a set $H \in (\mathcal{S} \otimes \mathcal{T})_{\sigma\delta}$ with the same measure of $E$ and the set $H \setminus E \in \mathcal{A}$ which is contained in a set $H_0 \in (\mathcal{S} \otimes \mathcal{T})_{\sigma\delta}$ having measure 0. Lemma 29.1 says the theorem holds for $H$ and the completeness of $\mu$ and $\nu$ combined with Lemma 29.1 imply that the theorem holds for $E \setminus H$. The additivity of measures and the fact that sums of measurable functions are measurable then completes the proof.

**Proof.** The proofs of (a) and (c) are symmetric with the proofs of (b) and (d). Thus we omit the latter. Similarly, the proofs of the first equality and second equality in (e) are virtually identical, so we omit the latter.

By Theorem 20.4, there is a set $H \in (\mathcal{S} \otimes \mathcal{T})_{\sigma\delta}$ such that $E \subseteq H$ and $\mu \otimes \nu(H) = \mu \otimes \nu(E) < \infty$. Then $H \setminus E \in \mathcal{A}$ and (by Theorem 5.2) $\mu \otimes \nu(H \setminus E) = 0$.

By applying Theorem 20.4 to $H \setminus E$, there is a set $H_0 \in (\mathcal{S} \times \mathcal{T})_{\sigma\delta}$ such that $H \setminus E \subseteq H_0$ and $\mu \otimes \nu(H_0) = \mu \otimes \nu(H \setminus E) = 0$. By Theorem 27.6, $H_0^y \in \mathcal{S}$ for every $y \in Y$. By Lemma 29.1, the function $g_0(y) = \mu(H_0^y)$ is $\mathcal{T}$-measurable and

$$\int \mu(H_0^y)d\nu(y) = \mu \otimes \nu(H_0) = 0.$$ 

Thus $\mu(H_0^y) = 0$ for $\nu$-a.e. $y \in Y$. Since $(X,\mathcal{S},\mu)$ is complete and $(E \setminus H)^y \subseteq H_0^y$, we have $(E \setminus H)^y \in \mathcal{S}$ and $\mu((E \setminus H)^y) = 0$ for $\nu$-a.e. $y \in Y$. By Theorem 27.6, $H^y \in \mathcal{S}$ for every $y \in Y$. Then, since $E^y = H^y \setminus (H \setminus E)^y$ (verify this), we have $E^y \in \mathcal{S}$ for $\nu$-a.e. $y \in Y$. This proves (a).

Let $B \in \mathcal{T}$ with $\nu(B^c) = 0$ and $E^y \in \mathcal{S}$ for every $y \in B$. Define $g$ as in (c). By Lemma 29.1, the function $g_H(y) = \mu(H^y)$ is $\mathcal{T}$-measurable. For each $y \in Y$, we have $H^y = E^y \cup (H \setminus E)^y$. Therefore,
for each $y \in B$, the countable additivity of $\mu$ gives

$$g_H(y) = \mu(H^y) = \mu(E^y) + \mu((H \setminus E^y)) = \mu(E^y) = g(y)$$

Thus $g_H(y) = g(y)$ for $\nu$-a.e. $y \in Y$. Since $g_H$ is $\mathcal{T}$-measurable and since $(Y, \mathcal{T}, \nu)$ is complete, Theorem 23.2 implies $g$ is $\mathcal{T}$-measurable. This proves (c).

Since $g_H(y) = g(y)$ for $\nu$-a.e. $y \in Y$, integrating gives

$$\int \mu(H^y) d\nu(y) = \int g_H(y) d\nu(y) = \int g(y) d\nu(y)$$

By Lemma 29.1,

$$\mu \otimes \nu(H) = \int \mu(H^y) d\nu(y).$$

Finally, recall that $\mu \otimes \nu(E) = \mu \otimes \nu(H)$. Putting everything together, we see that

$$\mu \otimes \nu(E) = \int g(y) d\nu(y).$$

This completes the proof.

