This is a preliminary version of the lecture notes for this course. A final version will be uploaded towards the end of the semester. In the mean time, if you have any questions about the content of the notes, please do email me.

MA32061: Measure theory and integration

Lecturer: Matthew Schrecker
Office: 6 West 1.13
Email: mris21@bath.ac.uk

2025-26, Semester 2

Aims: To lay the basic technical foundations and establish the main principles which underpin the classical notions of area, volume and the related idea of an integral. To familiarise students with measure as a tool in analysis, functional analysis and probability theory.

Content: Systems of measurable sets: σ-algebras, π-systems, d-systems, Dynkin's Lemma, Borel σ-algebras. Measure in the abstract: convergence properties, Uniqueness Lemma, Carathéodory's Theorem. Lebesgue measure on $\mathbb{R}^n$. Measurable functions. Monotone-Class Theorem. Lebesgue integration. Probability, Random variables, Expectation. Monotone-Convergence Theorem. Fatou's Lemma. Dominated-Convergence Theorem. Product measures. Tonelli's and Fubini's Theorem. Radon-Nikodým Theorem. Inequalities of Jensen, Hölder, Minkowski. Completeness of L^p. For some theorems the proofs will be optional reading material.

Learning Outcomes: On completing the course, students should be able to: * demonstrate a good knowledge and understanding of the main results and techniques in measure theory; * demonstrate an understanding of the Lebesgue Integral; * quote and apply the main inequalities of measure theory in a wide range of contexts.

References:

- G. Folland, *Real Analysis: Modern techniques and their Applications*, second edition, Wiley, New York, 1999.
- W. Rudin, *Real and Complex Analysis*, third edition, McGraw-Hill, New York, 1987.
- D. Williams, *Probability with Martingales.* Cambridge, 9th ed., 2005.

In-text exercises: Some statements made in the text are followed by "(HW)". This means they are left as exercises for you to do. Some of these will appear on problem sheets, but not all of them. You will be expected to know how to do them.

Potential typos: Unavoidably, some typos will have crept in. Please ask if anything is not clear.

Non-examinable: Any part that is *non-examinable* is marked as such, printed in smaller font size and indented. All else is examinable.

Contents

1 Preliminaries

Cardinality is a notion used to decide whether or not two sets have the same number of elements. As an example, consider the sets of natural numbers and integers:

$$\mathbb{N} = \{1, 2, 3, \ldots\} \qquad \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

You should recall that it is possible to construct a bijection (that is, a map that is both injective and surjective) between $\mathbb{N}$ and $\mathbb{Z}$, for example by

$$f : \mathbb{N} \to \mathbb{Z}, \qquad f(j) = \begin{cases} -\frac{n}{2}, & \text{if } n \text{ is even,} \\ \frac{n-1}{2}, & \text{if } n \text{ is odd.} \end{cases}$$

For this reason, we say that $\mathbb{N}$ and $\mathbb{Z}$ have the same cardinality. In general, we have the following definition

Definition 1.1. (i) Let X and Y be two sets (finite or infinite). If there exists an injective map $f : X \to Y$, then we say that the cardinality of X is less than or equal to the cardinality of Y and write $\#X \leq \#Y$. If both $\#X \leq \#Y$ and $\#Y \leq \#X$, we say that X and Y have the same cardinality and write $\#X = \#Y$.
(ii) A set X is called countable if $\#X = \#\mathbb{N}$.

It should be clear that $\#X = \#Y$ is equivalent to the existence of a bijection from X to Y. Note that finite sets are not countable according to this definition. This is a matter of convention; some authors instead define a set $\#X$ to be countable if $\#X \leq \#\mathbb{N}$.

Proposition 1.2. *(i) The set $\mathbb{Q}$ of rational numbers has the same cardinality as $\#\mathbb{N}$.*
(ii) The set $\mathbb{R}$ of real numbers has cardinality $\#\mathbb{N} \leq \#\mathbb{R}$ but $\#\mathbb{N} \neq \#\mathbb{R}$.

One can in fact show without difficulty that $\#(0,1) = \#\mathbb{R}$. This shows that while cardinality gives a very important measure of the size of the set, from the point of view of analysis (and in particular of integration theory), it does not distinguish between the size of sets that, from another point of view, we might want to consider as being of very different size.

Definition 1.3. (i) The *supremum* of a set $A \subset \mathbb{R}$ is the least upper bound of A. That is, for all $x \in A$, $x \leq \sup A$ and, if S is any other upper bound for A, $\sup A \leq S$.
(ii) The *infimum* of a set $A \subset \mathbb{R}$ is the greatest lower bound of A. That is, for all $x \in A$, $x \geq \inf A$ and, if S is any other lower bound for A, $\inf A \geq S$.
The completeness of $\mathbb{R}$ means that every non-empty set $A \subset \mathbb{R}$ that is bounded above has a supremum. Every non-empty set $A \subset \mathbb{R}$ that is bounded below has an infimum.

Definition 1.4. Let X be a set, $f : X \to \mathbb{R}$ and, for each $n \in \mathbb{N}$, let $f_n : X \to \mathbb{R}$. We say that $f_n \to f$ *pointwise* if for all $x \in X$, the sequence of real numbers $f_n(x) \to f(x)$. We say that $f_n \to f$ *uniformly* if: for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that, for all $n \geq N$, all $x \in X$, we have $|f_n(x) - f(x)| < \varepsilon$.

Recall that pointwise convergence is much weaker than uniform convergence of functions.

Definition 1.5. Let (M, d) be a metric space.

- A set $U \subset M$ is said to be **open**, if for all $x \in U$ there exists $r > 0$ such that $B_r(x) \subset U$.
- A set $F \subset M$ is said to be **closed**, if $\mathbb{R} \setminus F$ is open.
- Given a set $A \subset M$, a *cover* of A is a collection of sets $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ such that $A \subset \cup_{\alpha \in \mathcal{A}} U_\alpha$. It is an *open cover* if each U_α is an open set.
- A set $S \subset M$ is said to be **bounded** if there exist $r > 0$ and $x \in M$ such that $S \subset B_r(x)$.
- A set A in M is said to be **compact** if every open cover of A contains a finite subcover of A.

Theorem 1.6 (Heine-Borel)**.** *A set $A \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Definition 1.7. A normed vector space $(V, \|\cdot\|)$ is a *Banach space* if it is complete. That is, if every Cauchy sequence $(v_n)_{n=1}^{\infty} \subset V$ converges to some $v \in V$.

2 Introduction

Motivation for studying measure theory

- **Mathematical curiosity.** It is a natural question to ask whether every set in Euclidean space $\mathbb{R}^d$ has a well-defined volume. For example, we know that in $\mathbb{R}^2$, every rectangle has area given by length times height, every disk has area πr^2, and so on. Is it possible to extend this definition in a sensible manner to ALL sets in $\mathbb{R}^2$?

- **Probability theory.** Probability theory is important as a mathematical underpinning of theoretical statistics, mathematical finance, statistical physics or any situation where one has incomplete information. The theory of probability is heavily reliant on measure theory. Indeed, the foundations of probability theory can be viewed as abstract measure theory expressed in a different language.

- **Issues with Riemann integration** The Riemann integral constructed in earlier Analysis courses is excellent for integrating continuous or monotone functions. However, it does have a number of drawbacks.

  (i) Firstly, it is hard to characterise the set of Riemann integrable functions. There are many functions that are Riemann integrable but are neither continuous nor monotone. This makes it difficult to know whether a given function can be integrated with the Riemann integral and what its integral should be.

  (ii) Secondly, there are many functions that we would like to be able to integrate that are not Riemann integrable. A good example of such a function is the Dirichlet function on the interval $[0, 1]$, also known as the characteristic function of the rationals:
  $$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \cap [0, 1], \\ 0, & \text{else.} \end{cases}$$

  (iii) Exchanging limits of functions and Riemann integrals requires the very strong property of uniform convergence (i.e. to show that $\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$). This can often not be checked easily (or in fact fails). As an example of a sufficient condition for uniform convergence, compare the Arzelà–Ascoli Theorem which tells us that equicontinuity of a sequence of functions (and equiboundedness) gives uniform convergence. This is a very strong condition!

  NB: If you cannot remember the notions of pointwise and uniform convergence, you should revise these (see the Preliminaries section above) as they will be important throughout the second half of this course.

  To think more about the second and third of these issues, let us consider a sequence of functions that approximates (pointwise) the Dirichlet function f. Let $\{q_i\}_{i=1}^{\infty}$ be an enumeration of $\mathbb{Q} \cap [0, 1]$ and define the function
  $$f_n(x) = \begin{cases} 1, & \text{if } x \in \bigcup_{i=1}^{n} \{q_i\}, \\ 0, & \text{else.} \end{cases}$$

  Clearly $f_n \to f$ pointwise but not uniformly. In addition, each f_n is Riemann integrable and satisfies $\int_{[0,1]} f_n = 0$ (it is a good exercise to prove this for yourself) but $\int_{[0,1]} f$ is not defined. That each f_n has zero integral is clear as each f_n is zero except on a finite set of points. To consider how we might want to assign a value to the integral of f, let us consider how large (in a vague sense) the set of points on which f is non-zero is.

Let us fix a value $\varepsilon > 0$ and consider the cover of $\mathbb{Q} \cap [0, 1]$ given by

$$\mathcal{U} = \{U_n \,|\, n \in \mathbb{N}\}, \qquad U_n = \big(q_n - \frac{\varepsilon}{2^n}, q_n + \frac{\varepsilon}{2^n}\big),$$

where $\{q_n\}$ is the same enumeration of $\mathbb{Q} \cap [0, 1]$ as before. Clearly $\mathbb{Q} \cap [0, 1] \subset \bigcup_{n=1}^{\infty} U_n$.

On the other hand, we can think of each of the open intervals U_n as having a 'size' of $\frac{2\varepsilon}{2^n}$ (as this is the length of the interval). As the whole of $\mathbb{Q} \cap [0, 1]$ is contained in $\bigcup_{n=1}^{\infty} U_n$, we should therefore hope that the 'size' of this set is bounded by

$$\sum_{n=1}^{\infty} \frac{2\varepsilon}{2^n} = 2\varepsilon \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 2\varepsilon.$$

As ε was chosen arbitrarily, we conclude that the only sensible value we could give to the measure of $\mathbb{Q} \cap [0, 1]$ is 0. We should therefore hope that, if we can define an appropriate notion of integral, we would conclude

$$\int_{[0,1]} f = 0 = \lim_{n\to\infty} \int_{[0,1]} f_n.$$

In fact, this basic approach of covering sets with intervals (for which the appropriate notion of length or 'measure' is clear) is how we are going to construct the Lebesgue measure on $\mathbb{R}$.

Before we proceed to this most fundamental measure, though, we consider a possible wish-list that a measure would ideally achieve and an immediate paradox that arises.

A paradox. Suppose we have a function $\mu(A)$, representing 'length' or 'size' (μ is the Greek letter 'mu') of the set $A \subset \mathbb{R}$. To match our intuition about length we would also like μ to have the following wish list of properties:

(a) $\mu(A)$ is defined for all $A \subset \mathbb{R}$.

(b) $\mu((a, b]) = b - a$ for any real $a < b$. (Recall that $(a, b] := \{x \in \mathbb{R} : a < x \leq b\}$. We use $:=$ to denote definition.)

(c) *Nonnegativity:* $0 \leq \mu(A) \leq \infty$ for all $A \subset \mathbb{R}$.

(d) *Countable additivity*: $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $A_1, A_2, \ldots,$ are pairwise disjoint (i.e., satisfy $A_i \cap A_j = \varnothing$ whenever $i \neq j$.)

(e) *Translation invariance*: $\mu(A + x) = \mu(A)$ for all $A \subset \mathbb{R}$ and $x \in \mathbb{R}$. Here $A + x := \{a + x : a \in A\}$.

Reminder: Given $x_1, x_2, \ldots$ with $x_i \geq 0$ for each i, we define $\sum_{i=1}^{\infty} x_i := \lim_{n\to\infty} \sum_{i=1}^{n} x_i$. The limit (possibly $+\infty$) exists since it is the limit of an nondecreasing sequence of numbers. If any of the x_i are $+\infty$ we also define $\sum_{i=1}^{\infty} x_i = +\infty$. We are using these definitions in the property (d) above (and again in similar situations later on).

Theorem 2.1. *There exists a set $E \subset (0,1]$ such that, setting $E_q := E + q$ for all $q \in \mathbb{Q}$ and writing $\mathbb{Q} = \{q_n\}_{n=1}^{\infty}$, we have*

$$(0,1] \subset \bigcup_{n=1}^{\infty} E_{q_n} \tag{1}$$

and

$$\text{the sets } E_q, q \in (-1,1] \cap \mathbb{Q} \text{ are pairwise disjoint.} \tag{2}$$

Proof. *(Non-examinable)* Define the following relation between elements of the interval $(0,1]$: for $x, y \in (0,1]$, we write $x \sim y$, if $x - y \in \mathbb{Q}$. This is an equivalence relation:
(i) $x \sim x$, because $x - x = 0 \in \mathbb{Q}$;
(ii) $x \sim y$ implies $y \sim x$, because $y - x = -(x - y) \in \mathbb{Q}$;
(iii) $x \sim y$ and $y \sim z$ imply $x \sim z$, because $x - z = (x - y) + (y - z) \in \mathbb{Q}$.
Consider the equivalence classes with respect to the equivalence relation $\sim$. Let $E \subset (0,1]$ be a set that contains *exactly one* number from each equivalence class.

Suppose we had $x \in E_q \cap E_r$, with $q, r \in (-1,1] \cap \mathbb{Q}$. Then due to the definitions of E_q and E_r, we must have $x = y + q = z + r$ for some $y, z \in E$. But then $y - z = r - q \in \mathbb{Q}$, and hence $y \sim z$. But since E contains *exactly one* number from each equivalence class, we can only have $y \sim z$ if $y = z$ and hence $q = r$. This proves (2).

To see (1), let $x \in (0,1]$. Then there exists a unique $y \in E$ such that $x \sim y$. Let $q = x - y \in (-1,1] \cap \mathbb{Q}$. Then $x = y + q \in E_q$, so x is an element of the set on the right in (1). □

Corollary 2.2. *There does* **not** *exist any 'length function' μ satisfying properties (a)–(e) listed at the start of this section.*

Proof. By contradiction. Suppose such a function μ exists. Let $E \subset (0,1]$ be as given by the preceding Theorem (i.e. with properties (1) and (2)). Observe first that

$$(0,1] \subset \cup_{n=1}^{\infty} E_{q_n} \subset (-1,2], \tag{3}$$

where the first inclusion comes from (1), and the second is because $E \subset (0,1]$ and each q is in $(-1,1]$. Using (3), we obtain that

$$1 = \mu((0,1]) \leq \mu(\cup_{n=1}^{\infty} E_{q_n}) \leq \mu((-1,2]) = 3, \tag{4}$$

so $\mu(\cup_{n=1}^{\infty} E_{q_n})$ is finite and strictly positive. However, for all q we have $\mu(E_q) = \mu(E)$ by translation invariance. Hence using (2) and countable additivity we have

$$\mu(\cup_{n=1}^{\infty} E_{q_n}) = \sum_{n=1}^{\infty} \mu(E_{q_n}) = \sum_{n=1}^{\infty} \mu(E) \tag{5}$$

Hence $\mu(\cup_{q \in (-1,1] \cap \mathbb{Q}} E_q)$ is either 0 (if $\mu(E) = 0$) or is ∞ (if $\mu(E) > 0$). Either way, this contradicts (4). □

Similar problems arise when trying to define 'area' (respectively 'volume') for all sets in $\mathbb{R}^2$ (resp. $\mathbb{R}^3$).

To resolve this issue, later on we shall relax Condition (a) in our list, i.e. define the function $\mu(A)$, not for *all* $A \subset \mathbb{R}$, but for *a large class of* sets $A \subset \mathbb{R}$. For example, we'd like our class of sets to include all intervals in $\mathbb{R}$.

Remark. *(Non-examinable.)* In defining the set E we used the so-called *Axiom of choice*, that asserts that if there is a family of sets $\{A_\alpha : \alpha \in I\}$, then there exists a function $f : I \to \cup_{\alpha \in I} A_\alpha$ such that $f(\alpha) \in A_\alpha$ for each $\alpha \in I$. That is, there exists a choice function that selects one element from each set in the family. In our case the sets A_α are the equivalence classes with respect to $\sim$.

3 The Lebesgue (outer) measure

Based on the foregoing observations, we now begin to work towards constructing the Lebesgue measure. This follows a procedure that is commonly used to construct *many* measures, not just the Lebesgue measure (we will come back to this in Section **??**). The basic idea is to define a 'sensible' notion of length for every set in the real line, acknowledging that this notion of length will not define a measure for us by the paradox just described. Instead, we will try to prove as much as we can for this length (which we will call the *outer measure*) and then identify a good collection of subsets of $\mathbb{R}$ on which the outer measure does satisfy all the properties we want our measure to satisfy.

Definition 3.1 (Outer Lebesgue Measure)**.** Let $I = (a, b)$, $[a, b)$, $(a, b]$ or $[a, b]$ with $a < b$ be an interval. We define the length of I to be $|I| = b - a$.
Let $A \subset \mathbb{R}$. We define the outer measure of A to be

$$\mu^*(A) = \inf\{\sum_{j=1}^{\infty} |I_j| \mid \{I_j\}_{j=1}^{\infty} \text{ is a cover of } A \text{ by open intervals}\}.$$

Note that any set in $\mathbb{R}$ can be covered by intervals, and so this is well defined. Moreover, as the infimum is over a non-empty set of positive numbers, $\mu^*(A) \geq 0$ (but is possibly infinite) for each $A \subset \mathbb{R}$. Additionally, we do not have to take an infinite collection of intervals if it is not necessary as we may take most of the intervals to be empty (think of $I = (a, a) = \varnothing$).

The outer measure μ^* will not in fact have most of the properties that we want our measure to have. Instead, we are going to find a collection of subsets of $\mathbb{R}$ with 'nice' properties such that the full wish list holds once we restrict μ^* to this collection.

Example
$\mu^*(\mathbb{Q}) = 0$.
First, note that $\mu^*(\mathbb{Q}) \geq 0$ by definition of μ^*. It is therefore sufficient to show that $\mu^*(\mathbb{Q}) \leq 0$. To do this, we follow essentially the strategy outlined before in §1. Fix $\varepsilon > 0$, let $\{x_n\}_{n=1}^{\infty}$ be an enumeration of $\mathbb{Q}$ and define the interval $I_j = \left(x_j - \frac{\varepsilon}{2^j}, x_j + \frac{\varepsilon}{2^j}\right)$. Now clearly we have $\mathbb{Q} \subset \bigcup_{j=1}^{\infty} I_j$, and so

$$\mu^*(\mathbb{Q}) \leq \sum_{j=1}^{\infty} |I_j| = \sum_{j=1}^{\infty} \frac{2\varepsilon}{2^j} = 2\varepsilon.$$

As $\varepsilon > 0$ may be taken arbitrarily small, we conclude that $\mu^*(\mathbb{Q}) \leq 0$, and hence also $\mu^*(\mathbb{Q}) = 0$.

Example
$\mu^*([a, b]) = b - a$.
We need to check that the two definitions provided for the measure of an interval are consistent. We first show that $\mu^*([a, b]) \leq b - a$ by taking an open cover, for $\varepsilon > 0$, given by

$$I_1 = \left(a - \frac{\varepsilon}{2}, b + \frac{\varepsilon}{2}\right), \qquad I_j = \varnothing, \; j \geq 2.$$

Then

$$\mu^*([a, b]) \leq \sum_{j=1}^{\infty} |I_j| = |I_1| = b - a + \varepsilon.$$

We therefore conclude $\mu^*([a, b]) \leq b - a$.

To show the reverse inequality, suppose that $\varepsilon > 0$ and we have an open cover of $[a, b]$ by intervals $\{I_j\}_{j=1}^{\infty}$ such that

$$\sum_{j=1}^{\infty} |I_j| \leq \mu^*([a, b]) + \varepsilon.$$

As $[a, b]$ is compact, we may take a finite subcover of n intervals. By relabelling indices and removing intervals if necessary, we order the intervals so that each $I_j = (a_j, b_j)$, where $a \in I_1$, $b_j \in I_{j+1}$ for $j = 1 \dots, n-1$, $b \in I_n$. (NB: this can be done by a process of induction; Exercise). Then $a_1 < a < b_1 < b_2 < \cdots < b_{n-1} < b < b_n$. Now we note that

$$b - a < b_n - a_1 \leq \sum_{j=1}^{n} b_j - a_j \leq \sum_{j=1}^{\infty} |I_j| \leq \mu^*([a, b]) + \varepsilon.$$

Thus $b - a \leq \mu^*([a, b]) + \varepsilon$ for all $\varepsilon > 0$, and hence $\mu^*([a, b]) = b - a$.