**Theorem 29.3.** (Tonelli’s Theorem). Let $f : X \times Y \to \mathbb{R}$ be $\mathcal{A}$-measurable. Assume $f \geq 0$ and $\{f > 0\}$ is $\sigma$-finite with respect to $\mu \otimes \nu$.

(a) $f^y$ is $\mathcal{S}$-measurable for $\nu$-a.e. $y \in Y$. In other words, there is a set $B \in \mathcal{T}$ such that $\nu(B^c) = 0$ and $f^y$ is $\mathcal{T}$-measurable for every $y \in B$.

(b) $f_x$ is $\mathcal{T}$-measurable for $\mu$-a.e. $x \in X$. In other words, there is a set $A \in \mathcal{S}$ such that $\mu(A^c) = 0$ and $f_x$ is $\mathcal{S}$-measurable for every $x \in A$.

(c) There exists a set $B$ as in (a) such that the function $g(y) = \begin{cases} \int f^y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{T}$-measurable.

(d) There exists a set $A$ as in (b) such that the function $h(x) = \begin{cases} \int f_x(y) d\nu(y) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{S}$-measurable.

(e) $\int g(y) d\nu(y) = \int f(x, y) d(\mu \otimes \nu)(x, y) = \int h(x) d\mu(x)$.

**Proof Outline.** Lemma 29.2 implies the theorem holds for indicator functions. Then the linearity of the integral and the fact that linear combinations of measurable functions are measurable implies the theorem holds for non-negative simple functions. Finally, the simple approximation theorem, the monotone convergence theorem, and the fact limits of measurable functions are measurable implies the theorem holds for non-negative functions.

**Proof.**

The proofs of (a) and (c) are symmetric with the proofs of (b) and (d). Thus we omit the latter. Similarly, the proofs of the first equality and second equality in (e) are virtually identical, so we omit the latter.
Case 1: $f = 1_E$, $E \in \mathcal{A}$, $\mu \otimes \nu(E) < \infty$

This is Lemma 29.2.

Case 2: $f$ is a $\mathcal{A}$-measurable simple function such that $\mu \otimes \nu(\{f > 0\}) < \infty$.

Then $f = \sum_{i=1}^{n} c_i 1_{E_i}$, where $c_i \in (0, \infty)$, $E_i = f^{-1}(\{c_i\}) \in \mathcal{S} \otimes \mathcal{T}$, and $\mu \otimes \nu(E_i) < \infty$. Note that, for every $y \in Y$, $f^y = \sum_{i=1}^{n} c_i 1_{E_i}^y$. By Case 1, for each $i \in \{1, \ldots, n\}$, there is a set $B_i \in \mathcal{T}$ such that $\nu(B_i^c) = 0$, $1_{E_i}^y$ is $\mathcal{S}$-measurable for every $y \in B_i$, and the function

$$g_i(y) = \begin{cases} \int 1_{E_i}^y(x) d\mu(x) & \text{if } y \in B_i \\ 0 & \text{otherwise} \end{cases}$$

is $\mathcal{T}$-measurable. Let $B = \bigcap_{i=1}^{n} B_i$. Then $B \in \mathcal{T}$ with $\nu(B^c) = 0$, $1_{E_i}^y$ is $\mathcal{S}$-measurable for every $y \in B$ and every $i \in \{1, \ldots, n\}$, and the function

$$g_i(y) 1_B(y) = \begin{cases} \int 1_{E_i}^y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$$

is $\mathcal{T}$-measurable for every $i \in \{1, \ldots, n\}$. Therefore $f^y = \sum_{i=1}^{n} c_i 1_{E_i}^y$ is $\mathcal{S}$-measurable for every $y \in B$ and the function

$$g(y) = \begin{cases} \int f^y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \int \sum_{i=1}^{n} c_i 1_{E_i}^y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases} = \sum_{i=1}^{n} c_i g_i(y) 1_B(y)$$