Example
The middle third Cantor set.
The middle third Cantor set C is constructed iteratively as follows: At the first step of the iteration, we take the interval $C_0 = [0, 1]$ and remove the middle third $(\frac{1}{3}, \frac{2}{3})$ to define the set $C_1 := [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. At the second step, we remove the open middle third from each of the two disjoint closed intervals that compose C_1, which leaves us with $C_2 := [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. We proceed iteratively, so that at the $(n+1)$-th step, we take the obtained set C_n , which is the union of 2^n closed and disjoint intervals, each of length $\frac{1}{3^n}$, and remove from each of them the open middle third to obtain C_{n+1}. The Cantor set C is then the intersection of all of these sets:

$$C = \bigcap_{n=1}^{\infty} C_n.$$

Claim: C is uncountable (Exercise).

We now show that $\mu^*(C) = 0$ by proving that $\mu^*(C) \leq \left(\frac{2}{3}\right)^n$ for any $n \in \mathbb{N}$.
Observe firstly that $C \subset C_n$ for all n, so that any cover of one of the sets C_n by open intervals is necessarily an open cover of C by open intervals. Choose an $n \in \mathbb{N}$. As C_n is

a finite, disjoint union of 2^n closed intervals of length 3^{-n}, by a simple adaptation of the argument in the previous example, we may deduce that the outer measure of each C_n is $(\frac{2}{3})^n$ and construct a cover of C_n by open intervals I_j such that $\sum_{j=1}^{\infty} |I_j| \leq (\frac{2}{3})^n + \varepsilon$ for any given $\varepsilon > 0$. In particular, we deduce that $\mu^*(C) \leq \left(\frac{2}{3}\right)^n + \varepsilon$ for any $n \in \mathbb{N}$, $\varepsilon > 0$. Thus $\mu^*(C) = 0$.

We see from this example that measure does not mix well with cardinality!

Remark 3.2. Let $A \subset \mathbb{R}$ such that $\mu^*(A) < \infty$ and let $\varepsilon > 0$. By definition of $\mu^*(A)$, there exists an open cover $\{I_j\}_{j=1}^{\infty}$ of A such that

$$\sum_{j=1}^{\infty} |I_j| < \mu^*(A) + \varepsilon.$$

This observation is very helpful for many proofs!

Proposition 3.3. *The outer measure has the following properties.*

(i) For all $A \subset \mathbb{R}$, $0 \leq \mu^*(A) \leq \infty$.

(ii) For all $A, B \subset \mathbb{R}$ such that $A \subset B$, we have $\mu^*(A) \leq \mu^*(B)$.

(iii) For all $A, A_n \subset \mathbb{R}$ such that $A \subset \bigcup_{n=1}^{\infty} A_n$, we have $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

(iv) For all intervals $I = [a, b]$, $[a, b)$, $(a, b]$, (a, b) with $a \leq b$, we have $\mu^*(I) = |I| = b - a$.

(v) For all $A \subset \mathbb{R}$ and $h \in \mathbb{R}$, $\mu^*(A + h) = \mu^*(A)$.

Proof. Part (i) is obvious, while part (iv) has already been proven in the second example above.
(ii) Let $\varepsilon > 0$ and suppose $\{I_j\}_{j=1}^{\infty}$ is a cover of B by open intervals satisfying $\sum_{j=1}^{\infty} |I_j| < \mu^*(B) + \varepsilon$. Then as $A \subset B$, it is also an open cover of A, and hence $\mu^*(A) \leq \sum_{j=1}^{\infty} |I_j| < \mu^*(B) + \varepsilon$. As this holds for any $\varepsilon > 0$, we have $\mu^*(A) \leq \mu^*(B)$.
(iii) Let $\varepsilon > 0$ and suppose that for each n, $\{I_{n,j}\}_{j=1}^{\infty}$ is a cover of A_n by open intervals satisfying $\sum_{j=1}^{\infty} |I_j| < \mu^*(A_n) + \frac{\varepsilon}{2^n}$. By part (ii), we know that $\mu^*(A) \leq \mu^*(\bigcup_{n=1}^{\infty} A_n)$, and so we estimate this upper quantity.

As $\{I_{n,j}\}_{n,j=1}^{\infty}$ is a countable cover of $\bigcup_{n=1}^{\infty} A_n$ by open intervals, we may estimate

$$\mu^*(A) \leq \mu^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} |I_{n,j}| \leq \sum_{n=1}^{\infty} \left(\mu^*(A_n) + \frac{\varepsilon}{2^n}\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.$$

As this holds for all $\varepsilon > 0$, we conclude.
(v) The set of open covers of A by intervals is in bijection with the covers of A_h by open intervals via the correspondence $\{I_j\}_{j=1}^{\infty} \leftrightarrow \{I_j + h\}_{j=1}^{\infty}$. As $|I| = |I + h|$ for any interval I and $h \in \mathbb{R}$, we conclude the proof. □

NB: It is worth remarking at this point that we will frequently see expressions of the form $\sum_{n=1}^{\infty} \mu^*(A_n)$ throughout this course. As each of the terms $\mu^*(A_n)$ is non-negative, the sequence of partial sums $\sum_{j=1}^{n} \mu^*(A_j)$ is always a monotone increasing sequence and will therefore converge either to a non-negative real number or to ∞. Such expressions therefore always make sense provided we are careful to interpret inequalities involving them in such a way that we never consider expressions like $0 \cdot \infty$ or $\infty - \infty$.

With this Proposition, we have recovered points (a), (b), (c) and (e) of our wish list of properties for a measure on $\mathbb{R}$, but we now come to a problem: we cannot show (and it isn't true) that

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

We will rectify this issue soon by introducing the notion of a measurable set, but first we turn to identify what kinds of collections of sets are appropriate for a measure to be defined upon. We therefore take a detour at this point into the world of set theory.

4 σ-algebras

In this chapter, X denotes an arbitrary non-empty set. For example X could be $\mathbb{R}^d$, or in probability theory X might be a sample space of possible 'outcomes'. If A, B are sets then we write $A \setminus B$ for the set of all elements of A that are not elements of B (also known as $A - B$). We will write $X \setminus A = A^c$ for subsets $A \subset X$ where there is no confusion over the space X. We use capital letters (A, B,...) to denote sets, and calligraphic notation ($\mathcal{A}$, $\mathcal{B}$,...) to denote collections of sets (i.e. sets whose elements are themselves sets!)

As discussed in the last chapter, it is sometimes not possible to consistently define the 'size' (or in terminology to be used later, the **measure**), of ALL the sets in the space X, so as to have all desired properties such as countable additivity. One would like the 'size' of sets to be defined for a class of 'elementary' sets (for example in $\mathbb{R}$, the intervals) and also for 'nice' sets that can be built up from the elementary sets by operations such as taking unions or complements. This motivates the following which can be thought of as a criterion for a collection of 'nice' sets whose 'size' we might hope to be able to consistently define.

Definition 4.1 (Algebra)**.** A collection $\mathcal{A}$ of subsets of X is an algebra *if*

(i) $X \in \mathcal{A}$*;*

(ii) (Closure under complement) If $A \in \mathcal{A}$*, then* $X \setminus A \in \mathcal{A}$*;*

(iii) (Closure under finite union) If $A_1, \ldots, A_n \in \mathcal{A}$*, then* $\bigcup_{i=1}^n A_i \in \mathcal{A}$*.*

Much more important for our purposes is the notion of σ-algebra.

Definition 4.2 (σ-algebra)**.** A collection $\mathcal{A}$ of subsets of X is a σ-algebra *if*

(i) $X \in \mathcal{A}$*;*

(ii) (Closure under complement) If $A \in \mathcal{A}$*, then* $X \setminus A \in \mathcal{A}$*;*

(iii) (Closure under countable union) If $\{A_i\}_{i=1}^\infty \subset \mathcal{A}$*, then* $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$*.*

Remark 4.3. We could replace the requirement (i) in each definition with the requirement that $\mathcal{A}$ be non-empty. From this and (ii) and (iii) of each definition, it follows that if $E \in \mathcal{A}$, then so is E^c, and hence $X = E \cup E^c \in \mathcal{A}$ also. From this, we find $\varnothing = X^c \in \mathcal{A}$ for every algebra or σ-algebra.

Examples
(i) The *trivial σ-algebra* is given by $\mathcal{A} = \{\varnothing, X\}$.
(ii) The power set $\mathcal{P}(X)$ is always a σ-algebra.
(iii) If X is uncountable, then $\mathcal{A} = \{E \subset X \mid E \text{ or } E^c \text{ is at most countable}\}$ is a σ-algebra.
(iiv) If X is countable, then $\mathcal{A} = \{E \subset X \mid E \text{ or } E^c \text{ is finite}\}$ is an algebra but not a σ-algebra.
(Proving these facts will be a homework exercise.)

The next lemma lists some easy consequences of the definition of a σ-algebra.

Lemma 4.4. *Let $\mathcal{A}$ be a σ-algebra in X.*

(i) If $A_1, A_2 \in \mathcal{A}$, then also $A_1 \cup A_2 \in \mathcal{A}$ (and thus $\mathcal{A}$ is an algebra);

(ii) If $A_1, A_2, \dots \in \mathcal{A}$, then also $\cap_{n=1}^{\infty} A_n \in \mathcal{A}$;

Proof. (i) Suppose $A_1, A_2 \in \mathcal{A}$. Take $A_3 = A_4 = \cdots = \varnothing \in \mathcal{A}$. Then $\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \in \mathcal{A}$.
(ii) Suppose $A_1, A_2, \ldots \in \mathcal{A}$. For each $i \in \mathbb{N}$ we have $A_i^c \in \mathcal{A}$ by (ii). Then using De Morgan's law, and (iii), we have

$$(\cap_{n=1}^{\infty} A_n)^c = \cup_{n=1}^{\infty} A_n^c \in \mathcal{A}.$$

Then using (ii) again we have $\cap_{n=1}^{\infty} A_n = ((\cap_{n=1}^{\infty} A_n)^c)^c \in \mathcal{A}$. □

Roughly speaking, there are two ways we could go about trying to define a 'good' σ-algebra for us to define a measure on. We could start with a small collection of subsets (for example, intervals in $\mathbb{R}$) and define the measure on them (e.g., the length function). We could then try to extend this collection via unions, intersections, complements, etc, until we had added enough sets to the collection to have a true σ-algebra, and attempt to extend the definition of the measure onto this whole σ-algebra.

Alternatively, we can define something like a measure on all subsets (e.g., the outer measure defined on $\mathcal{P}(\mathbb{R})$ above) and then try to identify a σ-algebra of 'good' subsets on which the measure behaves the way we would like.

Both of these approaches can be extremely useful. For now, we will focus on the former, and so we introduce the notion of a *generated* σ-algebra.

First, we show a simple property of σ-algebras.

Lemma 4.5. *Suppose $\{\mathcal{A}_j\}_{j \in J}$ is a (not necessarily countable) collection of σ-algebras. Then the intersection $\mathcal{A} = \bigcap_{j \in J} \mathcal{A}_j$ is also a σ-algebra.*

Proof. We show that the three defining properties of a σ-algebra all hold.
(i) As $X \in \mathcal{A}_j$ for all $j \in J$, clearly $X \in \mathcal{A}$.
(ii) Suppose $A \in \mathcal{A}$. Then $A \in \mathcal{A}_j$ for all $j \in J$. As each $\mathcal{A}_j$ is a σ-algebra, we have $X \setminus A \in \mathcal{A}_j$ for all $j \in J$ also. Hence $A^c \in \mathcal{A}$.
(iii) Suppose that $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$. Then $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}_j$ for all $j \in J$ and so, as each $\mathcal{A}_j$ is a σ-algebra, we have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}_j$ for all $j \in J$, and so $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. □

Definition 4.6. Given any family of subsets $\mathcal{G} \subset \mathcal{P}(X)$, the σ-algebra $\sigma(\mathcal{G})$ generated by $\mathcal{G}$ *is the smallest σ-algebra containing $\mathcal{G}$ in the sense that, if $\mathcal{A}$ is any other σ-algebra containing $\mathcal{G}$, then $\sigma(\mathcal{G}) \subset \mathcal{A}$.*

Proof that the definition is well-posed. To show that such a generated σ-algebra exists, we first note that $\mathcal{P}(X)$ is a σ-algebra containing $\mathcal{G}$, and so the collection of all σ-algebras containing $\mathcal{G}$ is a non-empty (possibly not countable) collection, which we denote $\{\mathcal{A}_j\}_{j\in J}$. We let $\sigma(\mathcal{G}) = \bigcap_{j\in J} \mathcal{A}_j$ which is a σ-algebra by the previous proposition. As $\mathcal{G} \subset \mathcal{A}_j$ for all $j \in J$, we have $\mathcal{G} \subset \sigma(\mathcal{G})$.

To show that $\sigma(\mathcal{G})$ is the smallest such σ-algebra, suppose that $\mathcal{A}$ is another such σ-algebra containing $\mathcal{G}$. Then $\mathcal{A} = \mathcal{A}_j$ for some $j \in J$, and hence $\sigma(\mathcal{G}) \subset \mathcal{A}$, as required. □

Remark. The **algebra generated by** $\mathcal{C}$ is defined analogously, as the intersection of all algebras in X that contain $\mathcal{C}$. By a similar argument to the proof of the preceding theorem, this is indeed an algebra (HW).

Example. Suppose $\mathcal{C} = \{B\}$ for some $B \subset X$ with $B \neq \varnothing$ and $B \neq X$. Then

$$\sigma(\mathcal{C}) = \{\varnothing, B, B^c, X\}.$$

One particularly important σ-algebra is the Borel σ-algebra, which we now define.

Definition 4.7. The **Borel σ-algebra** in a metric space (M, d), denoted $\mathcal{B}_M$, is the σ-algebra generated by the collection of open sets. That is, setting

$$\mathcal{O} := \{U : U \subset M,\ U \text{ open}\},$$

so that $\mathcal{O}$ here denotes the collection of all open sets, we define

$$\mathcal{B}_M := \sigma(\mathcal{O}) = \text{smallest } \sigma\text{-algebra containing all open sets.}$$

If $A \subset M$ with $A \in \mathcal{B}_M$, then A is said to be a **Borel set in** M.

Let $\mathcal{I} \subset \mathcal{P}(\mathbb{R})$ be the class of all half-open intervals in $\mathbb{R}$ which are open on the left and closed on the right, that is,

$$\mathcal{I} := \{(a, b] : -\infty < a < b < \infty\} \cup \{\varnothing\}.$$

Theorem 4.8. *(a) Let $\mathcal{H}$ be the class of all closed sets in $\mathbb{R}$. Then $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{H})$.*
(b) Also $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{I})$.

Proof. (a) Suppose $A \in \mathcal{H}$, that is $A \subset \mathbb{R}$ is a closed set. Then A^c is open, so $A^c \in \mathcal{O}$. Therefore since $\mathcal{O} \subset \sigma(\mathcal{O}) = \mathcal{B}_\mathbb{R}$ we have $A^c \in \mathcal{B}_\mathbb{R}$, and hence also $A = (A^c)^c \in \mathcal{B}_\mathbb{R}$ (because $\mathcal{B}_\mathbb{R}$ is closed under complementation). Thus, for $A \in \mathcal{H}$ we have $A \in \mathcal{B}_\mathbb{R}$. Therefore $\mathcal{H} \subset \mathcal{B}_\mathbb{R}$, and hence $\sigma(\mathcal{H}) \subset \mathcal{B}_\mathbb{R}$ (since $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{O})$ is a σ-algebra).

We prove the reverse inclusion similarly. Suppose $B \in \mathcal{O}$, i.e. B is open. Then $B^c \in \mathcal{H} \subset \sigma(\mathcal{H})$ so $B = (B^c)^c \in \sigma(\mathcal{H})$ (because $\sigma(\mathcal{H})$ is closed under complementation). Thus $\mathcal{O} \subset \sigma(\mathcal{H})$, and since $\sigma(\mathcal{H})$ is a σ-algebra, therefore also $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{O}) \subset \sigma(\mathcal{H})$.

Combining the last two paragraphs gives $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{H})$.

(b) First we show that $\mathcal{I} \subset \mathcal{B}_\mathbb{R}$. For $-\infty < a < b$ we claim $(a, b] \in \mathcal{B}_\mathbb{R}$. This is because $(a, \infty) \in \mathcal{O} \subset \mathcal{B}_\mathbb{R}$, and likewise $(b, \infty) \in \mathcal{B}_\mathbb{R}$, so by Lemma 4.4 (d),

$$(a, b] = (a, \infty) \setminus (b, \infty) \in \mathcal{B}_\mathbb{R},$$

as claimed above. Hence $\mathcal{I} \subset \mathcal{B}_\mathbb{R}$, so that $\mathcal{B}_\mathbb{R}$ is a σ-algebra containing $\mathcal{I}$, and hence

$$\sigma(\mathcal{I}) \subset \mathcal{B}_\mathbb{R}. \tag{1}$$

Next we show that $\mathcal{O} \subset \sigma(\mathcal{I})$. Let $U \in \mathcal{O}$. Then for each $x \in U$ we can find $q \in \mathbb{Q}, r \in \mathbb{Q}$ with $q < x < r$ and $(q, r] \subset U$. Therefore since $\mathbb{Q} \times \mathbb{Q}$ is countable,

$$U = \cup_{(q,r)\in\mathbb{Q}\times\mathbb{Q}, q<r, (q,r]\subset U} (q, r] \in \sigma(\mathcal{I}).$$

Hence $\mathcal{O} \subset \sigma(\mathcal{I})$ and therefore $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{O}) \subset \sigma(\mathcal{I})$. Combined with (1) this gives the result. □

Remark. By the preceding theorem, we could equally well have defined the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ by $\mathcal{B}_{\mathbb{R}} := \sigma(\mathcal{H})$ or $\mathcal{B}_{\mathbb{R}} := \sigma(\mathcal{I})$. There are various other equivalent ways to define $\mathcal{B}_{\mathbb{R}}$, for example as the σ-algebra generated by the collection of all intervals of the form $(-\infty, x]$, $x \in \mathbb{R}$ (HW).

5 Measures

We often use the Greek letter μ (mu) to stand for a measure.

Definition 5.1 (Measure). Let $X \neq \varnothing$, $\mathcal{A}$ a σ-algebra on X. A function $\mu : \mathcal{A} \to [0, \infty]$ is a measure *iff*

(i) $\mu(\varnothing) = 0$;

(ii) (Countable additivity) If $A_i \in \mathcal{A}$ *for all* $i \in \mathbb{N}$ *and* $A_i \cap A_j = \varnothing$ *for* $i \neq j$, *then*

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

Definition 5.2. Let $X \neq \varnothing$, $\mathcal{A}$ a σ-algebra on X, and μ a measure. We call the pair $(X, \mathcal{A})$ a measurable space *and the triple* $(X, \mathcal{A}, \mu)$ *a* measure space.
If $\mu(X) = 1$, we say that μ is a **probability measure** and $(X, \mathcal{A}, \mu)$ is a **probability space**. In this case one often uses notation $(\Omega, \mathcal{F}, \mathbb{P})$ instead of $(X, \mathcal{A}, \mu)$.
If $\mu(X) < \infty$ we say that μ is a **finite measure**. If there exist sets $F_1, F_2, F_3, \ldots \in \mathcal{A}$ with $\cup_{i=1}^{\infty} F_i = X$ and $\mu(F_i) < \infty$ for all i, then we say that μ is a **σ-finite measure**.

Examples
(i) Let $(X, \mathcal{A})$ be any measurable space such that $\{x\} \in \mathcal{A}$ for all $x \in X$. The *counting measure* is defined by

$$\mu(A) = \begin{cases} \#\text{ (i.e. number) of elements of } A & \text{when } A \text{ is finite;} \\ +\infty & \text{when } A \text{ is infinite.} \end{cases}$$

If X is countable, then the counting measure on $(X, \mathcal{A})$ is σ-finite. Conversely, if X is *uncountable* (e.g. if $X = (0, 1]$), then the counting measure is *not* σ-finite. (In this course we mainly consider finite or σ-finite measures).

(ii) Let $(X, \mathcal{A})$ be any measurable space, $x_0 \in X$. Then the *Dirac measure* at x_0 is defined by setting, for $A \in \mathcal{A}$,

$$\mu(A) = \begin{cases} 1, & x_0 \in A, \\ 0, & x_0 \notin A. \end{cases}$$

(iii) In fact, both of these are special cases of the following situation. Let $f : X \to [0, \infty]$. Then the function defined as

$$\mu(A) = \begin{cases} \sum_{x \in A} f(x), & E \neq \varnothing, \\ 0, & A = \varnothing \end{cases}$$

is a measure. In the case that $f(x) = 1$ for all $x \in X$, this measure is the counting measure. In the case that there exists $x_0 \in X$ such that

$$f(x) = \begin{cases} 1, & x = x_0, \\ 0, & x \neq x_0, \end{cases}$$

then μ is the *Dirac measure* at x_0.
(iv) Scalar multiples of measures. If $(X, \mathcal{A}, \mu)$ is a measure space and $a \geq 0$ is a constant, then $a\mu$ (defined pointwise, i.e. $(a\mu)(A) := a\mu(A)$ for all $A \in \mathcal{A}$) is also a measure on $(X, \mathcal{A})$. [Here we need to use the convention $0 \times \infty := 0$ if $a = 0$. This convention may come up again later.]
(v) Countable sums of measures. If $(X, \mathcal{A})$ is a measurable space and $\mu_1, \mu_2, \mu_3, \ldots$ are measures on $(X, \mathcal{A})$ then $\sum_{i=1}^{\infty} \mu_i$ (defined pointwise, i.e. $(\sum_{i=1}^{\infty} \mu_i)(A) := \sum_{i=1}^{\infty} \mu_i(A)$ for all $A \in \mathcal{A}$) is a measure on $(X, \mathcal{A})$.

(HW): Check that (i)–(v) are indeed measures.

Non-example

Let X be infinite, $\mathcal{A} = \mathcal{P}(X)$ and define

$$\mu(A) = \begin{cases} 0, & A \text{ finite}, \\ \infty, & A \text{ infinite}. \end{cases}$$

The defined function μ is not a measure as it only finitely additive, not countable additive.

Theorem 5.3. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then*

(i) (Monotonicity) Let $A, B \in \mathcal{A}$ and $A \subset B$. Then $\mu(A) \leq \mu(B)$. If also $\mu(B) < \infty$ then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(ii) (Subadditivity) Let $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$. Then $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

(iii) (Continuity from below) Let $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ be such that $A_1 \subset A_2 \subset \ldots$. Then $\mu(\bigcup_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \mu(A_i)$.

(iv) (Continuity from above) Let $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ be such that $A_1 \supset A_2 \supset \ldots$ and there exists $k \in \mathbb{N}$ such that $\mu(A_k) < \infty$. Then $\mu(\bigcap_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \mu(A_i)$.

Proof. (i) Let A, B be as in the statement. Then, as $B = (B \setminus A) \cup A$, we have

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

where we have used finite additivity and non-negativity of the measure. In the case $\mu(B)$ is finite, we rearrange the equality to obtain $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(ii) Let $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$. We define $B_1 = A_1$ and, for $n \geq 2$, $B_n = A_n \setminus \bigcup_{j=1}^{n-1} A_j$. Then the sets B_n are pairwise disjoint, $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$, and clearly also $B_n \subset A_n$ for all n. Therefore, using countable additivity,

$$\mu\Big(\bigcup_{n=1}^\infty A_n\Big) = \mu\Big(\bigcup_{n=1}^\infty B_n\Big) = \sum_{n=1}^\infty \mu(B_n) \leq \sum_{n=1}^\infty \mu(A_n),$$

where we have also used monotonicity in the last step.

(iii) Let $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_1 \subset A_2 \subset \ldots$. Define $A_0 = \varnothing$. Then we have the simple identity $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty (A_n \setminus A_{n-1})$, where the latter union is over pairwise disjoint sets. Therefore, using countable additivity,

$$\begin{aligned}\mu\Big(\bigcup_{n=1}^\infty A_n\Big) &= \mu\Big(\bigcup_{n=1}^\infty (A_n \setminus A_{n-1})\Big) = \sum_{n=1}^\infty \mu(A_n \setminus A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{j=1}^n \mu(A_j \setminus A_{j-1}) = \lim_{n\to\infty} \mu\Big(\bigcup_{j=1}^n (A_j \setminus A_{j-1})\Big) = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

(iv) Suppose $A_n \supset A_{n+1}$ for all n, and $\mu(A_1) < \infty$. The idea is that we take complements and apply part (iii), but to make this work we should take complements *within a set of finite measure*, namely A_1.

Set $A = \cap_{n=1}^\infty A_n$, and for $k \in \mathbb{N}$ set $B_k := A_1 \setminus A_k$ (i.e. the complement of A_k within A_1). Then $B_k \subset B_{k+1}$ for each k, and

$$\cup_{n=1}^\infty B_n = A_1 \cap (\cup_{n=1}^\infty A_n^c) = A_1 \cap (\cap_{n=1}^\infty A_n)^c = A_1 \cap A^c = A_1 \setminus A.$$

Also, for $n \in \mathbb{N}$ we have $\mu(A_1) < \infty$ and $A_n \subset A_1$ so $\mu(B_n) = \mu(A_1) - \mu(A_n)$ by part (i). Similarly $A \subset A_1$ so $\mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$. Hence by (iii) and the Algebra of Limits, as $n \to \infty$ we have

$$\begin{aligned}\mu(A_n) &= \mu(A_1) - \mu(B_n) \\ &\to \mu(A_1) - \mu(A_1 \setminus A) = \mu(A).\end{aligned}$$

□

Now that we have seen the definition of a measure and some examples, we turn our attention to a particularly nice class of measures and show how they can be obtained from other measures. These measures are those called complete. Essentially, complete measures are those that can measure all subsets of sets with measure zero.

Definition 5.4. Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that a set $A \subset \mathcal{A}$ is null *if* $\mu(A) = 0$ *(equivalently, we say it has* measure zero*).*
A property (P) is said to hold almost everywhere *on a measure space (usually abbreviated as a.e.) if the set where (P) does not hold is of measure zero.*

For example, two functions f and g are equal a.e. if the set $\{x \mid f(x) \neq g(x)\}$ is of measure zero.

Definition 5.5. Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that μ is complete *if $\mathcal{A}$ contains every subset of every null set.*

Theorem 5.6 (Completing a measure)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space. Define*

$$\mathcal{N} = \{N \in \mathcal{A} \,|\, \mu(N) = 0\}, \qquad \overline{\mathcal{A}} = \{A \cup B \,|\, A \in \mathcal{A},\ B \subset N \text{ for some } N \in \mathcal{N}\}.$$

Then $\overline{\mathcal{A}}$ is a σ-algebra and there exists a unique extension of μ, $\bar{\mu}$, such that $(X, \overline{\mathcal{A}}, \bar{\mu})$ is a complete measure space.

Proof. Non-examinable. Step 1: We first show that $\overline{\mathcal{A}}$ is a σ-algebra.
To show that $\overline{\mathcal{A}}$ is closed under complements, take $(A \cup B) \in \overline{\mathcal{A}}$ such that $A \in \mathcal{A}$, $B \subset N$ for some $N \in \mathcal{N}$. If $A \cap N \neq \varnothing$, then we define a new null set $N' = N \setminus (A \cap N)$ and $B' = B \cap N'$ and observe

$$A \cup B = A \cup \underbrace{(B \cap (N \setminus N'))}_{\subset A} \cup \underbrace{(B \cap (N'))}_{= B'} = A \cup B'.$$

Then, as we now have $A \cap B = \varnothing$ we decompose the set as

$$\begin{aligned} A \cup B &= \big((A \cap N^c) \cup (N \cap N^c)\big) \cup \big((A \cap B) \cup (N \cap B)\big) \\ &= \big((A \cup N) \cap N^c\big) \cup \big((A \cup N) \cap B\big) \\ &= (A \cup N) \cap (N^c \cup B) \end{aligned}$$

so that $(A \cup B)^c = (A \cup N)^c \cup (N \cap B^c)$. As $\mathcal{A}$ is a σ-algebra, we have $(A \cup N)^c \in \mathcal{A}$, while clearly $(N \cap B^c) \subset N \in \mathcal{N}$, and we are done.

To show that $\overline{\mathcal{A}}$ is closed under countable unions, we take $\{A_j \cup B_j\}_{j=1}^{\infty}$ such that, for all $j \in \mathbb{N}$, we have $A_j \in \mathcal{A}$, $B_j \subset N_j$ for some $N \in \mathcal{N}$. Let $N = \bigcup_{j=1}^{\infty} N_j$ and observe $\mu(N) \leq \sum_{j=1}^{\infty} \mu(N_j) = 0$, so that

$$\bigcup_{j=1}^{\infty} B_j \subset \bigcup_{j=1}^{\infty} N_j = N \in \mathcal{N}.$$

Now as $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$, we have

$$\bigcup_{j=1}^{\infty} (A_j \cup B_j) = \Big(\bigcup_{j=1}^{\infty} A_j\Big) \cup \Big(\bigcup_{j=1}^{\infty} B_j\Big) \in \overline{\mathcal{A}}.$$

Step 2: We now construct the extension measure $\bar{\mu}$.
First, note that if $\bar{\mu}$ is an extension of μ, so that $\bar{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$, then if $B \subset N \in \mathcal{N}$, we have

$$\mu(A) = \bar{\mu}(A) \leq \bar{\mu}(A \cup B) \leq \bar{\mu}(A \cup N) = \mu(A \cup N) \leq \mu(A) + \mu(N) = \mu(A).$$

This firstly guarantees uniqueness of the extension, and also forces us to define

$$\bar{\mu}(A \cup B) = \mu(A) \text{ for all } A \cup B \in \overline{\mathcal{A}} \text{ with } A \in \mathcal{A},\ B \subset N \in \mathcal{N}.$$

To verify that this extension is well-defined, we take any set $E \in \overline{\mathcal{A}}$ and suppose that $E = A_1 \cup B_1 = A_2 \cup B_2$ where $A_i \in \mathcal{A}$, $B_i \subset N_i \in \mathcal{N}$, $i = 1, 2$. We need to prove that $\mu(A_1) = \mu(A_2)$ to show that $\bar{\mu}$ is well-defined. Now note that $A_1 \subset A_2 \cup B_2 \subset A_2 \cup N_2$, so that

$$\mu(A_1) \leq \mu(A_2 \cup N_2) \leq \mu(A_2) + \mu(N_2) = \mu(A_2)$$

and, similarly, $\mu(A_2) \leq \mu(A_1)$. Thus the extension is well-defined.
Exercise: Show that $\bar{\mu}$ satisfies the definition of a measure and is complete. □

6 Outer measures, Pre-measures, and Caratheodory's Extension Theorem

6.1 Outer Measures and Carathéodory's Lemma

Next, we turn out attention to the construction of measures. As we saw in Section 3, one convenient approach to constructing measures is to start with a collection of sets that we want to define the size of (such as intervals, rectangles, etc) and use this to define an outer measure of all other sets. This is in fact a well-defined approach, and so we first make precise the notion of outer measure.

Definition 6.1. Let $X \neq \varnothing$. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is an outer measure *if it satisfies*

(i) $\mu^*(\varnothing) = 0$,

(ii) $A \subset B \implies \mu^*(A) \leq \mu^*(B)$,

(iii) $\mu^*\big(\bigcup_{j=1}^{\infty} A_j\big) \leq \sum_{j=1}^{\infty} \mu^*(A_j)$.

Recall that these are exactly the properties that we showed for the Lebesgue outer measure on $\mathbb{R}$ in addition to the $\mathbb{R}$-specific properties (the $\mathbb{R}$-specific properties we mean are the translation invariance and the identity for the measure of intervals, neither of which generalises to an abstract measure space).

Following this analogy to the construction of Section 3, we now show how to obtain an outer measure from a pre-determined notion of size of a sub-collection of sets.

Proposition 6.2. *Let* $\xi \subset \mathcal{P}(X)$*,* $\varnothing \in \xi$*,* $X \in \xi$*. Let* $\rho : \xi \to [0, \infty]$ *be such that* $\rho(\varnothing) = 0$*. Then there exists an outer measure* μ^* *on* X *defined by, for all* $A \subset X$*,*

$$\mu^*(A) = \inf \Big\{ \sum_{j=1}^{\infty} \rho(E_j) \,\big|\, E_j \in \xi,\ A \subset \bigcup_{j=1}^{\infty} E_j \Big\}.$$

In this proposition, the key point in the assumptions on ξ is that, for any $A \subset X$, we can find at least one cover of A by elements of ξ because $X \in \xi$. We often refer to ξ as the covering class, and, when certain additional properties are satisfied (see Section 6.3), we call ρ the pre-measure.

The proof is formally identical to the proof of Proposition 3.3(i)–(iii). For completeness, we include it here anyway.

Proof. We need to check the three defining properties of an outer measure. Observe first that, by definition, $\mu^*(\varnothing) = 0$.
Property (ii) is also essentially trivial: given $A \subset B$, every cover of B is also a cover of A, so that the infimum in the definition of $\mu^*(A)$ is an infimum over a larger class of covers than that for $\mu^*(B)$.
(iii) Let $\varepsilon > 0$. For each A_j, we find a cover $\{E_{j,k}\}_{k=1}^{\infty}$ such that $A_j \subset \bigcup_{k=1}^{\infty} E_{j,k}$, all $E_{j,k} \in \xi$, and $\sum_{k=1}^{\infty} \rho(E_{j,k}) \leq \mu^*(A_j) + \frac{\varepsilon}{2^j}$. If we do not have $\mu^*(A_j) < \infty$ for all j, there is nothing to prove (the inequality holds trivially), so suppose that this holds.

Then $\bigcup_{j=1}^{\infty} A_j \subset \left(\bigcup_{j,k=1}^{\infty} E_{j,k}\right)$ (and note that the collection $\{E_{j,k}\}_{j,k=1}^{\infty}$ is countable), so that

$$\mu^*\Big(\bigcup_{j=1}^{\infty} A_j\Big) \leq \sum_{j,k=1}^{\infty} \rho(E_{j,k}) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \rho(E_{j,k}) \leq \sum_{j=1}^{\infty}\Big(\mu^*(A_j) + \frac{\varepsilon}{2^j}\Big) = \sum_{j=1}^{\infty} \mu^*(A_j) + \varepsilon.$$

As this holds for all $\varepsilon > 0$, we conclude the proof. □

As with the outer Lebesgue measure, this proposition gives a useful tool for constructing outer measures. However, as it stands, we need to take care with how we choose both ξ and ρ in order to obtain a measure that is actually useful for anything. You will recall from Section 3 that we worked quite hard to show that the outer measure agreed with the pre-measure on intervals ($|I|$ corresponds to ρ in this notation, and such objects are sometimes called pre-measures in the literature). If we do not have that $\mu^*(E) = \rho(E)$ for $E \in \xi$, it is not at all clear that μ^* is doing anything useful!

This bad scenario can, of course, occur.
Example
Let $X = \mathbb{R}$ and consider the covering class $\xi = \{\varnothing, [0,2], [2,4], [1,3], \mathbb{R}\}$. Let ρ be given as follows:

$$\rho(\varnothing) = 0, \quad \rho([0,2]) = 3, \quad \rho([2,4]) = 1, \quad \rho([1,3]) = 5, \quad \rho(\mathbb{R}) = \infty.$$

Then the outer measure μ^* given by Proposition 6.2 satisfies $\mu^*([1,3]) \leq 4$ because $[1,3] \subset [0,2] \cup [2,4]$ and $\rho([0,2]) + \rho([2,4]) = 4$ (actually $\mu^*([1,3]) = 4$ in this example, but the inequality ≤ 4 suffices for our purposes). Therefore $\mu^*([1,3]) < \rho([1,3])$ and μ^* does not extend ρ.

Given an outer measure, we now want to know how to construct a measure from it. Again, the intuition in the case of the Lebesgue measure on $\mathbb{R}$ is our guide, and so we come to the definition of measurability in order to construct a σ-algebra of measurable sets.

Definition 6.3. Let $X \neq \varnothing$ and suppose μ^* is an outer measure on X. We say that a set $A \subset X$ is *μ^*-measurable* (or simply *measurable* if no confusion can arise) provided

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset X.$$

Note that, from the definition, we always have X and $\varnothing$ are measurable. Moreover, if E is measurable, then so is E^c.

One way of interpreting this condition is to think of the set E as a set which we use to test A. The definition entails that A 'splits' every test set E well (in some sense) with respect to the outer measure μ^*. We now need to know that the collection of measurable sets has enough desirable properties to be useful, specifically, that the collection of measurable sets forms a σ-algebra. This is the content of Carathéodory's Theorem, which we will come to soon, but we first return to the example of the Lebesgue outer measure.

Lemma 6.4. *Let μ^* be the Lebesgue outer measure on $\mathbb{R}$. If $a \leq b$, then $I = (a,b)$, $[a,b)$, $(a,b]$, $[a,b]$ are all μ^*-measurable.*

Proof. consider the case $I = (a,b)$.

Let $E \subset \mathbb{R}$. We need to show $\mu^*(E) \geq \mu^*(E \cap I) + \mu^*(E \cap I^c)$. Take an open cover of E by intervals $\{I_k\}_{k=1}^{\infty}$ and observe that, for each k, the intersection

- $I \cap I_k$ is either $\varnothing$ or an open interval;
- $I^c \cap I_k$ is either $\varnothing$, one interval or two intervals (not necessarily open).

Let $\varepsilon > 0$. For each k, we take three open intervals, I_k', I_k'', I_k''' such that

$$(I \cap I_k) \subset I_k', \qquad (I^c \cap I_k) \subset (I_k'' \cup I_k''')$$

and also

$$|I_k'| + |I_k''| + |I_k'''| \leq |I_k| + \frac{\varepsilon}{2^k}.$$

As $\{I_k\}_{k=1}^\infty$ is an open cover of E, we observe

$$(E \cap I) \subset \bigcup_{k=1}^\infty I_k', \qquad (E \cap I^c) \subset \bigcup_{k=1}^\infty (I_k'' \cup I_k''').$$

Then

$$\mu^*(E \cap I) + \mu^*(E \cap I^c) \leq \sum_{k=1}^\infty (|I_k'| + |I_k''| + |I_k'''|) \leq \sum_{k=1}^\infty (|I_k| + \frac{\varepsilon}{2^k}) = \sum_{k=1}^\infty |I_k| + \varepsilon.$$

Now taking the infimum on the right over all covers of E by open intervals, we obtain

$$\mu^*(E \cap I) + \mu^*(E \cap I^c) \leq \mu^*(E) + \varepsilon,$$

and hence we conclude. □

We now provide the general theorem that ensures that the collection of measurable sets gives the 'right' σ-algebra on which to define a measure.

Theorem 6.5 (Carathéodory's Lemma)**.** *Let $X \neq \varnothing$, μ^* an outer measure on X. Then*

(i) The collection $\mathcal{A}$ of measurable sets is a σ-algebra;

(ii) The restriction μ of μ^* to $\mathcal{A}$ is a complete measure.

Non-examinable. (i) We show $\mathcal{A}$ is a σ-algebra by showing that it is closed under complements and countable unions. That $\mathcal{A}$ is closed under complements is trivial from the definition of measurability and the fact that $(A^c)^c = A$.
Step 1: We show that $\mathcal{A}$ is closed under finite unions.
Let $A, B \in \mathcal{A}$, $E \subset X$. As $A \in \mathcal{A}$, we have $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$. As $B \in \mathcal{A}$, each term in this identity extends as

$$\mu^*(E) = \mu^*\big((E\cap A)\cap B\big) + \mu^*\big((E\cap A)\cap B^c\big) + \mu^*\big((E\cap A^c)\cap B\big) + \mu^*\big((E\cap A^c)\cap B^c\big). \quad (1)$$

Noting that the last term on the right is equal to $\mu^*(E \cap (A \cup B)^c)$, we try to relate the first three terms to $\mu^*(E \cap (A \cup B))$. Observe that we may decompose $A \cup B$ as the disjoint union

$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B),$$

so that

$$E \cap (A \cup B) = \big(E \cap (A \cap B)\big) \cup \big(E \cap (A \cap B^c)\big) \cup \big(E \cap (A^c \cap B)\big)$$

and, as μ^* is an outer measure, by monotonicity,

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

Substituting this now into (1), we have obtained

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

As the reverse inequality is always true by sub-additivity, we have obtained

$$\mu^*(E) = \mu^*\big(E \cap (A \cup B)\big) + \mu^*\big(E \cap (A \cup B)^c\big),$$

and so $A \cup B$ is measurable.