is $\mathcal{T}$-measurable. Also by Case 1, we have that $\int g_i(y) d\nu(y) = \mu \otimes \nu(E_i)$. Notice that $g_i = g_i 1_B$ for $\nu$-a.e. $y \in Y$. So $\int g_i(y) 1_B(y) d\nu(y) = \int 1_{E_i}(x, y) = d(\mu \otimes \nu)(x, y) = \mu \otimes \nu(E_i)$. Therefore

$$\int g(y) d\nu(y) = \sum_{i=1}^{n} c_i \int g_i(y) 1_B(y) d\nu(y) = \sum_{i=1}^{n} c_i \mu \otimes \nu(E_i) = \int f(x, y) d\mu \otimes \nu(x, y).$$

Case 3: $f \geq 0$ and $\{f \geq 0\}$ is $\sigma$-finite with respect to $\mu \otimes \nu$.

By the Simple Approximation Theorem (Theorem 7.5), there is a sequence $(s_n)$ of $\mathcal{A}$-measurable simple functions such that $0 \leq s_n \uparrow f$ pointwise on $X \times Y$. Since $\{f > 0\}$ is $\sigma$-finite with respect to $\mu \otimes \nu$, there is an increasing sequence of sets $F_1, F_2, \ldots \in \mathcal{A}$ such that $X \times Y = \bigcup_{n=1}^{\infty} F_n$ and $\mu \otimes \nu(F_n) < \infty$ for each $n$. Define $f_n = s_n 1_{F_n}$. Then $(f_n)$ is a sequence of $\mathcal{A}$-measurable simple functions such that $0 \leq f_n \uparrow f$ pointwise on $X \times Y$ and $\mu \otimes \nu(\{f_n > 0\}) < \infty$ for each $n$.

By Case 2, for each $n$, there is a set $B_n \in \mathcal{T}$ such that $\nu(B_n^c) = 0$, $f_n^y$ is $\mathcal{S}$-measurable for each $y \in B_n$, and the function

$$g_n(y) = \begin{cases} \int f_n^y(x) d\mu(x) & \text{if } y \in B_n \\ 0 & \text{otherwise} \end{cases}$$

is $\mathcal{T}$-measurable. Let $B = \bigcap_{n=1}^{\infty} B_n$. Then $B \in \mathcal{T}$ with $\nu(B^c) = 0$, $f_n^y$ is $\mathcal{S}$-measurable for every $y \in B$ and every $n$, and the function

$$g_n(y) 1_B(y) = \begin{cases} \int f_n^y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$$
is $\mathcal{T}$-measurable for every $n$. Note that $0 \leq f^+_n \uparrow f^+ y$ pointwise on $X$ for every $y \in Y$. Therefore $f^+ y$ is $\mathcal{T}$-measurable for every $y \in B$. Moreover, by the monotone convergence theorem, $0 \leq g_n 1_B \uparrow g$ pointwise on $Y$, where

$$g(y) = \begin{cases} \int f^+_n(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}. $$

Therefore $g$ is $\mathcal{T}$-measurable. Also by Case 2, we have that $\int g_n(y) d\nu(y) = \int f_n(x,y) d(\mu \otimes \nu)(x,y)$. Notice that $g_n = g_n 1_B$ for $\nu$-a.e. $y \in Y$. So $\int g_n(y) 1_B(y) d\nu(y) = \int f_n(x,y) d(\mu \otimes \nu)(x,y)$. Then, applying the monotone convergence theorem twice,

$$\int g(y) d\nu(y) = \lim_n \int g_n(y) d\nu(y) = \lim_n \int f_n(x,y) d(\mu \otimes \nu)(x,y) = \int f(x,y) d\mu \otimes \nu(x,y).$$

\[ \square \]

**Theorem 29.4. (Fubini’s Theorem).** Let $f : X \times Y \rightarrow \mathbb{R}$ be $\mathcal{A}$-measurable. Assume $f$ is integrable with respect to $\mu \otimes \nu$.