By induction, $\mathcal{A}$ is closed under all finite unions and, as it is closed under complements, is also closed under finite intersections.

Step 2: We show $\mathcal{A}$ is closed under countable unions. In fact, as we have already shown that $\mathcal{A}$ is an algebra, it is enough to show that $\mathcal{A}$ is closed under countable *disjoint* unions, as the remaining case follows in the usual way.
Therefore, we take $\{A_j\}_{j=1}^{\infty} \subset \mathcal{A}$, $A_i \cap A_j = \varnothing$ for $i \neq j$, and we define

$$B_n = \bigcup_{j=1}^{n} A_j, \qquad B = \bigcup_{j=1}^{\infty} A_j.$$

Now $B_n \cap A_n = A_n$ and $B_n \cap A_n^c = B_{n-1}$. So, by measurability of A_n,

$$\mu^*(E \cap B_n) = \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) = \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}).$$

Inductively,

$$\mu^*(E \cap B_n) = \sum_{j=1}^{n} \mu^*(E \cap A_j).$$

As $\mathcal{A}$ is an algebra, we know that B_n, $B_n^c \in \mathcal{A}$, and hence

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) + \mu^*(E \cap B^c) = \sum_{j=1}^{n} \mu^*(E \cap A_j) + \mu^*(E \cap B^c),$$

where we have used that $B^c \subset B_n^c$ and the previous identity.

As $\sum_{j=1}^{n} \mu^*(A_j)$ is monotone increasing, its limit as $n \to \infty$ exists (but may be infinite) and so, as the other terms in the inequality are all independent of n, we may pass to the limit to obtain

$$\mu^*(E) \geq \sum_{n=1}^{\infty} \mu^*(E \cap A_n) + \mu^*(E \cap B^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c)$$

where we have used sub-additivity in the second inequality as $E \cap B \subset \bigcup_{n=1}^{\infty}(E \cap A_n)$. As the reverse inequality is trivial, we conclude that $B \in \mathcal{A}$, and so $\mathcal{A}$ is a σ-algebra.

(ii) We need to show that the restriction of μ^* to $\mathcal{A}$ is a complete measure. Clearly $\mu^*(\varnothing) = 0$ already, so we require countable additivity.

Adopting the notation from Step 2, we let A_j, B be as in that step, and note that we showed there

$$\mu^*(E) \geq \sum_{n=1}^{\infty} \mu^*(E \cap A_n) + \mu^*(E \cap B^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) = \mu^*(E),$$

where the final equality follows from $B \in \mathcal{A}$. Therefore all inequalities here are actually equalities, and so

$$\sum_{n=1}^{\infty} \mu^*(E \cap A_n) = \mu^*(E \cap B).$$

Now taking $E = B$, as all the A_n are pairwise disjoint, $B \cap A_n = A_n$, and so this becomes

$$\sum_{n=1}^{\infty} \mu^*(A_n) = \mu^*(B)$$

as required. Thus the restriction μ is a measure.

It remains only to show that μ is complete. Suppose that $\mu(A) = 0$ and $B \subset A$. For $\mu(A)$ to be defined, we must have $A \in \mathcal{A}$. By subadditivity of μ^*, we also have $\mu^*(B) \leq \mu^*(A) = \mu(A) = 0$, so $\mu^*(B) = 0$. We therefore need to show that $\mu^*(B) = 0$ implies $B \in \mathcal{A}$.

Let $E \subset X$ and consider $\mu^*(E \cap B) \leq \mu^*(B) = 0$. Then

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

by subadditivity, and hence we obtain that B is measurable. □

With Carathéodory's Theorem, we now have a full recipe for defining measures on a set X:
Step 1: Define a collection of sets ξ and a pre-measure ρ satisfying some mild compatibility conditions.
Step 2: Use Proposition 6.2 to extend the pre-measure to an outer measure on X.
Step 3: Use Carathéodory's Lemma, Theorem 6.5, to find the σ-algebra of measurable sets and obtain the complete measure on this σ-algebra.

As a specific example, we have the following.

Definition 6.6. The *Lebesgue measure* on $\mathbb{R}$ is defined as the restriction of the Lebesgue outer measure to the σ-algebra of measurable sets. It is commonly denoted by μ, m, or λ^1. The σ-algebra of Lebesgue measurable sets is usually denoted Σ or Σ_1.

You may well wonder at this point whether we could equally well have just restricted the Lebesgue outer measure to the Borel σ-algebra to produce this measure. As the Borel σ-algebra is generated by the open sets, and the open sets are all measurable, clearly $\mathcal{B} \subset \Sigma$. However, it turns out that this inclusion is strict: there are Lebesgue measurable sets that are not Borel sets, and the restriction of the Lebesgue measure to the Borel σ-algebra is not complete.

6.2 Pre-measures and Carathéodory's Theorem

Having seen that we can always obtain a measure from an *outer* measure by restricting to the measurable sets, the next obvious question is how can we get hold of 'good' outer measures? After all, an outer measure is defined on the whole power set, i.e. on all subsets of our space. We would like a general recipe for generating an outer measure from something defined on a simpler collection of sets, like we used for the Lebesgue measure and in Proposition 6.2. However, we have seen from examples that this proposition can lead to bizarre objects if we are not careful. The goal of the following definitions is to define the 'right' way of constructing outer measures.

Definition 6.7. A collection $\mathcal{S} \subset \mathcal{P}(X)$ is a *semi-ring* or *semi-algebra* on X if
(i) $\varnothing \in \mathcal{S}$;
(ii) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$;
(iii) If $A, B \in \mathcal{S}$, then there exist finitely many disjoint $D_1, \ldots, D_N \in \mathcal{S}$ such that $A \setminus B = \bigcup_{j=1}^{N} D_j$.

The definition of semi-algebra is similar to the earlier definition of an algebra (Definition 4.1), except that (iv) above is a weaker version of the 'closed under complementation' condition for an algebra. It can be shown that every algebra is a semi-algebra (HW).

When we discussed the construction of outer measures using Carathéodory's Lemma (Theorem 6.5 in the previous section, we mentioned that there are certain compatibility conditions that we should ask for on our initial assignment of values in order to guarantee that the constructed outer measure really does extend the initial set function (see the discussion and example after Proposition 6.2). These compatibility conditions are precisely those that define a pre-measure:

Definition 6.8. Let $\mathcal{S} \subset \mathcal{P}(X)$ be a semi-ring. A function $\rho : \mathcal{S} \to [0, \infty]$ is called a *pre-measure* if it satisfies:
(i) $\rho(\varnothing) = 0$;
(ii) Whenever $\{R_j\}_{j=1}^{\infty} \subset \mathcal{S}$ are pairwise disjoint and satisfy $\bigcup_{j=1}^{\infty} R_j \in \mathcal{S}$ also, then $\rho(\bigcup_{j=1}^{\infty} R_j) = \sum_{j=1}^{\infty} \rho(R_j)$.

As $\mathcal{S}$ is not a σ-algebra, it may not always be the case that countable unions are contained in $\mathcal{S}$, but the second condition in the definition of pre-measure guarantees that whenever such a union is contained, then the pre-measure is compatible with the requirement of σ-additivity.

You may have noticed that we are not imposing the usual monotonicity condition. This actually follows directly from the definitions of semi-ring and pre-measure: Suppose $A \subset B$. Then there exists a finite, disjoint set $D_1, \ldots, D_N \in \mathcal{S}$ such that $B = A \cup \bigcup_{j=1}^{N} D_j$. Then, using condition (ii) of the pre-measure,

$$\rho(B) = \rho(A) + \sum_{j=1}^{N} \rho(D_j) \geq \rho(A).$$

Theorem 6.9 (Carathéodory's Extension Theorem)**.** *Let $\mathcal{S} \subset \mathcal{P}(X)$ be both a covering class and a semi-ring and suppose that $\rho : \mathcal{S} \to [0, \infty]$ is a pre-measure. Then there exists a measure μ defined on the measurable space $(X, \sigma(\mathcal{S}))$ such that μ and ρ are equal on $\mathcal{S}$. More precisely, denote by μ^* the outer measure constructed from Proposition 6.2. Then ρ and μ^* agree on $\mathcal{S}$ and every set $S \in \mathcal{S}$ is μ^*-measurable.*

Sketch proof. The proof that ρ and μ^* agree on $\mathcal{S}$ is formally identical to that of Lemma 3.3(iv), while the fact that every $S \in \mathcal{S}$ is μ^*-measurable follows as in Lemma 6.4. We therefore leave it as an exercise (that you should do!) In fact, an inspection of those proofs will show you that the conditions that are used are precisely those contained in the definition of a pre-measure.

To conclude the Theorem, we now simply apply Carathéodory's Lemma, Theorem 6.5 to deduce (a) that the collection of μ^*-measurable sets is a σ-algebra, and hence contains $\sigma(\mathcal{S})$; and (b) that the restriction of μ^* to $\sigma(\mathcal{S})$ is a measure. $\square$

Another interesting question that we might ask is whether there is another way of extending the pre-measure, either to a measure on some σ-algebra or as an outer measure on the full power set. This question of uniqueness of the extension is a subtle one, but we can give some kind of an answer with the following general theorem (which is sometimes stated as part of Carathéodory's Extension Theorem).

Theorem 6.10. *Let $\mathcal{S}$ be a semi-ring on X and ρ a σ-finite pre-measure. Let μ^* be the outer measure defined from ρ through Proposition 6.2 and let Σ be the set of μ^*-measurable sets. Suppose that $\tilde{\mu}^* : \mathcal{P}(X) \to [0, \infty]$ is another outer measure such that $\tilde{\mu}^* = \rho$ on $\mathcal{S}$. Then we necessarily have that $\mu^* = \tilde{\mu}^*$ on Σ.*

Note that Theorem 6.9 already guarantees that $\mu^* = \rho$ on $\mathcal{S}$ as ρ was a pre-measure.

The σ-finiteness in this theorem is necessary, as is shown by the following example:

Example

Let $X = \mathbb{R}$ and consider the semi-algebra (and covering class) $\mathcal{S} = \{\varnothing, \mathbb{R}\}$ and the pre-measure on it defined by $\rho(\varnothing) = 0$, $\rho(\mathbb{R}) = \infty$. Then the outer measure μ^* defined through Proposition 6.2 produces

$$\mu^*(A) = \begin{cases} 0 & \text{if } A = \varnothing, \\ \infty & \text{if } A \neq \varnothing. \end{cases}$$

However, we can also extend ρ via the counting outer measure

$$\tilde{\mu}^*(A) = \begin{cases} \#A & \text{if } A \text{ is finite,} \\ \infty & \text{else.} \end{cases}$$

A third possible extension is λ^1, the one-dimensional Lebesgue measure on $\mathbb{R}$. The σ-algebra of measurable sets in each of the first two cases is the whole power set, $\mathcal{P}(\mathbb{R})$, yet the outer measures do not agree on this set.

On the other hand, the strength of the uniqueness statement in this theorem is limited to the σ-algebra of measurable sets which, in general, may be very small, as is shown by the following example.

Example

Let $X = [0, 1]$ and consider the semi-ring (and covering class) $\mathcal{S} = \{\varnothing, [0, 1]\}$ and the pre-measure on it defined by $\rho(\varnothing) = 0$, $\rho([0, 1]) = 1$. Then the Carathéodory extension μ^* defined through Proposition 6.2 produces the outer measure

$$\mu^*(A) = \begin{cases} 0, & \text{if } A = \varnothing, \\ 1, & \text{if } A \neq \varnothing. \end{cases}$$

Another outer measure that extends ρ is the 1-dimensional Lebesgue measure λ^1 restricted to the interval $[0, 1]$. Note that the σ-algebra of μ^*-measurable sets is just $\Sigma = \{\varnothing, [0, 1]\}$. Indeed, whenever $A \notin \{\varnothing, [0, 1]\}$, choose B that has non-empty intersection with both A and $[0, 1] \setminus A$: then $\mu^*(B) = 1$, $\mu^*(B \cap A) = 1$ and $\mu^*(B \setminus A) = 1$, so A is not μ^*-measurable. Note, however, that λ^1 has the same action as μ^* on the σ-algebra $\Sigma = \{\varnothing, [0, 1]\}$, so that this does not violate the uniqueness theorem!

The next natural thing we would like to consider is the Lebesgue measure in more than one dimension, i.e. on $\mathbb{R}^n$. The theory we have just developed makes this very straightforward.

6.3 Lebesgue measure on $\mathbb{R}^n$

To construct the Lebesgue measure on $\mathbb{R}^n$, we follow a very similar approach to that taken to construct the Lebesgue measure on $\mathbb{R}$. In place of the set of open intervals, we turn to the space of half-open rectangles $\mathcal{I}^n = \{[a,b) \mid a,b \in \mathbb{R}^n\}$, where the notation

$$[a,b) = [a_1,b_1) \times [a_2,b_2) \times \cdots \times [a_n,b_n), \text{ where } a_j \le b_j,\ j = 1,\dots,n.$$

We leave the verification that $\mathcal{I}^n$ is a semi-ring as a (non-examinable) exercise (compare the proof of Lemma 6.4 for an idea of the third bullet point in the definition).

We begin by assigning a function $\lambda : \mathcal{I}^n \to [0,\infty)$ which captures the 'natural' volume by setting

$$\lambda([a,b)) = \prod_{j=1}^{n} (b_j - a_j), \tag{2}$$

where if any $b_j = a_j$, we note that the rectangle is empty and we have $\lambda([a,b)) = \lambda(\varnothing) = 0$, while in any other case each factor is positive.

Lemma 6.11. *The function $\lambda : \mathcal{I}^n \to [0,\infty)$ defined by (2) is a pre-measure on $\mathcal{I}^n$.*

We omit the proof, which is based on an induction in n.

We now have the Lebesgue measure on $\mathbb{R}^n$.

Definition 6.12. Let μ^* be the outer measure on $\mathbb{R}^n$ given by Theorem 6.9 applied to λ on $\mathcal{I}^n$. Then the restriction of μ^* to the σ-algebra Σ_n of measurable sets is the Lebesgue measure on $\mathbb{R}^n$, which we denote λ^n.

We would like to know now that the σ-algebra of Lebesgue measurable sets on $\mathbb{R}^n$ includes the Borel σ-algebra (it does). To get there, though, we need first to turn to the theory of product spaces.

7 Product measures

Throughout this section, we let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces.

Recall that for any two sets A, B, the *cartesian product* (or just *product*) of A and B is the set $A \times B := \{(x,y) : x \in A, y \in B\}$.

We would like to define a *product measure* φ on an appropriate σ-algebra $\mathcal{F}$ in the space $X \times Y$, with the property that

$$\varphi(A \times B) = \mu(A)\nu(B), \quad \text{for all } A \in \mathcal{A},\ B \in \mathcal{B}. \tag{1}$$

where we use the convention $\infty \times 0 = 0 \times \infty = 0$ if needed for the product on the right. In this section we shall prove that this can indeed be done.

One reason this is important is the special case where $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu) = (\mathbb{R}, \Sigma_1, \lambda_1)$ (the real line with the Lebesgue σ-algebra and Lebesgue measure). By taking A and B both to be intervals, the product measure φ in this case has the property that $\varphi(R)$ equals the area of R for any rectangular set of the form $R = (a,b] \times (c,d]$ (since such a set has area $(b-a)(d-c)$). The measure φ in this special case coincides with the *two-dimensional Lebesgue measure* λ_2. It extends our notion of area for rectangles to a larger class of sets. This is important, for example, if we want to define integrals of functions via 'area under the curve'.

Definition 7.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A set of the form $A \times B \subset X \times Y$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$, is called a **measurable rectangle**. We denote by $\mathcal{R}$ the collection of all measurable rectangles, that is, we set

$$\mathcal{R} := \{A \times B : A \in \mathcal{A},\, B \in \mathcal{B}\}.$$

We write $\mathcal{A} \otimes \mathcal{B}$ for $\sigma(\mathcal{R})$, the σ-algebra in $X \times Y$ generated by the class of measurable rectangles, and call it the **product σ-algebra** (of $\mathcal{A}$ and $\mathcal{B}$).

Caution: some authors write $\mathcal{A} \times \mathcal{B}$ rather than $\mathcal{A} \otimes \mathcal{B}$ for the product σ-algebra. However, $\mathcal{A} \otimes \mathcal{B}$ is *not* a Cartesian product of the collections $\mathcal{A}$ and $\mathcal{B}$.

Note that it follows from the definition that

$$(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F),$$
$$(A \times B)^c = (X \times B^c) \cup (A^c \times B) = (A \times B^c) \cup (A^c \times Y).$$

Given we want our product measure φ to have property (1), we would like $\varphi(S)$ to be defined whenever S is a measurable rectangle, i.e. $S \in \mathcal{R}$. Therefore our σ-algebra $\mathcal{F}$ on $X \times Y$ should include all such S. It is simplest to take the *smallest* σ-algebra containing $\mathcal{R}$, namely the product σ-algebra $\mathcal{A} \otimes \mathcal{B}$.

In general, it will not be the case that every set in the product σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is of the form $A \times B$ (or even a disjoint union of such sets). However, if $\mathcal{R}$ is a semi-ring in the sense of Definition 6.7, then we can try to define a pre-measure on $\mathcal{R}$ (recall Definition 6.8) and apply Carathéodory's Theorem to extend it to a measure on $\mathcal{A} \otimes \mathcal{B}$.

Proposition 7.2. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measurable spaces. Then the collection $\mathcal{R} = \{A \times B \,|\, A \in \mathcal{A},\, B \in \mathcal{B}\}$ is a semi-ring on $X \times Y$.*

Proof. It should be clear that $\varnothing = \varnothing \times \varnothing \in \mathcal{R}$.
We have already noted that if $A, E \in \mathcal{A}$, $B, F \in \mathcal{B}$, then $(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F) \in \mathcal{R}$ as required.
The final point to check is the complementation. We need to show that if two rectangles $R_1 = A_1 \times B_1$ and $R_2 = A_2 \times B_2 \in \mathcal{R}$, then there exist finitely many disjoint $T_1, \ldots, T_N \in \mathcal{A} \times \mathcal{B}$ such that $R_1 \setminus R_2 = \bigcup_{j=1}^{N} T_j$. We use the identity

$$\begin{aligned} R_1 \setminus R_2 &= \{(x, y) \,|\, x \in A_1,\, y \in B_1\} \setminus \{(x, y) \,|\, x \in A_2,\, y \in B_2\} \\ &= \{(x, y) \,|\, x \in A_1 \setminus A_2,\, y \in B_1\} \cup \{(x, y) \,|\, x \in A_1 \cap A_2,\, y \in B_1 \setminus B_2\} \\ &= \big((A_1 \setminus A_2) \times B_1\big) \cup \big((A_1 \cap A_2) \times (B_1 \setminus B_2)\big) \end{aligned}$$

which is a disjoint union of sets in $\mathcal{R}$. □

Following the intuition from the Lebesgue measure, we are going to define the function $\rho : \mathcal{R} \to [0, \infty]$ by

$$\rho(A \times B) = \mu(A)\nu(B) \qquad \text{for all } A \times B \in \mathcal{R}.$$

The next proposition shows that this is a pre-measure.

Proposition 7.3. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces, let $\mathcal{R} = \{A \times B \,|\, A \in \mathcal{A},\, B \in \mathcal{B}\}$, and define*

$$\rho(A \times B) = \mu(A)\nu(B) \qquad \textit{for all } A \times B \in \mathcal{R}.$$

Then the function ρ is a pre-measure on the semi-ring $\mathcal{R}$.

There are several ways of proving this proposition, but it is most convenient to use tools of integration that we have not yet defined (in particular, the proof becomes much easier once we have the Monotone Convergence Theorem). For this reason, the proof is omitted.

Therefore we may apply the Carathéodory method to construct a measure π on $\mathcal{A}\otimes\mathcal{B}$. The measure π is the restriction of the outer measure

$$\pi^*(E) = \inf\Big\{\sum_{j=1}^{\infty}\rho(A_j\times B_j)\,|\,E\subset\bigcup_{j=1}^{\infty}(A_j\times B_j),\ A_j\in\mathcal{A},\ B_j\in\mathcal{B}\Big\}$$

constructed as in Proposition 6.2 to the collection of π^*-measurable sets. By Proposition 6.9, every set in $\mathcal{S}$ is π^*-measurable, π agrees with ρ on $\mathcal{S}$ and so, as $\mathcal{A}\otimes\mathcal{B}$ is generated by $\mathcal{S}$, π is defined on all of $\mathcal{A}\otimes\mathcal{B}$.

Observe that if both μ and ν are σ-finite, then so is π as we may write

$$X = \bigcup_{j=1}^{\infty} A_j, \qquad Y = \bigcup_{j=1}^{\infty} B_j$$

where all $\mu(A_j) < \infty$, $\nu(B_j) < \infty$ and then $X\times Y = \bigcup_{j,k=1}^{\infty} A_j\times B_k$ and $\pi(A_j\times B_k) = \mu(A_j)\nu(B_k) < \infty$.

In the case that both spaces are σ-finite, Theorem 6.10 guarantees the uniqueness of the product measure in the sense that any other measure $\tilde{\pi}$ that agrees with π on $\mathcal{S}$ must agree with π on $\mathcal{A}\otimes\mathcal{B}$ (actually on all of the π-measurable sets).

Definition 7.4. Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be measure spaces. The measure π defined as in the above construction is called the product measure *on* $\mathcal{A}\otimes\mathcal{B}$ *and denoted* $\pi=\mu\otimes\nu$. *The* product measure space *is the triple* $(X\times Y,\mathcal{A}\otimes\mathcal{B},\mu\otimes\nu)$.

Corollary 7.5. *Whenever* $k+d=n$*, the Lebesgue measure on* $\mathbb{R}^n$*, restricted to the Borel* σ*-algebra, is the product measure of the Lebesgue measures on* $\mathbb{R}^k$ *and* $\mathbb{R}^d$*. Specifically,*

$$(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\lambda^n) = (\mathbb{R}^k\times\mathbb{R}^d,\mathcal{B}(\mathbb{R}^k)\otimes\mathcal{B}(\mathbb{R}^d),\lambda^k\otimes\lambda^d).$$

Note that we have restricted the Lebesgue measure here to the Borel measurable sets, not the Lebesgue measurable sets. In fact, it is *not* the case that the product σ-algebra $\Sigma_k\otimes\Sigma_d$ is the same as Σ_n. However, the product measure $\pi=\lambda^k\otimes\lambda^d$ obtained from Carathéodory's Theorem, when restricted to all of the π^*-measurable sets is the Lebesgue measure λ^n, and the σ-algebra of π^*-measurable sets is Σ_n.

What this tells us is that even if we start with complete measure spaces, the product measure restricted to the product σ-algebra may not be complete (but its completion will, by definition, be complete).

8 Measurable functions and their properties

8.1 Motivation for measurability

Recall that at the beginning of the course, we tried to think about how we might define the integral of the so-called Dirichlet function, also known as the indicator function of the rationals: $f : [0,1] \to \mathbb{R}$ given by

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

We can think of this function as a pointwise limit of Riemann integrable functions f_n with $\int_0^1 f_n = 0$ for all n, but Riemann integral theory does not allow us to exchange the integral and the limit. However, we have now seen that $\mathbb{Q}$ is Lebesgue measurable and has measure zero. So if the integral of f is somehow to capture the 'area under the graph', the only sensible way to assign a value to the integral of f should be to have $\int_0^1 f = 0 = \lim_{n\to\infty} \int_0^1 f_n$.

More generally, we can consider the indicator function of a set $A \subset X$, where $(X, \mathcal{A}, \mu)$ is a measure space:

$$\mathbb{1}_A = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

If we want to integrate $\mathbb{1}_A$ with respect to the measure μ, then provided the set $A \in \mathcal{A}$ (i.e., is μ-measurable), we can try to assign the value

$$\int_X \mathbb{1}_A \, \mathrm{d}\mu = \mu(A).$$

Once we have assigned the value of integrals to such functions, we can try to treat them (or linear combinations of them) as the equivalents of the step functions in the construction of the Riemann integral.

One advantage of this approach is that it will vastly increase the range of functions that we can integrate. However, as we saw when measuring sets, we cannot hope in general to be able to integrate all functions (just as we could not measure all sets). This leads us naturally to the notion of *measurable functions*, which is where this chapter begins. We will soon see that such functions are exactly those that can be well approximated by linear combinations of characteristic functions of measurable sets.

Notation (range, pre-image): Suppose Y is a set and $f : X \to Y$ is a function. The **image** of f (also called the **range** of f), denoted $f(X)$, is the set of possible values taken by the function f, i.e.

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

The set $f(X)$ could be strictly contained in Y. The whole set Y is called the **codomain** of f. Some authors (not these notes!) use the word 'range' to mean the codomain, rather than the image, of f.

Given $B \subset Y$, we define the **pre-image** (or **inverse image**) **of** B **under** f as

$$f^{-1}(B) := \{x \in X : f(x) \in B\}.$$

Note that *nothing* is implied about f having an inverse function: in general, it will not have one. For example, $f^{-1}(B)$ could be empty, even if B is not. The notation $f^{-1}(B)$ is simply a shorthand for the set of points in X that f maps into the set B. For example if $X = Y = \mathbb{R}$ and $f(x) = x^2$ then $f^{-1}([1,9]) = [1,3] \cup [-3,-1]$ while $f^{-1}((-\infty,0)) = \varnothing$.

Before we continue, we should note the following facts concerning the action of pre-images (independent of measure theory):

(i) f^{-1} preserves complements: $f^{-1}(E^c) = (f^{-1}(E))^c$;

(ii) f^{-1} preserves unions: $f^{-1}\big(\bigcup_{\alpha\in A} E_\alpha\big) = \bigcup_{\alpha\in A}\big(f^{-1}(E_\alpha)\big)$;

(iii) f^{-1} preserves intersections: $f^{-1}\big(\bigcap_{\alpha\in A} E_\alpha\big) = \bigcap_{\alpha\in A}\big(f^{-1}(E_\alpha)\big)$.

Notation: Extended real line. We are mainly interested in functions f from X to $\mathbb{R}$ but sometimes it is useful to allow f to take values $\pm\infty$ as well. We write $\overline{\mathbb{R}}$ or $[-\infty,\infty]$ to denote the set $\mathbb{R}\cup\{-\infty,+\infty\}$, known as the **extended real line**.

8.2 Measurable functions

We begin with the definition of a measurable function.

Definition 8.1. Let $(X,\mathcal{A})$ and $(Y,\mathcal{B})$ be two measurable spaces. A function $f : X \to Y$ is $(\mathcal{A},\mathcal{B})$-measurable *(or simply* measurable *if no confusion can arise) iff $f^{-1}(E) \in \mathcal{A}$ for all $E \in \mathcal{B}$.*

This definition should recall that of continuity between metric (or topological) spaces. Moreover, it should be obvious from the definition that the composition of measurable functions is measurable. As a test for measurability of a given function, it is helpful to be able to reduce to testing a generating set of the σ-algebras. This is content of the next lemma.

Lemma 8.2. *Let $(X,\mathcal{A})$ and $(Y,\mathcal{B})$ be two measurable spaces and suppose that $\mathcal{B}$ is generated by $\mathcal{E}$. Then f is measurable if and only if $f^{-1}(E) \in \mathcal{A}$ for all $E \in \mathcal{E}$.*

Proof. Note that if f is measurable, then as $\mathcal{E} \subset \mathcal{B}$, we immediately have that $f^{-1}(E) \in \mathcal{A}$ for all $E \in \mathcal{E}$. It remains to prove the reverse implication.

For the converse, we note that

$$\mathcal{M} = \{E \subset Y \mid f^{-1}(E) \subset \mathcal{A}\}$$

is a σ-algebra in Y: by the properties (i)–(ii) above, we have that if $E \in \mathcal{M}$, then as $\mathcal{A}$ is a σ-algebra, also $f^{-1}(E^c) = (f^{-1}(E))^c \in \mathcal{A}$, so that $E^c \in \mathcal{M}$, and similarly for countable unions. As $\mathcal{E} \subset \mathcal{M}$ by assumption, we therefore have that $\sigma(\mathcal{E}) \subset \mathcal{M}$, as required. □

Corollary 8.3. *Let X and Y be metric spaces, $\mathcal{A} = \mathcal{B}_X$, $\mathcal{B} = \mathcal{B}_Y$. Then every continuous function $f : X \to Y$ is measurable.*

Proof. As $\mathcal{B}_Y$ is generated by the open sets in Y and preimages of open sets under continuous maps are open, we simply apply the definitions and conclude. □

Although the definition we have just given is the general definition of measurability, in practice, there are certain specific σ-algebras that occur so often that we give measurability with respect to them its own terminology.

Definition 8.4. Let $(X, \mathcal{A})$ be a measurable space and $f : X \to \mathbb{R}$. We say that f is Borel measurable *if it is* $(\mathcal{A}, \mathcal{B}_{\mathbb{R}})$*-measurable.*

Let Σ_n *denote the Lebesgue measurable sets on* $\mathbb{R}^n$ *and* $f : \mathbb{R}^n \to \mathbb{R}$. *We say that* f *is Lebesgue measurable if it is* $(\Sigma_n, \mathcal{B}_{\mathbb{R}})$ *measurable.*

Remark 8.5. It is very important to note that the σ-algebras on the domain and target in Lebesgue measurability are not the same, even when $n = 1$! This means that the composition of Lebesgue measurable functions is not necessarily measurable; if f and g are both Lebesgue measurable from $\mathbb{R}$ to $\mathbb{R}$, then $g^{-1}(B)$ for a Borel set B may be Lebesgue measurable but not Borel, and so we have no knowledge of the preimage $f^{-1}(g^{-1}(B))$. (This of course relies on the existence of sets which are Lebesgue measurable but not Borel, which we will shortly prove).

Notation: Positive and negative parts. For $y \in \mathbb{R}$, we define $y^+ := \max(y, 0)$ (the **positive part** of y) and $y^- := \max(-y, 0)$ (the **negative part** of y). Note that $y = y^+ - y^-$ and $|y| = y^+ + y^-$. Similarly, given a function $f : X \to \mathbb{R}$, we define $f^\pm$ pointwise as $(f(x))^\pm$.

Corollary 8.6. *Let* $a \in \mathbb{R}$. *If* $f : X \to \mathbb{R}$ *is measurable, so are* af, $|f|$, f^+ *and* f^-.

Proof. This follows directly from observing that each of these functions is a composition of the measurable function f with the continuous (and hence measurable) function g given, in each case, by $g(y) = ay$, $g(y) = |y|$, $g(y) = y^+$ and $g(y) = y^-$ respectively. □

Theorem 8.7 (Limits of measurable functions)**.** *If* $f_n : X \to \overline{\mathbb{R}}$, *are measurable functions, defined for* $n \in \mathbb{N}$, *then the functions* $g : X \to \overline{\mathbb{R}}$ *and* $h : X \to \overline{\mathbb{R}}$ *defined by*

$$g := \sup_{n \geq 1} f_n \qquad \textit{and} \qquad h := \limsup_{n \to \infty} f_n$$

are also measurable. Similarly for $\inf_{n \geq 1} f_n$, *and* $\liminf_{n \to \infty} f_n$.

Proof. We begin by setting some notation:

$$\{f > a\} = f^{-1}((a, \infty)), \qquad \{f < b\} = f^{-1}((-\infty, b)).$$

For each of the combinations h in the statement of the proposition, we will show that $\{h > a\} \in \mathcal{A}$. As the intervals (a, ∞) generate $\mathcal{B}_{\mathbb{R}}$, this then suffices to deduce measurability of the function h.

For any $a \in \mathbb{R}$ we have

$$g^{-1}((a, \infty]) = \bigcup_{n=1}^{\infty} \{x \in X : f_n(x) > a\} = \bigcup_{n=1}^{\infty} f_n^{-1}((a, \infty]) \in \mathcal{A},$$

where we have used that each set in the union is in $\mathcal{A}$ since each f_n is measurable. Therefore g is measurable. Also $\widetilde{g} := \inf_{n \geq 1} f_n = -(\sup_{n \geq 1} -f_n)$ is also measurable.

We can write

$$h = \lim_{n \to \infty} \sup_{k \geq n} f_k = \inf_{n \geq 1} \sup_{k \geq n} f_k,$$

and this is measurable by the previous paragraph. Similarly, $\liminf_n f_n = \sup_{n \geq 1} \inf_{k \geq n} f_k$ is also measurable. □

Corollary 8.8. *If $f_n : X \to \mathbb{R}$ are measurable, and $f(x) := \lim_{n\to\infty} f_n(x)$ exists in $\overline{\mathbb{R}}$ for each $x \in X$, then f is also measurable.*

For the proof of this, just note that if $f(x) := \lim_{n\to\infty} f_n(x)$ exists, then $f = \limsup_{n\to\infty} f_n$.

Definition 8.9. A function $f : X \to \mathbb{R}$ is said to be *simple*, if (i) it is measurable and (ii) the range $f(X)$ is a finite set, i.e. f takes only finitely many values.

Note 1: Here we exclude $\pm\infty$ from the possible values.
Note 2: It is convenient to include measurability in the definition of a simple function, though not all authors do so.

Theorem 8.10 (Existence of Approximating Simple Functions)**.** *Let $f : X \to [0, \infty]$ be measurable. There exist nonnegative simple functions f_n, $n \in \mathbb{N}$, such that $f_n \uparrow f$ pointwise, or in other words, such that for all $x \in X$:*
(a) $0 \le f_n(x) \le f_{n+1}(x)$ for all $n \in \mathbb{N}$;
(b) $f_n(x) \to f(x)$ as $n \to \infty$.

That is, every nonnegative measurable function can be expressed as an increasing limit of simple functions.

Proof. For each $n \ge 1$, define the function $f_n : X \to \mathbb{R}$ by:

$$f_n(x) := \begin{cases} (k-1)10^{-n} & \text{if } (k-1)10^{-n} \le f(x) < k10^{-n} \text{ for some integer } 1 \le k \le n10^n; \\ n & \text{if } f(x) \ge n. \end{cases}$$

In other words, $f_n(x)$ is obtained by rounding $f(x)$ down to the nth decimal place if $f(x) < n$, and setting $f_n(x) = n$ if $f(x) \ge n$. For example, if $f(x) = \pi - 2 = 1.14159...$, then $f_1(x) = 1$, $f_2(x) = 1.14$, $f_3(x) = 1.141$, $f_4(x) = 1.1415$ and so on.

For each n the function f_n is simple. Indeed, the possible values taken by f_n all lie in the finite set $\{k10^{-n} : k \in \mathbb{Z} \cap [0, n10^{-n}]\}$; moreover, f_n is measurable because for $1 \le k \le n10^n$,

$$f_n^{-1}(\{(k-1)10^{-n}\}) = f^{-1}([(k-1)10^{-n}, k10^{-n})) \in \mathcal{A}$$

and $f_n^{-1}(\{n\}) = f^{-1}([n, \infty]) \in \mathcal{A}$, and each set of the form $f_n^{-1}(\alpha, \infty)$ is a finite union of such sets.

Moreover for all $x \in X$:

- $0 \le f_1(x) \le f_2(x) \le \cdots \le f(x)$, so (a) holds;
- $f_n(x) \to f(x)$ as $n \to \infty$ for all x (so (b) holds). The convergence follows because if $0 \le f(x) \le n$ then $f(x) - 10^{-n} \le f_n(x) \le f(x)$.

□

Theorem 8.11 (Sums and products of measurable functions)**.** *If $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ are measurable, then so are $f + g$ and fg.*

Proof. First suppose f and g are simple. Let $\alpha \in \mathbb{R}$. Then

$$(f+g)^{-1}((\alpha,\infty)) = \cup_{a\in f(X), b\in g(X): a+b>\alpha} f^{-1}(\{a\}) \cap g^{-1}(\{b\}),$$

which is a finite union of sets in $\mathcal{A}$ and therefore in $\mathcal{A}$. Thus $f+g$ is measurable in this case.

In the general case, by Theorem 8.10 we can take simple $f_n^+ : X \to \mathbb{R}$ and $f_n^- : X \to \mathbb{R}$ with $0 \le f_n^+ \uparrow f^+$ and $0 \le f_n^- \uparrow f^-$ pointwise. Set $f_n := f_n^+ - f_n^-$. Then f_n is simple (for each n) and $f_n \to f$ pointwise. Similarly we can find simple g_n with $g_n \to g$ pointwise.

But then $f_n + g_n \to f + g$ pointwise and since f_n and g_n are simple, $f_n + g_n$ is measurable by the first case considered. Therefore by Corollary 8.8, $f+g$ is also measurable.

The argument for measurability of fg is the same. □

Remark 8.12. It can further be shown that if $f : X \to [-\infty, \infty]$ is measurable, then so are $f^+ := \max\{f, 0\}$ and $f^- := \max\{-f, 0\}$. Also, if $f, g : X \to [0,\infty]$ are measurable, then so are $f+g$ and fg. (HW)

Before moving on to integration, we introduce one more useful proposition.

Proposition 8.13. *Let $(X, \mathcal{A}, \mu)$ be a complete measure space.*
(i) Let $f : X \to \mathbb{R}$ be Borel (or Lebesgue in the case $X = \mathbb{R}^n$) measurable. Suppose that $g : X \to \mathbb{R}$ is such that $g = f$ almost everywhere. Then g is measurable.
(ii) Suppose that for all $n \in \mathbb{N}$, $f_n : X \to \mathbb{R}$ is Borel (or Lebesgue in the case $X = \mathbb{R}^n$) measurable and that $f_n \to f$ almost everywhere. Then f is measurable.

Recall that the terminology of almost everywhere was introduced at the beginning of Section 6.

9 The Lebesgue integral for non-negative functions

Consider a Borel function $f : \mathbb{R} \to \mathbb{R}$. If $f(x) \ge 0$ for all x, we may define $\int_{-\infty}^{\infty} f(x)dx$ as the 'area under the curve', i.e. the two-dimensional Lebesgue measure (λ_2) of the region between the x-axis and the graph $y = f(x)$ (we call this region the *hypograph* of f). Then for general f we shall define $\int_{-\infty}^{\infty} f(x)dx$ by considering f^+ and f^- separately.

This strategy works just as well for $f : X \to \mathbb{R}$ with $(X, \mathcal{A}, \mu)$ an arbitrary measure space; instead of λ_2 we use the product measure $\mu \otimes \lambda_1$. We denote the integral by $\int f d\mu$. The integral $\int_{-\infty}^{\infty} f(x)dx$ amounts to the special case where $X = \mathbb{R}$, $\mathcal{A} = \Sigma_1$ and μ is Lebesgue measure. In the case where μ is a probability measure, the function f is usually called a *random variable* and the integral $\int f d\mu$ is called the *expected value* (or *expectation*) of f (often denoted $\mathbb{E}[f]$).

We are now in a position to begin to define the Lebesgue integral on a measure space. Due to the difficulties with functions that can take infinite values and still be measurable, we begin by focusing on the integral for non-negative functions. To shorten the notation, we introduce the following space of functions.

Definition 9.1. Let $(X, \mathcal{A})$ be a measurable space. We define the set $L^+(X, \mathcal{A})$ (or simply L^+ when clear) by

$$L^+ := \{f : X \to [0,\infty] \,|\, f \text{ is } \textit{measurable}\}.$$

Throughout this section, $(X, \mathcal{A}, \mu)$ denotes an arbitrary σ-finite measure space.

Definition 9.2. Given $f : X \to [0, \infty]$ we define the *hypograph* of f to be a subset of $X \times \mathbb{R}$ given by

$$\text{hyp}(f) := \{(x, y) \in X \times \mathbb{R} : 0 < y < f(x)\}.$$

Lemma 9.3. *If $f : X \to [0, \infty]$ is measurable, then* $\text{hyp}(f) \in \mathcal{A} \otimes \mathcal{B}$.