(a) $f^+ y$ is $\mathcal{S}$-measurable for $\nu$-a.e. $y \in Y$. In other words, there is a set $B \in \mathcal{T}$ such that $\nu(B^c) = 0$ and $f^+ y$ is $\mathcal{T}$-measurable for every $y \in B$.

(b) $f_x$ is $\mathcal{T}$-measurable for $\mu$-a.e. $x \in X$. In other words, there is a set $A \in \mathcal{S}$ such that $\mu(A^c) = 0$ and $f_x$ is $\mathcal{S}$-measurable for every $x \in A$.

(c) There exists a set $B$ as in (a) such that the function $g(y) = \begin{cases} \int f^+ y(x) d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{T}$-measurable.

(d) There exists a set $A$ as in (b) such that the function $h(x) = \begin{cases} \int f_x(y) d\nu(y) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$ is $\mathcal{S}$-measurable.

(e) $\int g(y) d\nu(y) = \int f(x,y) d(\mu \otimes \nu)(x,y) = \int h(x) d\mu(x)$.

**Proof Outline.** Apply Tonelli’s theorem to $f^+$ and $f^-$.

\[ \square \]

**Proof.** Since $f$ is integrable with respect to $\mu \otimes \nu$ and since

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

$f^+$ and $f^-$ are integrable with respect to $\mu \otimes \nu$. \{ $f^+ > 0$ \} and \{ $f^- > 0$ \} $\sigma$-finite with respect to $\mu \otimes \nu$. Tonelli’s theorem implies (a)-(e) hold with $f$ replaced by $f^+$ and $f^-$. In particular, there are sets $B^+, B^- \in \mathcal{T}$ such that $\nu((B^+)^c) = \nu((B^-)^c) = 0$, $(f^+)^y$ is $\mathcal{S}$-measurable for all $y \in B^+$, $(f^-)^y$ is $\mathcal{S}$-measurable for all $y \in B^-$, the functions $g^+_0(y) = \{ \int (f^+)^y(x) d\mu(x) | y \in B^+ \}$ and $g^-_0(y) = \{ \int (f^-)^y(x) d\mu(x) | y \in B^- \}$ are both $\mathcal{T}$-measurable, and

$$\int g^+_0 d\nu(y) = \int f^+ d(\mu \otimes \nu) \quad \text{and} \quad \int g^-_0 d\nu(y) = \int f^- d(\mu \otimes \nu).$$ (29.2)
Define $B = B^+ \cap B^-$. Then $B \in \mathcal{T}$, $\nu(B) = 0$, $(f^+)^y$ and $(f^-)^y$ are $\mathcal{S}$-measurable for all $y \in B$, and the functions $g_0^+ 1_B$ and $g_0^- 1_B$ are both $\mathcal{T}$-measurable. Since $f = f^+ - f^-$, we have $f^y = (f^+)^y - (f^-)^y$ for every $y \in Y$. Therefore $f^y$ is $\mathcal{S}$-measurable for all $y \in B$. This proves (a). Moreover, for every $y \in Y$, 

$$g(y) = \begin{cases} \int f^y(x)d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \int (f^y)^+(x)d\mu(x) - \int (f^y)^-(x)d\mu(x) & \text{if } y \in B \\ 0 & \text{otherwise} \end{cases}$$

$$= g_0^+(y) 1_B(y) - g_0^-(y) 1_B(y).$$

So $g$ is $\mathcal{T}$-measurable. This proves (c). (As an aside, since $g_0^+, g_0^- \geq 0$, we also have $g^+ = g_0^+ 1_B$ and $g^- = g_0^- 1_B$.) Finally, since $g_0^+ 1_B = g_0^+$ and $g_0^- 1_B = g_0^-$ hold $\nu$-a.e., (29.2) implies 

$$\int g(y)d\nu(y) = \int g_0^+(y) 1_B(y)d\nu(y) - \int g_0^-(y) 1_B(y)d\nu(y) = \int g_0^+(y)d\nu(y) - \int g_0^-(y)d\nu(y)$$

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

This proves the first equality in (e). The proofs of (b),(d), and the second equality in (e) are similar. 