Proof. If $0 < y < f(x)$ we can find a rational number $q \in (y, f(x))$. Therefore

$$\begin{aligned} \text{hyp}(f) &= \cup_{q \in \mathbb{Q} \cap (0,\infty)} \{(x, y) \in X \times \mathbb{R} : 0 < y < q < f(x)\} \\ &= \cup_{q \in \mathbb{Q} \cap (0,\infty)} [f^{-1}((q, \infty]) \times (0, q)], \end{aligned}$$

which is a countable union of sets in $\mathcal{A} \otimes \mathcal{B}$, and hence itself is in $\mathcal{A} \otimes \mathcal{B}$. □

Remark 9.4. Conversely, it can be proved that if $f : X \to [0, \infty]$ satisfies $\text{hyp}(f) \in \mathcal{A} \otimes \mathcal{B}$, then f is measurable. We leave the proof of this as an exercise.

Recall that we are assuming that the measure space $(X, \mathcal{A}, \mu)$ is σ-finite (and so is the measure space $(\mathbb{R}, \mathcal{B}, \lambda_1)$). Therefore the product measure $\mu \otimes \lambda_1$ is well defined, as the unique measure on $(X \times \mathbb{R}, \mathcal{A} \otimes \mathcal{B})$ that agrees with the 'area' function for measurable rectangles (see Definition 7.4).

We may now define the integral of a nonnegative measurable function on X as the product measure of its hypograph, and then extend to general real-valued functions on X by taking positive and negative parts.

Definition 9.5. If $f \in L^+(X, \mathcal{A})$, then we define the *integral of f with respect to μ* to be

$$\int f \, d\mu = (\mu \otimes \lambda_1)(\text{hyp}(f)). \tag{2}$$

It is important to note at this stage that we have not assumed that the integral of a non-negative function is finite! It could very well be infinite. This causes some problems in terminology: we want to be able to say that the integral of a function is infinite, but we also want to say that a function is integrable if it has finite integral. In other words, we will end up in a situation where some functions are not integrable but can be integrated (and have infinite integral).

Notation: Given a function $f \in L^+$ and a set $A \in \mathcal{A}$, we define

$$\int_A f \, d\mu = \int f \mathbb{1}_A \, d\mu.$$

Lemma 9.6. *Suppose $f, g \in L^+(X, \mathcal{A})$.*
(a) If $f \le g$ pointwise then $\int f \, d\mu \le \int g \, d\mu$.
(b) If $\int f \, d\mu < \infty$, then $\mu(f^{-1}(\{\infty\})) = 0$.

Proof. (a) Suppose $0 \le f \le g$. Then $\text{hyp}(f) \subset \text{hyp}(g)$ so $(\mu \otimes \lambda_1)(\text{hyp}(f)) \le (\mu \otimes \lambda_1)(\text{hyp}(g))$, or in other words $\int f \, d\mu \le \int g \, d\mu$.
(b) Set $E := f^{-1}(\{\infty\})$. Then $\text{hyp}(f) \supset E \times (0, \infty)$, and if $\mu(E) > 0$ then $(\mu \otimes \lambda_1)(\text{hyp}(f)) \ge (\mu \otimes \lambda_1)(E \times [0, \infty)) = \infty$, and the result follows. □

Theorem 9.7 (Monotone Convergence Theorem). *Let* $\{f_n\}_{n=1}^{\infty}$ *be a sequence such that* $f_n \in L^+$ *for all* $n \in \mathbb{N}$ *and such that* $0 \le f_1 \le f_2 \le \dots$. *Define* $f(x) = \lim_{n\to\infty} f_n(x)$. *Then*

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

As the sequence f_n is monotone, the limit function f is well-defined and in L^+.

Proof. Since $f_n \le f_{n+1}$ we have $\mathrm{hyp}(f_n) \subset \mathrm{hyp}(f_{n+1})$ for each $n \in \mathbb{N}$. Also, we claim

$$\mathrm{hyp}(f) = \cup_{n=1}^{\infty} \mathrm{hyp}(f_n).$$

Indeed, if $(x,y) \in \mathrm{hyp}(f)$ then $0 < y < f(x)$, and since $f_n(x) \to f(x)$ we have $f_n(x) > y$ for some n, and hence $(x,y) \in \cup_{n=1}^{\infty} \mathrm{hyp}(f_n)$. Hence $\mathrm{hyp}(f) \subset \cup_{n=1}^{\infty} f_n$ and the reverse inclusion holds because $f_n \le f$ pointwise so $\mathrm{hyp}(f_n) \subset \mathrm{hyp}(f)$ for all n, justifying the claim. Therefore by continuity from below of $\mu \otimes \lambda_1$ (compare Theorem 5.3(iii)),

$$\int f \,\mathrm{d}\mu = (\mu \otimes \lambda_1)(\mathrm{hyp}(f)) = \lim_{n\to\infty} ((\mu \otimes \lambda_1)(\mathrm{hyp}(f_n))) = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu,$$

as required. □

Lemma 9.8 (Integrating simple functions). *(i) If* $f : X \to [0,\infty]$ *is simple then it has a representation*

$$f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} \tag{3}$$

for some $n \in \mathbb{N}$ *and* $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, *and* $A_1, \dots, A_n \in \mathcal{A}$, *pairwise disjoint.*

(ii) If f *is given by* (3) *for some* $n \in \mathbb{N}$ *with* $\alpha_1, \dots, \alpha_n \in [0,\infty]$, *and* $A_1, \dots, A_n \in \mathcal{A}$ *pairwise disjoint, then*

$$\int f \,\mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i). \tag{4}$$

Proof. (i) Suppose f is simple. Let $\alpha_1, \dots, \alpha_n$ be an enumeration of the range of f, i.e. $f(X) = \{\alpha_1, \dots, \alpha_n\}$. Set $A_i = f^{-1}(\{\alpha_i\})$. Then (3) holds and the sets A_i are in $\mathcal{A}$ and pairwise disjoint.

(ii) Note we do not assume here that the α_i are distinct. For $(x,y) \in X \times \mathbb{R}$, to have $(x,y) \in \mathrm{hyp}(f)$ we must have for some $i \le n$ that $x \in A_i$ and $0 < y < f(x) = \alpha_i$. Thus

$$(x,y) \in \mathrm{hyp}(f) \Leftrightarrow \exists i \le n : (x,y) \in A_i \times (0,\alpha_i).$$

Therefore $\mathrm{hyp}(f) = \cup_{i=1}^{n} (A_i \times (0,\alpha_i))$ (interpreting $(0,\alpha_i) = \varnothing$ in any cases with $\alpha_i = 0$) and the sets in the union are pairwise disjoint (since the A_i are pairwise disjoint), so that using (2) we have

$$\int f \,\mathrm{d}\mu = \sum_{i=1}^{n} \mu \otimes \lambda_1(A_i \times (0,\alpha_i)) = \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

□

Remark 9.9. We make several remarks:
(i) In (4), there might be some indices i for which $\alpha_i = 0$ and $\mu(A_i) = \infty$ or $\alpha_i = \infty$ and $\mu(A_i) = 0$. To interpret the right hand side of (4) in these cases, we take the product to be zero.
(ii) In the case where μ is a probability measure and denoted $\mathbb{P}$, (4) gives us the well-known formula for the expectation of a discrete random variable:

$$\mathbb{E}[f] = \sum_{\alpha} \alpha \mathbb{P}[\{f = \alpha\}]$$

where $\{f = \alpha\}$ is short for $\{x \in X : f(x) = \alpha\}$, and the sum is over all possible values α for f, i.e. over $\alpha \in f(X)$.
(iii) By working a bit harder, we could have defined the integral for simple functions directly through (4) and then defined, for functions $f \in L^+$,

$$\int f \, \mathrm{d}\mu = \sup \Big\{ \int \varphi \, \mathrm{d}\mu \,|\, \varphi \text{ is simple and } 0 \leq \varphi \leq f \Big\}.$$

Such a definition would not have required us to assume that $(X, \mathcal{A}, \mu)$ is σ-finite, but would have needed more work to show that the integral is well-defined for simple functions, independent of the choice of representation, and then to prove the Monotone Convergence Theorem.

We would like to show that $\int (af + bg) d\mu = a \int f \mathrm{d}\mu + b \int g \mathrm{d}\mu$, for any $a, b \in \mathbb{R}$ and $f, g \in \mathbb{R}(X)$ such that $\int f \mathrm{d}\mu$ and $\int g \mathrm{d}\mu$ are finite (the function $af + bg$ is defined pointwise, i.e. $(af + bg)(x) = af(x) + bg(x)$ for all $x \in X$). As a first step, we consider the special case where f and g are nonnegative simple functions, and $a, b \geq 0$.

Lemma 9.10. *Suppose $f : X \to [0, \infty)$ and $g : X \to [0, \infty)$ are simple functions, and $c \in [0, \infty)$ is a constant. Then*
(i) $\int (cf) \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$, and
(ii) $\int (f + g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.

Proof. (i) By Lemma 9.8(i), we can write $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ with $A_i \in \mathcal{A}$ disjoint. Then $cf = \sum_{i=1}^{n} c\alpha_i \mathbb{1}_{A_i}$, and by Lemma 9.8(ii) $\int (cf) \, \mathrm{d}\mu = \sum_i (c\alpha_i)\mu(A_i) = c \int f \, \mathrm{d}\mu$.

(ii) Suppose $f(X) = \{\alpha_1, \dots, \alpha_m\}$ and $g(X) = \{\beta_1, \dots, \beta_n\}$. Set $A_i = f^{-1}(\{\alpha_i\})$ and $B_j = g^{-1}(\{\beta_j\})$. Then $f = \sum_{i=1}^{m} \alpha_i \mathbb{1}_{A_i}$ and $g = \sum_{j=1}^{n} \beta_j \mathbb{1}_{B_j}$ with $\alpha_1, \dots, \alpha_m$ distinct and $\beta_1, \dots, \beta_n$ distinct. Also $A_1, \dots, A_m$ form a partition of X (that is, they are pairwise disjoint and their union is X) and so do $B_1, \dots, B_n$. For each $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$, set $E_{ij} = A_i \cap B_j$. Then $A_i = \cup_{j=1}^{n} E_{ij}$ (disjoint union) so

$$\int f \, \mathrm{d}\mu = \sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \mu(E_{ij})$$

and similarly $\int g \, \mathrm{d}\mu = \sum_{j=1}^{n} \sum_{i=1}^{m} \beta_j \mu(E_{ij})$. Hence

$$\int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu = \sum_{i=1}^{m} \sum_{j=1}^{n} (\alpha_i + \beta_j)\mu(E_{ij}). \tag{5}$$

Now $(f+g)(x) = \alpha_i + \beta_j$ for all $x \in E_{ij}$; in other words

$$f + g = \sum_{i=1}^{m} \sum_{j=1}^{n} (\alpha_i + \beta_j) \mathbb{1}_{E_{ij}},$$

and the sets $E_{11}, \dots, E_{mn}$ are pairwise disjoint, so by Lemma 9.8(ii),

$$\int (f+g) \, \mathrm{d}\mu = \sum_{i=1}^{m} \sum_{j=1}^{n} (\alpha_i + \beta_j) \mu(E_{ij}).$$

Comparing this with (5) gives us the result. □

Lemma 9.11. *Suppose $f, g \in L^+$. Let $c \in [0, \infty)$. Then: (i) $\int (cf) \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$, and (ii) $\int (f+g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.*

Proof. By Theorem 8.10, there exist simple functions $0 \le f_n \uparrow f$ and $0 \le g_n \uparrow g$. To prove (i), note that $cf_n \uparrow cf$, and cf_n is simple, so that by the MCT and Lemma 9.10,

$$\int cf d\mu = \lim_{n\to\infty} \int cf_n d\mu \overset{\text{Lemma 9.10}}{=} \lim_{n\to\infty} c \int f_n d\mu = c \lim_{n\to\infty} \int f_n d\mu = c \int f d\mu.$$

For (ii), note that $0 \le (f_n + g_n) \uparrow (f + g)$. Applying MCT and the additivity of integration for nonnegative simple functions (Lemma 9.10), we get:

$$\begin{aligned} \int (f+g) \, \mathrm{d}\mu &\overset{\text{MCT}}{=} \lim_{n\to\infty} \int (f_n + g_n) \, \mathrm{d}\mu \overset{\text{Lemma 9.10}}{=} \lim_{n\to\infty} \left[\int f_n \, \mathrm{d}\mu + \int g_n \, \mathrm{d}\mu \right] \\ &= \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu + \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu \overset{\text{MON}}{=} \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu. \end{aligned}$$

□

Using the Monotone Convergence Theorem, we can easily exchange integral with sums for non-negative functions. In fact, we can do this even for infinite sums of functions.

Theorem 9.12. *Let $f_n \in L^+$ for $n \in \mathbb{N}$ and define $f = \sum_{n=1}^{\infty} f_n$. Then $\int f \, \mathrm{d}\mu = \sum_{n=1}^{\infty} \int f_n \, \mathrm{d}\mu$.*

Proof. We define the sequence of partial sums pointwise as $g_k(x) := \sum_{n=1}^{k} f_n(x)$, and set $g(x) := \sum_{n=1}^{\infty} f_n(x)$. Then $g_k \uparrow g_\infty$ pointwise and g_∞ is measurable by Corollary 8.8. Hence by the MCT, and Lemma 9.11,

$$\int g_\infty \mathrm{d}\mu = \lim_{k\to\infty} \int g_k \mathrm{d}\mu = \lim_{k\to\infty} \sum_{n=1}^{k} \int f_n \mathrm{d}\mu = \sum_{n=1}^{\infty} \int f_n d\mu,$$

as required. □

We would like to recover some of the classical results of integration theory from the Riemann integral, such as the very useful lemma that states that if $f \ge 0$, f is continuous, and $\int f = 0$, then $f = 0$ everywhere. As we have already seen that measurable functions do not have to be continuous (or even close to continuous), we would like a stronger version of this lemma that drops the assumption of continuity. In general, of course, such a result would then be false (consider a function that takes the value 1 at a single point and is zero elsewhere), but with the language of measure theory, we can make a statement that is almost as good.

Recall the terminology *almost everywhere* (abbreviated a.e.) from Section 6: a function $f = 0$ a.e. if $\mu(\{x \mid f(x) \neq 0\}) = 0$.

Lemma 9.13. *Let $f \in L^+$. Then $\int f \, \mathrm{d}\mu = 0$ if and only if $f = 0$ a.e.*

The proof of this lemma is left as a homework exercise (HW).

There are numerous consequences of this lemma, which basically says that to understand the integral of a function (or sequence of functions), we can neglect sets of measure zero. One example is the following corollary.

Corollary 9.14. *Let $f_n \in L^+$ for all $n \in \mathbb{N}$ be a sequence that is increasing a.e. and define the limit function $f = \lim_{n\to\infty} f_n$ on the set where f_n is increasing (and defined arbitrarily on the remaining set) such that $f \in L^+$. Then*

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

Proof. Define the set

$$E = \{x \in X \mid f_n(x) \text{ is not increasing}\}, \qquad \mu(E) = 0.$$

Then as $f = f\mathbb{1}_{E^c}$ a.e., we have $f - f\mathbb{1}_{E^c} = 0$ a.e. and $f - f\mathbb{1}_{E^c} \in L^+$. Thus $\int (f - f\mathbb{1}_{E^c}) \mathrm{d}\mu = 0$, and so, by adding $\int f\mathbb{1}_{E^c} \, \mathrm{d}\mu$ to both sides,

$$\begin{aligned} \int f \, \mathrm{d}\mu = \int f\mathbb{1}_{E^c} \, \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \mathbb{1}_{E^c} \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mathbb{1}_{E^c} \, \mathrm{d}\mu & \qquad \text{by MCT} \\ = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu, & \end{aligned}$$

where we used that $f_n - f_n\mathbb{1}_{E^c} = 0$ a.e. and $f_n - f_n\mathbb{1}_{E^c} \in L^+$ for all n in the last step. □

The assumption in this corollary that the sequence f_n be increasing is essential; we can easily construct a sequence of functions that converges a.e. but where the integrals fail to converge to the integral of the limit function. A simple example is on $\mathbb{R}$ equipped with the Lebesgue measure. Let $f_n = \mathbb{1}_{(n,n+1)}$. Then $f_n \to 0$ a.e. but $\int f_n = 1$ for all n.

While this simple example shows that we cannot take limits without at least some kind of additional assumption, the following, highly significant lemma shows that the integral of the limit can never exceed the limit of the integrals.

Lemma 9.15 (Fatou's Lemma)**.** *Let $\{f_n\}_{n=1}^\infty$ be a sequence of functions in L^+. Then*

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

Note that even if the function $\lim_{n\to\infty} f_n$ exists, the sequence of integrals $\int f_n \, \mathrm{d}\mu$ may not have a limit (**Exercise:** construct such a sequence).

Proof. First note that, by definition,

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

as the sequence of infima is increasing: $\inf_{n\ge k} f_n \le \inf_{n\ge k+1} f_n$. In addition, $\inf_{n\ge k} f_n \le f_j$ for all $j \ge k$.

Therefore,

$$\int \inf_{n\ge k} f_n \, \mathrm{d}\mu \le \int f_j \, \mathrm{d}\mu \text{ for all } j \ge k.$$

Taking the infimum in j, we therefore obtain

$$\int \inf_{n\geq k} f_n \, d\mu \leq \inf_{j\geq k} \int f_j \, d\mu.$$

Both the left and right hand sides of this inequality are increasing sequences of real numbers, and so we may take limits (in the extended real line) to find

$$\lim_{k\to\infty} \int \inf_{n\geq k} f_n \, d\mu \leq \lim_{k\to\infty} \inf_{j\geq k} \int f_j \, d\mu.$$

The right hand side is clearly $\liminf_{n\to\infty} \int f_n \, d\mu$, as required. For the left hand side, we use the monotone convergence theorem to see

$$\lim_{k\to\infty} \int \inf_{n\geq k} f_n \, d\mu = \int \lim_{k\to\infty} \inf_{n\geq k} f_n \, d\mu = \int \liminf_{n\to\infty} f_n \, d\mu.$$

□

Example 9.16. Another example of a situation in which we have a strict inequality in Fatou's lemma is as follows. Suppose $X = \{0, 1\}$ with the counting measure. Suppose $f_n(1) = (1 + (-1)^n)/2$ and $f_n(0) = 1 - f_n(1)$. That is, $f_n(1)$ equals 1 for even n and zero for odd n, and $f_n(0)$ is the other way round.

Then $\int f_n \, d\mu = 1 + 0 = 1$ for all n, while $\liminf_{n\to\infty} f_n(x) = 0$ for all $x \in X$ so $\int \liminf_{n\to\infty} f_n = 0$. Thus, in this case the inequality in the statement of Fatou's lemma is strict.

Corollary 9.17. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions in L^+ such that $f_n \to f$ a.e. Then*

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

10 Integration of general measurable functions

Having established a method for integrating non-negative integrable functions, we now extend this theory to real or complex valued functions without a definite sign. We begin with real-valued functions.

Definition 10.1. Let $f : X \to \mathbb{R}$ be measurable. We define the positive and negative parts $f^+ := \max\{f, 0\}$, $f^- := \max\{-f, 0\}$, so that $f = f^+ - f^-$. We define the *integral of f* to be

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

when at most one of $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ is $+\infty$.

This definition holds for many measurable functions. We now define the particular class of functions we call integrable.

Definition 10.2. Let $f : X \to \mathbb{R}$ be measurable. We say that f is *integrable* if

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

NB: Note that the integral is still defined for many functions that are not integrable!

Observe also that, by definition, f is integrable if and only if $\left|\int f\,\mathrm{d}\mu\right| < \infty$. This holds if and only if both $\int f^+\,\mathrm{d}\mu < \infty$ and $\int f^-\,\mathrm{d}\mu < \infty$. Thus f is integrable if and only if $|f|$ has a finite integral: $\int |f|\,\mathrm{d}\mu = \int (f^+ + f^-)\,\mathrm{d}\mu < \infty$.

Recall from the theory of the Riemann integral that the space of Riemann integrable functions is a vector space. The same is true for Lebesgue integrable functions.