Remark 29.5. Concerning the hypotheses of Tonelli’s theorem, if $X$ is $\sigma$-finite with respect to $\mu$ and $Y$ is $\sigma$-finite with respect to $\nu$, then Theorem 28.3(d) implies $\{f > 0\}$ is $\sigma$-finite with respect to $\mu \otimes \nu$.

Remark 29.6. The conclusion (e) in Fubini’s and Tonelli’s theorem is often stated as 

$$\int \int f(x,y)d\mu(x)d\nu(y) = \int f(x,y)d\mu \otimes \nu(x,y) = \int \int f(x,y)d\nu(x)d\mu(y).$$

The statement (e) is the precise version.

Remark 29.7. Fubini’s and Tonelli’s theorems are frequently used together. Typically one wishes to reverse the order of integration in a double integral $\int \int f d\mu d\nu$. First one verifies that $\int |f|d(\mu \otimes \nu) < \infty$ by using Tonelli’s theorem to evaluate this integral as the iterated integral $\int \int |f|d\mu d\nu$. Then one applies Fubini’s theorem to conclude that $\int \int f d\mu d\nu = \int \int f d\nu d\mu$. Some examples are given in the exercises.

Remark 29.8. If $A = \mathcal{S} \otimes \mathcal{T}$, the hypothesis that $(X, \mathcal{S}, \mu)$ and $(X, \mathcal{T}, \mu)$ are complete can be omitted. Moreover, in Theorems 29.3 and 29.4, the conclusions can be modified as follows:

(a) $f^y$ is $\mathcal{S}$-measurable for every $y \in Y$.

(b) $f_x$ is $\mathcal{T}$-measurable for every $x \in X$.

(c) The function $g(y) = \int f^y(x)d\mu(x)$ is $\mathcal{T}$-measurable.

(d) The function $h(x) = \int f_x(y)d\mu(y)$ is $\mathcal{S}$-measurable.

(e) $\int g(y)d\nu(y) = \int f(x,y)d(\mu \otimes \nu)(x,y) = \int h(x)d\mu(x)$.

An analogous modification of Lemma 29.2 also holds.
30 Product Spaces With More Than Two Factors

All the integration theory we have developed for product spaces of two factors $X \times Y$ generalizes naturally to products spaces of $n$ factors $\prod_{i=1}^{n} X_i = X_1 \times \cdots \times X_n$. We state the main definitions and theorems, but omit the proofs and any inessential details.

Let $X_1, \ldots, X_n$ be sets.

**Notation.** If $S_i \subseteq \mathcal{P}(X_i)$ ($i = 1, \ldots, n$), we define

\[
\prod_{i=1}^{n} S_i = S_1 \times \cdots \times S_n = \left\{ \prod_{i=1}^{n} A_i = A_1 \times \cdots \times A_n : A_i \in S_i \right\}
\]

Note that this is not the usual Cartesian product of $S_i$ ($i = 1, \ldots, n$)

**Definition 30.1.** Let $\mathcal{S}_i$ be a $\sigma$-algebra on $X_i$ ($i = 1, \ldots, n$).

(a) The sets in $\prod_{i=1}^{n} \mathcal{S}_i$ are called **measurable rectangles**.

(b) The $\sigma$-algebra on $\prod_{i=1}^{n} X_i$ generated by $\prod_{i=1}^{n} \mathcal{S}_i$ is called the **product $\sigma$-algebra** of $\mathcal{S}_1, \ldots, \mathcal{S}_n$.