Definition 10.3. Define the space $\mathcal{L}^1(\mu)$ (or simply $\mathcal{L}^1$ when no confusion can arise) as

$$\mathcal{L}^1(\mu) := \left\{ f : X \to \mathbb{R} \,\middle|\, \left| \int f\,\mathrm{d}\mu \right| < \infty \right\}.$$

Proposition 10.4. *The space $\mathcal{L}^1(\mu)$ is a vector space and the integral is a linear functional on this space. Moreover, $|\int f\,\mathrm{d}\mu| \leq \int |f|\,\mathrm{d}\mu$.*

Proof. Let $f \in \mathcal{L}^1(\mu)$. Let $a > 0$. Then $(af)^+ = af^+$, $(af)^- = af^-$. Thus

$$\int af\,\mathrm{d}\mu = \int (af)^+\,\mathrm{d}\mu - \int (af)^-\,\mathrm{d}\mu = \int af^+\,\mathrm{d}\mu - \int af^-\,\mathrm{d}\mu = a\Big(\int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu\Big) = a\int f\,\mathrm{d}\mu.$$

Now suppose $a < 0$ (the case $a = 0$ is straightforward). Then $(af)^+ = |a|f^-$ and $(af)^- = |a|f^+$, leading to

$$\int af\,\mathrm{d}\mu = -|a|\int f\,\mathrm{d}\mu = a\int f\,\mathrm{d}\mu.$$

Next, take $f, g \in \mathcal{L}^1(\mu)$. Then $|f+g|, |f|, |g| \in L^+$ and $|f+g| \leq |f| + |g|$, so that

$$\int |f+g|\,\mathrm{d}\mu \leq \int |f|\,\mathrm{d}\mu + \int |g|\,\mathrm{d}\mu.$$

Hence $f + g \in \mathcal{L}^1(\mu)$ and $\mathcal{L}^1(\mu)$ is a vector space.

To show that $\int$ is a linear functional, we need to show that $\int af\,\mathrm{d}\mu = a\int f\,\mathrm{d}\mu$ (which we have already shown) and $\int (f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$.

Let $h = f + g$, $h = h^+ - h^-$, $f = f^+ - f^-$, $g = g^+ - g^-$. Note that we have no guarantee that $h^+ = f^+ + g^+$. However, we may observe that

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

so that

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

where both sides are sums of functions in L^+.

Then

$$\begin{aligned}\int h^+\,\mathrm{d}\mu + \int f^-\,\mathrm{d}\mu + \int g^-\,\mathrm{d}\mu &= \int (h^+ + f^- + g^-)\,\mathrm{d}\mu \\ &= \int (h^- + f^+ + g^+)\,\mathrm{d}\mu = \int h^-\,\mathrm{d}\mu + \int f^+\,\mathrm{d}\mu + \int g^+\,\mathrm{d}\mu.\end{aligned}$$

Rearranging, we obtain

$$\int h\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu,$$

as required.

The final claim that $|\int f\,\mathrm{d}\mu| \leq \int |f|\,\mathrm{d}\mu$ is left as an **exercise** (HW). □

Now that we have defined the class of integrable functions, we are going to show some very basic properties that have to hold for such functions to show that they are not too bad.

Proposition 10.5. *(i) Let $f \in \mathcal{L}^1(\mu)$. Then $\{x \in X \mid f(x) \neq 0\}$ is σ-finite and $\{x \in X \mid f(x) = \pm\infty\}$ has measure zero.*
(ii) Let $f, g \in \mathcal{L}^1(\mu)$. Then $\int_E f \,\mathrm{d}\mu = \int_E g \,\mathrm{d}\mu$ for all $E \in \mathcal{A}$ if and only if $\int |f - g| \,\mathrm{d}\mu = 0$ if and only if $f = g$ a.e.

Proof. (i) Suppose without loss of generality that $f \geq 0$ as $\{f \neq 0\} = \{|f| \neq 0\}$. Note that

$$\{x \mid f(x) > 0\} = \bigcup_{n=1}^{\infty} \{x \mid f(x) > \frac{1}{n}\}.$$

Now observe that

$$\infty > \int f \,\mathrm{d}\mu \geq \int_{\{f > \frac{1}{n}\}} f \,\mathrm{d}\mu \geq \int_{\{f > \frac{1}{n}\}} \frac{1}{n} \,\mathrm{d}\mu = \frac{1}{n} \mu\big(\{f > \frac{1}{n}\}\big),$$

so that we see $\{f > \frac{1}{n}\}$ has finite measure and hence $\{f \neq 0\}$ is σ-finite.

Moreover, if the set $\{x \in X \mid f(x) = \infty\}$ has positive measure, then we have

$$\infty > \int f \,\mathrm{d}\mu \geq \int_{\{f = \infty\}} f \,\mathrm{d}\mu = \infty,$$

a contradiction. Similarly, $\{x \in X \mid f(x) = -\infty\}$ has measure zero also, concluding the proof of (i).

The proof of (ii) is left as a homework **exercise** (HW). □

As we have shown that $\mathcal{L}^1(\mu)$ is a vector space with some nice properties, one obvious next step might be to try to create a norm on this space that allows us to bring functional analysis techniques to bear on the integrable functions. This is indeed what we are going to do, but we first need to address a certain technical difficulty.

The clearest candidate for a norm on $\mathcal{L}^1(\mu)$ is the functional

$$\|\cdot\|_{\mathcal{L}^1(\mu)} : \mathcal{L}^1(\mu) \to \mathbb{R}_{\geq 0} \qquad \text{defined by } \|f\|_{\mathcal{L}^1(\mu)} = \int |f| \,\mathrm{d}\mu.$$

Clearly we have, for $\lambda \in \mathbb{R}$, $\|\lambda f\|_{\mathcal{L}^1(\mu)} = |\lambda| \|f\|_{\mathcal{L}^1(\mu)}$ and also $\|f + g\|_{\mathcal{L}^1(\mu)} \leq \|f\|_{\mathcal{L}^1(\mu)} + \|g\|_{\mathcal{L}^1(\mu)}$. However, the final property of a norm is that $\|f\|_{\mathcal{L}^1(\mu)} = 0$ implies $f = 0$. This clearly fails for any function f which is non-zero on a non-empty set of measure zero! For example, with $X = \mathbb{R}$ and the Lebesgue measure, we have $\int \mathbb{1}_{\mathbb{Q}} \,\mathrm{d}\mu = 0$.

To handle this issue, we are instead going to work with equivalence classes of functions (this is less strange that it may at first seem for reasons that will become apparent soon).

Definition 10.6 ($L^1(\mu)$)**.** We define the equivalence relation: for $f, g \in \mathcal{L}^1(\mu)$,

$$f \sim g \text{ if and only if } f = g \text{ a.e.}$$

The quotient space $\mathcal{L}^1(\mu)/\sim$ is the collection of equivalence classes and is the space $L^1(\mu)$.

We define a norm on $L^1(\mu)$ by

$$\|[f]\|_{L^1(\mu)} := \int_X |f| \,\mathrm{d}\mu.$$

NB: By Proposition 10.5, this is well-defined. Indeed, if two functions are in the same equivalence class, then they agree a.e., and so they have the same integral (and the same is true for their absolute values).

We have claimed in the definition that $\|\cdot\|_{L^1(\mu)}$ is in fact a norm. We now prove that this is so.

Lemma 10.7. *The map $\|\cdot\|_{L^1(\mu)} : L^1(\mu) \to \mathbb{R}$ is a norm.*

Proof. Clearly $\|[f]\|_{L^1(\mu)} \geq 0$ for all $[f] \in L^1$. If $\|[f]\|_{L^1(\mu)} = 0$, then we may take a representative f of the equivalence class and see that $\|[f]\|_{L^1(\mu)} = \int |f| \,\mathrm{d}\mu$, so that $f = 0$ a.e. by Proposition 10.5(ii). Thus $[f] = [0]$ (the equivalence class generated by the zero function) which is the zero element in the quotient space $L^1(\mu)$.

The remaining properties of a norm follow from the same arguments as above once we note that $[\lambda f] = \lambda[f]$ and $[f + g] = [f] + [g]$. □

11 Dominated convergence and sets of measure zero

We have already seen one way of exchanging limits with integrals: via the Monotone Convergence Theorem. However, in general, we cannot expect that sequences of functions will be monotone (or even monotone a.e.) and so we want a theorem that gives sufficient conditions to exchange limits and integrals without needing monotonicity. It turns out that the space of integrable functions is a good one in which to work for this.

We saw earlier that sequences such as $\mathbb{1}_{(n,n+1)}$ do not allow us to pass the limit through the integral. What is happening here, in a sense, is that the mass of the function is escaping off to infinity and being lost in the limit. An alternative way of 'losing' integral in the limit (equivalently, having a strict inequality in Fatou's Lemma) is to think of a sequence of functions where the mass is concentrating, such as $n\mathbb{1}_{[0,\frac{1}{n}]}$. Again the integral is always 1, but the mass accumulates at the origin and is lost in the a.e. limit.

The key assumption that will prevent this in the next theorem in the assumption that *all* of the functions in the sequence can be bounded by a single integrable function. When this happens, we say that this upper bound dominates the sequence, hence the name of the theorem: the Dominated Convergence Theorem. This dominating function prevents both the accumulation of mass and the possibility of mass escaping to infinity.

Theorem 11.1 (Dominated Convergence Theorem)**.** *Let $\{f_n\}_{n=1}^{\infty} \subset \mathcal{L}^1(\mu)$, $f \in \mathcal{L}^1(\mu)$ be such that $f_n \to f$ a.e. and suppose that there exists a function $g \in \mathcal{L}^1(\mu)$ such that $g \geq 0$ and $|f_n(x)| \leq g(x)$ for all $x \in X$. Then*

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

Proof. First, observe that as $|f_n(x)| \leq g(x)$ for all $n \in \mathbb{N}$ and $f_n \to f$ a.e. we have that $|f(x)| \leq g(x)$ a.e. Thus $f \in \mathcal{L}^1(\mu)$ ($\int |f| \,\mathrm{d}\mu < \infty$).

Next, using again the bound $|f_n| \leq g$, observe that the functions $g - f_n, g + f_n \geq 0$ for all n. We may therefore apply Fatou's lemma to each of these two sequences and find

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

Rearranging these inequalities, we have found

$$\limsup_{n\to\infty}\int f_n \, \mathrm{d}\mu \le \int f \, \mathrm{d}\mu \le \liminf_{n\to\infty}\int f_n \, \mathrm{d}\mu,$$

and so we conclude

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

□

Theorem 11.2. *Let $\{f_n\}_{n=1}^{\infty} \subset \mathcal{L}^1(\mu)$ be such that $\sum_{n=1}^{\infty}\int |f_n| \, \mathrm{d}\mu < \infty$. Then $\sum_{n=1}^{\infty} f_n$ converges a.e. to a function in $\mathcal{L}^1(\mu)$ and*

$$\int \sum_{n=1}^{\infty} f_n \, \mathrm{d}\mu = \sum_{n=1}^{\infty}\int f_n \, \mathrm{d}\mu.$$

Proof. We begin by considering the function (valued in the extended real line) $\sum_{j=1}^{\infty}|f_n|$. By the Monotone Convergence Theorem, we have

$$\int \sum_{n=1}^{\infty} |f_n| \, \mathrm{d}\mu = \sum_{n=1}^{\infty}\int |f_n| \, \mathrm{d}\mu,$$

where, by the assumption in the theorem, the right hand side is finite, and so the left hand side is finite also. Thus we have shown $\sum_{n=1}^{\infty}|f_n| \in \mathcal{L}^1(\mu)$ and therefore the sum is finite a.e.

Now we notice that, for all $n \in \mathbb{N}$

$$\Big|\sum_{k=1}^{n} f_k(x)\Big| \le \sum_{n=1}^{\infty}|f_n| =: g(x)$$

so that g is a dominating function in the sense of the Dominated Convergence Theorem.

Let $g_n(x) = \sum_{k=1}^{n} f_k(x)$, so that $|g_n| \le g$, $g \in \mathcal{L}^1(\mu)$ satisfies the assumptions of the DCT. By that theorem, we therefore take

$$\int \lim_{n\to\infty} g_n \, \mathrm{d}\mu = \lim_{n\to\infty}\int g_n = \lim_{n\to\infty}\int \sum_{k=1}^{n} f_k \, \mathrm{d}\mu = \lim_{n\to\infty}\sum_{k=1}^{n}\int f_k \, \mathrm{d}\mu = \sum_{n=1}^{\infty}\int f_n \, \mathrm{d}\mu.$$

□

When we were defining measurable functions, we saw that there was a sense in which they could be well approximated by simple functions. The question now arises whether the same can be said in the space $L^1(\mu)$ with the topology induced by the norm. The answer is that it can, though we will not prove it in this course.

Theorem 11.3. *(i) Let $f \in \mathcal{L}^1(\mu)$, $\varepsilon > 0$. Then there exists a simple function φ such that $\int |f - \varphi| \, \mathrm{d}\mu < \varepsilon$. In particular, the set of equivalence classes of simple functions is dense in $L^1(\mu)$.*
(ii) If μ is the Lebesgue measure on $\mathbb{R}$, then, writing $\varphi = \sum_{j=1}^{n} a_j \mathbb{1}_{E_j}$, we may take each of the sets E_j to be a finite union of open intervals. Additionally, there exists a continuous function g such that $\int |f - g| \, \mathrm{d}\mu < \varepsilon$.

We have compared the Lebesgue integral to the Riemann integral repeatedly so far, but have not yet made the comparison precise. The next theorem not only shows that the two integrals coincide on the set of Riemann integrable functions, it also gives us a characterisation of the Riemann integrable functions (something that was not possible with the classical theory of analysis developed in the 19th Century).

Theorem 11.4. *Let $f : [a,b] \to \mathbb{R}$.*
(i) If f is Riemann integrable, then f is also Lebesgue integrable and the two integrals are the same.
(ii) f is Riemann integrable if and only if the set $\{x \in [a,b] \mid f \text{ is not continuous at } x\}$ has Lebesgue measure zero.

The proof of this theorem is non-examinable (but ask if you would like to see a proof; the proof of (i) is quite direct from the DCT).

Theorem 11.5. *Let $f : X \to [0,\infty]$ be measurable, and for $E \in \mathcal{A}$, put $\varphi(E) := \int_E f \, \mathrm{d}\mu$. Then φ is a measure on $(X, \mathcal{A})$.*

Proof. It is clear that $\varphi : \mathcal{A} \to [0,\infty]$, and that $\varphi(\varnothing) = 0$. Suppose that $E_1, E_2, \ldots \in \mathcal{A}$ are disjoint and set $E := \cup_{n=1}^{\infty} E_n$. Then $\mathbb{1}_E f = \sum_{n=1}^{\infty} \mathbb{1}_{E_n} f$ so by Theorem 9.12,

$$\varphi(E) = \int \mathbb{1}_E f \, \mathrm{d}\,\mu = \sum_{n=1}^{\infty} \int \mathbb{1}_{E_n} f \, \mathrm{d}\,\mu = \sum_{n=1}^{\infty} \varphi(E_n).$$

This proves that φ is a measure. □

Remark. It is customary to write $f \, \mathrm{d}\mu$ for $\mathrm{d}\varphi$, and call f the **density function of φ with respect to μ.**

There is an important converse to the above theorem that we now state.

Definition 11.6. If ν and μ are both measures on $(X, \mathcal{A})$, we say that ν **is absolutely continuous with respect to** μ, denoted $\nu \ll \mu$, if whenever $\mu(E) = 0$ then also $\nu(E) = 0$.

Observe that φ has this property: if $\mu(E) = 0$, then $\varphi(E) = \int_E f \, \mathrm{d}\mu = 0$, by Proposition 10.5.

The following theorem states that under a σ-finiteness assumption, any finite measure that is absolutely continuous with respect to μ has a density.

Theorem 11.7 (Radon-Nikodým Theorem)**.** *Suppose ν and μ are both measures on a measurable space $(X, \mathcal{A})$, and $\nu \ll \mu$. If μ is σ-finite, and $\nu(X) < \infty$, then there exists a measurable function $f : X \to [0,\infty)$ such that $\nu(E) = \int_E f \, \mathrm{d}\,\mu$ for all $E \in \mathcal{A}$.*

> *(Non-examinable)* For the proof see any of the books mentioned at the start of the course.

Example 11.8. Suppose $\mathbb{P}$ is a probability measure on $(\mathbb{R}, \mathcal{B})$ such that $\mathbb{P}(B) = 0$ for all $B \in \mathcal{B}$ with $\lambda_1(B) = 0$. Then by the Radon-Nikodým Theorem, the measure $\mathbb{P}$ has a *probability density function*, that is, a Borel function $f : \mathbb{R} \to [0,\infty)$ such that $\mathbb{P}(A) = \int_A f(x) \, \mathrm{d}x$ for all $A \in \mathcal{B}$. In particular taking $A = \mathbb{R}$ shows that $\int_{-\infty}^{\infty} f(x) \, \mathrm{d}x = 1$.

12 Fubini-Tonelli Theorem

Fubini's theorem is an important result which gives conditions under which we are guaranteed to be able to interchange the order of a double integral.

Before we state and prove the Fubini–Tonelli Theorem for exchanging the order of integration, we gather some preliminary notation and results for the slicings of sets and functions in the product space.

Notation:

If $E \subset X \times Y$, $x \in X$, $y \in Y$, we let $E_x := \{y \in Y \mid (x, y) \in E\}$ and $E^y = \{x \in X \mid (x, y) \in E\}$.

If $f : X \times Y \to \mathbb{R}$, we let $f_x(y) = f(x, y)$ for each fixed $x \in X$, $f^y(x) = f(x, y)$ for each fixed $y \in Y$.

Proposition 12.1. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces.*
(i) If $E \in \mathcal{A} \otimes \mathcal{B}$, then $E_x \in \mathcal{B}$, $E^y \in \mathcal{A}$ for all $x \in X$, $y \in Y$.
(ii) Let $f : X \times Y \to \mathbb{R}$ be $\mathcal{A} \otimes \mathcal{B}$-measurable. Then $f_x(y)$ is $\mathcal{B}$-measurable and $f^y(x)$ is $\mathcal{A}$-measurable.

Proof. (i) Define a collection of sets

$$\mathcal{R} = \big\{E \subset X \times Y \mid E_x \in \mathcal{B},\ E^y \in \mathcal{A}\big\}.$$

We aim to show that $\mathcal{R} \supset \mathcal{A} \otimes \mathcal{B}$.

First, note that if $A \in \mathcal{A}$, $B \in \mathcal{B}$, then

$$(A \times B)_x = \begin{cases} B & \text{if } x \in A, \\ \varnothing & \text{if } x \notin A, \end{cases} \quad \text{and} \quad (A \times B)^y = \begin{cases} A & \text{if } y \in B, \\ \varnothing & \text{if } y \notin B. \end{cases}$$

Thus $A \times B \in \mathcal{R}$.

It now suffices to show that $\mathcal{R}$ is closed under complements and countable unions, as it must then contain the σ-algebra generated by rectangles $A \times B$ (which is $\mathcal{A} \otimes \mathcal{B}$).

Let $\{E_j\}_{j=1}^{\infty} \subset \mathcal{R}$. Then

$$\Big(\bigcup_{j=1}^{\infty} E_j\Big)_x = \{y \in Y \mid (x, y) \in \bigcup_{j=1}^{\infty} E_j\} = \bigcup_{j=1}^{\infty} \{y \in Y \mid (x, y) \in E_j\} = \bigcup_{j=1}^{\infty} (E_j)_x \in \mathcal{B}.$$

Similarly, $\Big(\bigcup_{j=1}^{\infty} E_j\Big)^y \in \mathcal{A}$.

Now suppose $E \in \mathcal{R}$. Then

$$(E^c)_x = \{y \in Y \mid (x, y) \notin E\} = (E_x)^c \in \mathcal{B}$$

and similarly $(E^y)^c = (E^c)^y \in \mathcal{A}$.

(ii) To show that $f_x(y)$ is $\mathcal{B}$-measurable, we need to show that $(f_x)^{-1}(B) \in \mathcal{B}$ for all $B \in \mathcal{B}(\mathbb{R})$ (the Borel σ-algebra). But it is clear from the definition that

$$(f_x)^{-1}(B) = (f^{-1}(B))_x.$$

By part (i), as $f^{-1}(B) \in \mathcal{A} \otimes \mathcal{B}$ (by measurability of f), this must be in $\mathcal{B}$, as required. The proof for f^y is similar. □

Remark 12.2. This proposition lets us make more precise the claim above that even when μ and ν are complete, $\mu \otimes \nu$ may not be. Suppose that $\mathcal{B} \neq \mathcal{P}(Y)$ and that there exists a non-empty set $A \in \mathcal{A}$ with $\mu(A) = 0$. As $\mathcal{A}$ is complete, we may assume that $A = \{x\}$ for a single $x \in X$. Then for $E \in \mathcal{P}(Y) \setminus \mathcal{B}$, we must have that $A \times E \notin \mathcal{A} \otimes \mathcal{B}$ (as if $A \times E \in \mathcal{A} \otimes \mathcal{B}$ then Proposition 12.1(i) implies $(A \times E)_x = (\{x\} \times E)_x = E \in \mathcal{B}$, a contradiction). But $A \times E \subset A \times Y$, and $\mu \otimes \nu(A \times Y) = \mu(A) \times \nu(Y) = 0$. So $\mu \otimes \nu$ is not complete.