It is denoted by $\bigotimes_{i=1}^{n} \mathcal{S}_i = \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$. In other words, $\bigotimes_{i=1}^{n} \mathcal{S}_i = \sigma(\prod_{i=1}^{n} \mathcal{S}_i)$.

**Lemma 30.2.** If $\mathcal{E}_i$ is a demi-ring on $X_i$ for $i = 1, \ldots, n$, then $\prod_{i=1}^{n} \mathcal{E}_i$ is a demi-ring on $\prod_{i=1}^{n} X_i$.

**Corollary 30.3.** If $\mathcal{S}_i$ is a $\sigma$-algebra on $X_i$ for $i = 1, \ldots, n$, then $\prod_{i=1}^{n} \mathcal{S}_i$ is a demi-ring on $\prod_{i=1}^{n} X_i$.

**Theorem 30.4.** Let $(X_i, \mathcal{S}_i, \mu_i)$ be a measure space for $i = 1, \ldots, n$. Define $\pi_0 : \prod_{i=1}^{n} \mathcal{S}_i \to [0, \infty]$ by

\[
\pi_0 \left( \prod_{i=1}^{n} A_i \right) = \prod_{i=1}^{n} \mu_i(A_i).
\]

Let $\pi_0$ be the outer measure generated by $\pi_0$.

(a) $\pi_0$ is finitely additive and countably monotone.

(b) $\prod_{i=1}^{n} \mathcal{S}_i \subseteq M(\pi^*)$ and $\pi^* = \pi_0$ on $\prod_{i=1}^{n} \mathcal{S}_i$.

(c) Let $\mathcal{A}$ be a $\sigma$-algebra on $\prod_{i=1}^{n} X_i$ such that $\prod_{i=1}^{n} \mathcal{S}_i \subseteq \mathcal{A} \subseteq M(\pi^*)$. Let $\bigotimes_{i=1}^{n} \mu_i = \pi^* \vert \mathcal{A}$ (i.e., $\bigotimes_{i=1}^{n} \mu_i$ is the restriction of $\pi^*$ to $\mathcal{A}$). Then $\bigotimes_{i=1}^{n} \mu_i$ is a measure on $\mathcal{A}$ and

\[
\bigotimes_{i=1}^{n} \mu_i \left( \prod_{i=1}^{n} A_i \right) = \prod_{i=1}^{n} \mu_i(A_i) \quad \text{for all } \prod_{i=1}^{n} A_i \in \prod_{i=1}^{n} \mathcal{S}_i. \tag{30.1}
\]

Moreover, if $\mathcal{A} = M(\pi^*)$, then $\bigotimes_{i=1}^{n} \mu_i$ is complete. We call $\bigotimes_{i=1}^{n} \mu_i$ the **product measure** of $\mu_1, \ldots, \mu_n$ on $\mathcal{A}$.

(d) Under the assumptions of (c), if $X_i$ is $\sigma$-finite with respect to $\mu_i$ for $i = 1, \ldots, n$, then $\prod_{i=1}^{n} X_i$ is $\sigma$-finite with respect to $\pi_0$ and $\bigotimes_{i=1}^{n} \mu_i$ is the unique measure on $\mathcal{A}$ that satisfies (30.1).

**Theorem 30.5.** (Fubini and Tonelli Theorems) Let $(X_i, \mathcal{S}_i, \mu_i)$ be a complete measure space for $i = 1, \ldots, n$. Let $\mathcal{A}$ be a $\sigma$-algebra on $\prod_{i=1}^{n} X_i$ such that $\prod_{i=1}^{n} \mathcal{S}_i \subseteq \mathcal{A} \subseteq M(\pi^*)$. Let $\bigotimes_{i=1}^{n} \mu_i$ the product of $\mu_1, \ldots, \mu_n$ on $\mathcal{A}$. Let $f : \prod_{i=1}^{n} X_i \to \mathbb{R}$ be $\mathcal{A}$ measurable. Assume at least one of the following two conditions holds:
(i) $f \geq 0$ and $\{f > 0\}$ is $\sigma$-finite with respect to $\bigotimes_{i=1}^{n} \mu_i$

(ii) $f$ is integrable with respect to $\bigotimes_{i=1}^{n} \mu_i$.

Then, for any permutation $p : \{1, \ldots, n\} \to \{1, \ldots, n\}$,

$$\int \cdots \int f(x_1, \ldots, x_n) d\mu_1(x_1) \cdots d\mu_n(x_n) = \int f(x_1, \ldots, x_n) d(\bigotimes_{i=1}^{n} \mu_i)(x_1, \ldots, x_n)$$

$$= \int \cdots \int f(x_1, \ldots, x_n) d\mu_{p(1)}(x_{p(1)}) \cdots d\mu_{p(n)}(x_{p(n)}).$$
31 Lebesgue Measure on $\mathbb{R}^n$

In this section, we construct Lebesgue measure on $\mathbb{R}^n$. We use the product measure construction of Theorem 30.4.

**Theorem 31.1.** Let $\lambda$ be the Lebesgue measure on $\mathcal{L}(\mathbb{R})$. Define $\lambda^n_0 : \mathcal{L}(\mathbb{R})^n \rightarrow [0, \infty]$ by

$$
\lambda^n_0 \left( \prod_{i=1}^{n} A_i \right) = \prod_{i=1}^{n} \lambda(A_i)
$$

Let $\lambda^{n,*}$ be the outer measure on $\mathbb{R}^n$ generated by $\lambda^n_0$. It is called the Lebesgue outer measure on $\mathbb{R}^n$. Let $\mathcal{E}$ be the collection of all intervals of the form $(a, b]$ where $a, b \in \mathbb{R}$ and $a \leq b$. Let

$$
\mathcal{E}^n = \mathcal{E} \times \cdots \times \mathcal{E} = \left\{ \prod_{i=1}^{n} (a_i, b_i] : (a_i, b_i] \in \mathcal{E} \right\}.
$$

(a) $\mathcal{L}(\mathbb{R})^n$ is a demi-ring on $\mathbb{R}^n$.
(b) $\lambda^n_0$ is finitely additive.
(c) $\lambda^n_0$ is countably monotone.
(d) $\mathbb{R}^n$ is $\sigma$-finite with respect to $\lambda^n_0$.
(e) $\lambda^{n,*} = \lambda^n_0$ on $\mathcal{L}(\mathbb{R})^n$. In particular, $\lambda^{n,*}$ assigns any product of intervals its volume.
(f) Define $\mathcal{L}(\mathbb{R}^n) = M(\lambda^{n,*})$. We call $\mathcal{L}(\mathbb{R}^n)$ the Lebesgue $\sigma$-algebra on $\mathbb{R}^n$. It satisfies $\mathcal{L}(\mathbb{R})^n \subseteq \mathcal{L}(\mathbb{R}^n)$ and $\mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{L}(\mathbb{R}^n)$.
(g) $\lambda^n|_{\mathcal{L}(\mathbb{R}^n)}$ is a complete measure on $\mathcal{L}(\mathbb{R}^n)$. In fact, it is equal to the product measure $\bigotimes_{i=1}^{n} \lambda$ on $\mathcal{L}(\mathbb{R}^n)$. It is the unique measure on $\mathcal{L}(\mathbb{R}^n)$ that agrees with $\lambda^n_0$ on $\mathcal{E}^n$. It is called Lebesgue measure on $\mathcal{L}(\mathbb{R}^n)$. It is denoted by $\lambda^n$.
(h) $\lambda^n|_{\mathcal{B}(\mathbb{R}^n)}$ is a measure on $\mathcal{B}(\mathbb{R}^n)$. It is the unique measure on $\mathcal{B}(\mathbb{R}^n)$ that agrees with $\lambda^n_0$ on $\mathcal{E}^n$. It is called Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$. It is also denoted by $\lambda^n$.

**Proof.**

(a),(b),(c): Immediate from Theorem 30.4.
(d): $\mathbb{R}^n = \bigcup_{k=1}^{\infty} (-k, k]^n$ and $\lambda^n_0 \left( \bigcup_{k=1}^{\infty} (-k, k]^n \right) = (2k)^n < \infty$ for all $k \in \mathbb{N}$.
(e): $\lambda^{n,*} = \lambda^n_0$ on $\mathcal{L}(\mathbb{R}^n)$ is immediate from Theorem 30.4. Combining this with Theorem 22.1(e) gives that $\lambda^{n,*}$ assigns any product of intervals its volume.
(f): The containment $\mathcal{L}(\mathbb{R})^n \subseteq \mathcal{L}(\mathbb{R}^n)$ comes from Theorem 30.4. A simple modification of the proof of Theorem 3.9 shows that $\mathcal{B}(\mathbb{R}^n)$ is generated by the collection of all open rectangles

$$
\mathcal{R} = \left\{ \prod_{i=1}^{n} (a_i, b_i] : a_i, b_i \in \mathbb{R}, a_i \leq b_i \right\}.
$$
Since $\mathcal{R} \subseteq \mathcal{L}(\mathbb{R})^n \subseteq \mathcal{L}(\mathbb{R}^n)$, and since $\mathcal{L}(\mathbb{R}^n)$ is a $\sigma$-algebra it follows that $\mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{L}(\mathbb{R}^n)$.

(g): Everything except the uniqueness assertion follows immediately from Theorem 30.4. Now we prove uniqueness. By theorem 22.1(a), $\mathcal{E}$ is a demi-ring on $\mathbb{R}$. By Lemma 30.2, $\mathcal{E}^n$ is a demi-ring on $\mathbb{R}^n$. Note that $\mathcal{E}^n \subseteq \mathcal{L}(\mathbb{R})^n \subseteq \mathcal{L}(\mathbb{R}^n)$. So, by (e), $\lambda^0_n = \lambda^n_* = \lambda^n$ on $\mathcal{E}^n$. Since $\lambda^n$ is a measure, it is finitely additive and countably monotone on $\mathcal{E}^n$. But since $\lambda^0_n = \lambda^n$ on $\mathcal{E}^n$, we have that $\lambda^0_n$ is finitely additive and countably monotone on $\mathcal{E}^n$. Therefore the desired uniqueness assertion follows from (d) and Theorem 21.1(c).

(h): Since $\mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{L}(\mathbb{R}^n)$ and since $\lambda^*|_{\mathcal{E}^n}$ is a measure on $\mathcal{L}(\mathbb{R}^n)$, $\lambda^*|_{\mathcal{B}(\mathbb{R}^n)}$ is a measure on $\mathcal{B}(\mathbb{R}^n)$. Since $\mathcal{E}^n \subseteq \mathcal{B}(\mathbb{R}^n)$, the uniqueness assertion can be proved just like in (g). \qed

Since $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ is complete and $\sigma$-finite, Theorem 30.5 implies the following.

**Theorem 31.2.** (Fubini and Tonelli Theorems on $(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \lambda^n)$). Let $f : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{L}(\mathbb{R}^n)$-measurable. Assume at least one of the following two conditions holds:

(i) $f \geq 0$

(ii) $f$ is integrable with respect to $\lambda^n$.

Then, for any permutation $p : \{1, \ldots, n\} \to \{1, \ldots, n\}$,

$$\int \cdots \int f(x_1, \ldots, x_n)d\lambda(x_1) \cdots d\lambda(x_n) = \int f(x_1, \ldots, x_n)d\lambda^n(x_1, \ldots, x_n) = \int \cdots \int f(x_1, \ldots, x_n)d\lambda(x_{p(1)}) \cdots d\lambda(x_{p(n)}).$$