Theorem 12.3. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces and suppose $E \in \mathcal{A} \otimes \mathcal{B}$. Then the maps*

$$x \mapsto \nu(E_x) \qquad \textit{and} \qquad y \mapsto \mu(E^y)$$

are measurable on X, Y, respectively, and

$$\int_X \nu(E_x)\,\mathrm{d}\mu = \int_Y \mu(E^y)\,\mathrm{d}\nu = \pi(E) = \mu \otimes \nu(E).$$

Proof. Non-examinable. First suppose that μ and ν are finite and let $\mathcal{C}$ be the set of all $E \in \mathcal{A} \otimes \mathcal{B}$ such that the Theorem holds. Now if $E = A \times B$, by definition, we have $\nu(E_x) = \mathbb{1}_A(x)\nu(B)$ and $\mu(E^y) = \mu(A)\mathbb{1}_B(y)$, so clearly $E \in \mathcal{C}$. By additivity of the measures, finite disjoint unions of rectangles $A \times B$ are in $\mathcal{B}$.

Claim: $\mathcal{C}$ is continuous from above and below (i.e., is closed under countable increasing unions and countable decreasing intersections).
Assuming the claim, we conclude from Homework 5, Q4 that $\mathcal{C}$ is a σ-algebra and hence contains $\mathcal{A} \otimes \mathcal{B}$.
Proof of Claim: Let $E_1 \subset E_2 \subset \cdots$, all $E_j \in \mathcal{C}$ and let $E = \bigcup_{j=1}^\infty E_j$. Then the functions $f_n(y) = \mu((E_n)^y)$ are all measurable and increase pointwise to $f(y) = \mu(E^y)$ (as E_n is an increasing sequence of sets). Hence f is measurable by Proposition 8.13 and, by the Monotone Convergence Theorem,

$$\int_Y \mu(E^y)\,\mathrm{d}\nu(y) = \lim_{n\to\infty} \int \mu\big((E_n)^y\big)\,\mathrm{d}\nu(y) = \lim_{n\to\infty} \mu \otimes \nu(E_n) = \mu \otimes \nu(E).$$

A similar argument gives $\mu \otimes \nu(E) = \int \nu(E_x)\,\mathrm{d}\mu(x)$, so we conclude $E \in \mathcal{C}$.

Now if $\{E_n\}$ is a decreasing sequence in $\mathcal{C}$ and $E = \bigcap_{n=1}^\infty E_n$, then the function $y \mapsto \mu((E_1)^y) \in L^1(Y, \nu)$ because $\mu((E_1)^y) \le \mu(X) < \infty$ and $\nu(Y) < \infty$. Then we may apply the Dominated Convergence Theorem to show, in a similar way to the above, that $E \in \mathcal{C}$.

Finally, to remove the assumption that μ and ν are finite, we write $X \times Y$ as a union of an increasing sequence of finite measure rectangles $X_j \times Y_j$, and the result just proved for finite measure spaces applies to each $E \cap (X_j \times Y_j)$, yielding

$$\mu \otimes \nu\big(E \cap (X_j \times Y_j)\big) = \int_X \mathbb{1}_{X_j}(x)\nu(E_x \cap Y_j)\,\mathrm{d}\mu(x) = \int_Y \mathbb{1}_{Y_j}(y)\nu(E^y \cap X_j)\,\mathrm{d}\nu(y).$$

Applying the Monotone Convergence Theorem again, we conclude the claim, and hence the Theorem. □

Theorem 12.4 (Fubini–Tonelli)**.** *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces.*
(i) (Tonelli) Let $f \in L^+(X \times Y)$ be $\mathcal{A} \otimes \mathcal{B}$-measurable. Then the functions

$$g(x) := \int_Y f_x(y)\,\mathrm{d}\nu(y), \qquad h(x) := \int_X f^y(x)\,\mathrm{d}\mu(x)$$

are measurable and in $L^+(X)$, $L^+(Y)$ respectively and

$$\int_{X\times Y} f\,\mathrm{d}\pi = \int_X \Big(\int_Y f(x,y)\,\mathrm{d}\nu(y)\Big)\mathrm{d}\mu(x) = \int_X g\,\mathrm{d}\mu \tag{1}$$

$$= \int_Y \Big(\int_X f(x,y)\,\mathrm{d}\mu(x)\Big)\mathrm{d}\nu(y) = \int_Y h\,\mathrm{d}\nu. \tag{2}$$

(ii) (Fubini) Let $f \in \mathcal{L}^1(\mu\otimes\nu)$. Then the functions $f_x \in \mathcal{L}^1(\nu)$ for μ-a.e. $x \in X$, $f^y \in \mathcal{L}^1(\mu)$ for ν-a.e. $y \in Y$, and the a.e. defined functions

$$g(x) := \int_Y f_x(y)\,\mathrm{d}\nu(y), \qquad h(x) := \int_X f^y(x)\,\mathrm{d}\mu(x)$$

are measurable and in $\mathcal{L}^1(\mu)$, $\mathcal{L}^1(\nu)$ respectively and (1) *holds.*

Proof. (i) Notice first that when $f = \mathbb{1}_E$ for some measurable E, then the result follows from Theorem 12.3 as, in this case, the function g is the map $x \mapsto \nu(E_x)$ and h is $y \mapsto \mu(E^y)$.

By linearity of the integrals (and recalling that linear combinations of measurable functions are measurable), we therefore have Tonelli's Theorem for simple functions.

Now for a general $f \in L^+$, we take a sequence of simple functions $\{f_n\}_{n=1}^\infty$ such that f_n is monotone increasing and converges to f using Theorem 8.10. We define a sequence

$$g_n(x) := \int_Y (f_n)_x(y)\,\mathrm{d}\nu(y), \qquad h_n(x) := \int_X (f_n)^y(x)\,\mathrm{d}\mu(x).$$

By the Monotone Convergence Theorem, $g_n \to g$ pointwise and $h_n \to h$ pointwise. Therefore, starting from (1) for simple functions, we have

$$\int_{X\times Y} f_n\,\mathrm{d}\pi = \int_X \underbrace{\Big(\int_Y f_n(x,y)\,\mathrm{d}\nu(y)\Big)}_{=g_n(x)}\mathrm{d}\mu(x) = \int_Y \underbrace{\Big(\int_X f_n(x,y)\,\mathrm{d}\mu(x)\Big)}_{=h_n(y)}\mathrm{d}\nu(y)$$

and, sending $n \to \infty$, we again apply the Monotone Convergence Theorem three times to see that

$$\int_{X\times Y} f_n\,\mathrm{d}\pi \to \int_{X\times Y} f\,\mathrm{d}\pi,$$
$$\int_X g_n(x)\mathrm{d}\mu(x) \to \int_X g(x)\mathrm{d}\mu(x) = \int_X \Big(\int_Y f(x,y)\,\mathrm{d}\nu(y)\Big)\mathrm{d}\mu(x),$$
$$\int_X h_n(y)\mathrm{d}\nu(y) \to \int_Y h(y)\mathrm{d}\nu(y) = \int_Y \Big(\int_X f(x,y)\,\mathrm{d}\mu(x)\Big)\mathrm{d}\nu(y),$$

which concludes the proof of Tonelli's Theorem.

(ii) We write $f = f^+ - f^-$ where $f^+, f^- \in L^+$. Applying Tonelli's Theorem to each of f^+ and f^- separately and then using linearity of the integral, we conclude the proof of Fubini's Theorem. ☐

Remark 12.5. We can slightly weaken the assumptions of the theorem to say that if any of the three integrals

$$\int_{X\times Y} |f|\,\mathrm{d}\pi, \qquad \int_X \Big(\int_Y |f|(x,y)\,\mathrm{d}\nu(y)\Big)\mathrm{d}\mu(x), \qquad \int_Y \Big(\int_X |f|(x,y)\,\mathrm{d}\mu(x)\Big)\mathrm{d}\nu(y)$$

is finite, then all three are finite and coincide and the conclusion of Fubini's Theorem holds.

It is reasonable to ask whether there is a straightforward way to see that a function is in $L^1(\mu)$ in order to be able to apply Fubini. As a general rule, one first considers $|f|$ and applies Tonelli to compute the integral $\int_{X\times Y} |f| \, \mathrm{d}\pi$. If this is finite, then $f \in L^1(\mu)$ and we may apply Fubini to compute $\int_{X\times Y} f \, \mathrm{d}\pi$. If it is not, we do not know what to do!

In general, if $f \notin L^1(\mu)$, the exchange of integrals may well fail. There will be examples of this below and in your homework.

Recall from above that the product of the Lebesgue σ-algebras on $\mathbb{R}^m$ and $\mathbb{R}^k$ is not the Lebesgue σ-algebra on $\mathbb{R}^n$. Nonetheless, Fubini's Theorem still holds for Lebesgue measurable functions.

Theorem 12.6. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be Lebesgue measurable and suppose $n = m + k$. Then $f \in \mathcal{L}^1(\lambda^n)$ if and only if one of the integrals*

$$\int_{\mathbb{R}^k} \Big(\int_{\mathbb{R}^m} |f|(x,y) \, \mathrm{d}\lambda^m(x) \Big) \, \mathrm{d}\lambda^k(y), \qquad \int_{\mathbb{R}^m} \Big(\int_{\mathbb{R}^k} |f|(x,y) \, \mathrm{d}\lambda^k(y) \Big) \, \mathrm{d}\lambda^m(x)$$

is finite. In this case

$$\int_{\mathbb{R}^n} f(x,y) \, \mathrm{d}\lambda^n = \int_{\mathbb{R}^k} \Big(\int_{\mathbb{R}^m} f(x,y) \, \mathrm{d}\lambda^m(x) \Big) \, \mathrm{d}\lambda^k(y) = \int_{\mathbb{R}^m} \Big(\int_{\mathbb{R}^k} f(x,y) \, \mathrm{d}\lambda^k(y) \Big) \, \mathrm{d}\lambda^m(x)$$

where all the inner integrals are well-defined a.e.

We now give a couple of **Counterexamples** where Fubini's theorem does not apply.

Example 12.7. Let $X = Y = \{1, 2, \ldots\}$ with $\mu = \nu =$ counting measure (then $\mu \otimes \nu$ is also counting measure on $X \times Y$). Put $f(m,m) = 1$, $f(m, m+1) = -1$ for all $m \geq 1$, and put $f(m,n) = 0$ otherwise. Then

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} f(m,n) = 0 \qquad \text{and} \qquad \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} f(m,n) = 1.$$

Here $f \not\geq 0$ and $f \notin L^1(\mu \otimes \nu)$. (The integrals are equal to sums, i.e. $\int f(n,m)\nu(dm) = \sum_m f(n,m)$ and so on, by one of the Exercises on the problem sheets).

Example 12.8. Let $X = (0,1)$, $\mathcal{A} = \mathcal{B}_{(0,1)}$ (Borel), $\mu =$ Lebesgue measure, and let $Y = (0,1)$, $\mathcal{B} = \mathcal{P}((0,1))$ (all subsets of $(0,1)$), and $\nu =$ counting measure. Let $f(x,y) = 1$ for $x = y$ and $f(x,y) = 0$ otherwise. Then

$$\begin{aligned} \int_X \left[\int_Y f(x,y) \, \nu(dy) \right] \mu(dx) &= \int_X 1 \, \mathrm{d}\mu = 1 \\ \int_Y \left[\int_X f(x,y) \, \mu(dx) \right] \nu(dy) &= \int_Y 0 \, d\nu = 0. \end{aligned}$$

Here ν is not σ-finite (actually we've not defined integration with respect to non-σ-finite measures, but if we did, we'd start with the formula for integrating simple functions).

13 The $L^p(\mu)$ spaces

In this final section, we will extend the theory developed above for the space L^1 to a wider family of spaces: the $L^p(\mu)$ spaces. These spaces are of enormous importance in modern analysis and, in fact, provide the language in which many of the most significant theorems of modern analysis are stated.

Throughout this section we will have an exponent $1 \leq p \leq \infty$ and a measure space $(X, \mathcal{A}, \mu)$. Similarly to the case of $\mathcal{L}^1(\mu)$ and $L^1(\mu)$, we begin by defining the functionals

$$\|f\|_{L^p(\mu)} = \begin{cases} \Big(\int_X |f|^p \, \mathrm{d}\mu \Big)^{\frac{1}{p}} & \text{for } 1 \leq p < \infty, \\ \inf \big\{ a \geq 0 \,|\, \mu(\{x \in X \,|\, |f(x)| > a\}) = 0 \big\} & \text{for } p = \infty. \end{cases} \tag{1}$$

Then the spaces $\mathcal{L}^p(\mu)$ are defined as

$$\mathcal{L}^p(\mu) := \big\{ f : X \to \mathbb{R} \,|\, f \text{ is measurable, } \|f\|_{L^p(\mu)} < \infty \big\}. \tag{2}$$

The *Lebesgue spaces* $L^p(\mu)$ are defined to be the quotients

$$L^p(\mu) := \mathcal{L}^p(\mu)/\sim, \qquad \text{where} \qquad f \sim g \text{ if } f = g \text{ a.e.} \tag{3}$$

From henceforth, we will restrict attention to $L^p(\mu)$ and, in a slight abuse of notation, will refer directly to measurable functions $f \in L^p$ rather than equivalence classes of functions.

Notation: Where the measure space $(X, \mathcal{A}, \mu)$ is clear, we will commonly write $\|\cdot\|_p$ for $\|\cdot\|_{L^p(\mu)}$.

Example 13.1. When $\mu = \lambda_1$ is Lebesgue measure on the Borel sets in $\mathbb{R}$, we write $L^p(\mathbb{R})$, or even just L^p, for $L^p(\lambda_1)$.

Likewise, if $W \subset \mathbb{R}$ with $W \in \mathcal{B}$ (e.g. $W = [0,1]$) we write $L^p(W)$ for the L^p space on $(X, \mathcal{A}, \mu) = (W, \mathcal{B}_W, \lambda_1|_W)$, where $\lambda_1|_W$ denotes the restriction of 1-dimensional Lebesgue measure to $(W, \mathcal{B}_W)$.

Example 13.2. Suppose $\mathcal{A} = \mathcal{P}(X)$ and μ is counting measure.

(a) If $X = \{1, \ldots, d\}$ for some $d \in \mathbb{N}$, then for Borel measurable functions $f : X \to \mathbb{R}$, by a homework exercise,

$$\|f\|_p = \left(\sum_{i=1}^{d} |f(i)|^p \right)^{1/p}$$

which is called the ℓ^p-norm of the d-dimensional vector $(f(1), \ldots, f(d))$. For example, the ℓ^2-norm is the same as the Euclidean norm on $\mathbb{R}^d$.

(b) If $X = \mathbb{N}$ then by homework exercise, for Borel measurable functions $f : X \to \mathbb{R}$,

$$\|f\|_p = \left(\sum_{i=1}^{\infty} |f(i)|^p \right)^{1/p}$$

which is called the ℓ^p norm of the infinite sequence $(f(1), f(2), f(3), \ldots)$. In this case we write ℓ^p for $L^p(\mu)$, which is in effect the space of infinite sequences $(x_1, x_2, \ldots)$ such that $\sum_{i=1}^{\infty} |x_i|^p < \infty$) (here writing x_i for $f(i)$).

The first thing to check is that $L^p(\mu)$ is a vector space.

Lemma 13.3. *For $1 \le p \le \infty$, $L^p(\mu)$ is a vector space.*

Proof. Clearly if $\lambda \in \mathbb{R}$, then we have, for $f \in L^p(\mu)$, $\|\lambda f\|_p = |\lambda|\|f\|_p < \infty$, so $\lambda f \in L^p(\mu)$.
If $f, g \in L^p(\mu)$ and $p < \infty$, then we may estimate, pointwise a.e.,

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

Thus

$$\int |f+g|^p \,\mathrm{d}\mu \le 2^p \int |f|^p \,\mathrm{d}\mu + 2^p \int |g|^p \,\mathrm{d}\mu < \infty.$$

If $p = \infty$, then we may let A_f and A_g respectively be the null sets such that $|f| \le \|f\|_\infty$ on A_f^c, $|g| \le \|g\|_\infty$ on A_g^c. Then on $(A_f \cup A_g)^c$ (which is again the complement of a null set), we have $|f+g| \le \|f\|_\infty + \|g\|_\infty$, and so $f+g \in L^\infty(\mu)$. $\square$

We see from the above proof that $\|\cdot\|_\infty$ satisfies the triangle inequality, and hence defines a norm on L^∞. However, we have not shown the triangle inequality when $p < \infty$. This is because our estimate on sums of L^p functions is rather crude. To show that the triangle inequality does hold, we will need to do a bit more work. First, we prove the following technical lemma.

Lemma 13.4. *Let $a \ge 0$, $b \ge 0$, $\lambda \in (0,1)$. Then $a^\lambda b^{1-\lambda} \le \lambda a + (1-\lambda)b$.*

Proof. First, note that if $b = 0$, the inequality is trivial.
Suppose $b > 0$. Then let $f(t) := t^\lambda - \lambda t$. Differentiating f, we have $f'(t) = \lambda t^{\lambda-1} - \lambda$. This vanishes when $t^{\lambda-1} = 1$, i.e. only for $t = 1$. $f''(1) = \lambda(\lambda-1) < 0$, so $t = 1$ is a maximum for f. Thus $t^\lambda - \lambda t \le f(1) = 1 - \lambda$. Therefore evaluating at $t = \frac{a}{b}$, we obtain

$$\frac{a^\lambda}{b^\lambda} - \lambda\frac{a}{b} \le 1 - \lambda,$$

which rearranges to the desired inequality. $\square$

The next result we will prove is Hölder's inequality. While this inequality may look strange initially, it is very useful in practice and crops up frequently throughout Functional Analysis, PDE Theory, and a range of other fields related to analysis. Essentially, it tells us how to relate the L^1 norm of products of functions in different L^p spaces.

Proposition 13.5 (Hölder Inequality)**.** *Let $p, q \in [1,\infty]$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$ (if $q = \infty$, then $p = 1$ and vice versa). Then, for any measurable $f \in L^p(\mu)$, $g \in L^q(\mu)$,*

$$\int fg \,\mathrm{d}\mu \le \|f\|_p \|g\|_q.$$

Proof. First, consider the case $p = 1$, $q = \infty$. Then on the full measure set on which $|g| \le \|g\|_\infty$, we have the pointwise bound $|fg| \le |f|\|g\|_\infty$, and hence the trivial bound

$$\int fg \,\mathrm{d}\mu \le \int |fg| \,\mathrm{d}\mu \le \int |f| \,\mathrm{d}\mu \|g\|_\infty = \|f\|_1 \|g\|_\infty.$$

Next, suppose $p, q > 1$ and that $\|f\|_p = 0$ (or $\|g\|_q = 0$). Then by definition of $\|f\|_p$, we conclude

$$\int |f|^p \,\mathrm{d}\mu = 0 \text{ and hence } |f| = 0 \text{ a.e.}$$

which means $f = 0$ in L^p. Thus also

$$\int fg \,\mathrm{d}\mu = \int 0 \,\mathrm{d}\mu = 0 = \|f\|_p\|g\|_q.$$

We now move onto the non-trivial cases. Again suppose $p, q > 1$ and that $\|f\|_p, \|g\|_q > 0$. Then, by considering the functions $\bar{f} = \frac{f}{\|f\|_p}$ and $\bar{g} = \frac{g}{\|g\|_q}$, we observe that $\|\bar{f}\|_p = \|\bar{g}\|_q = 1$ and that Hölder's inequality is equivalent to

$$\int \bar{f}\bar{g} \,\mathrm{d}\mu \leq 1.$$

We therefore seek to prove this inequality, assuming without loss of generality that $\|f\|_p = \|g\|_q = 1$.

By Lemma 13.4, applied with $a = |f|^p$ and $b = |g|^q$, $\lambda = \frac{1}{p}$, $1 - \lambda = \frac{1}{q}$, we have

$$|f||g| \leq \frac{1}{p}|f|^p + \frac{1}{q}|g|^q.$$

Integrating, we have

$$\begin{aligned}\Big|\int fg \,\mathrm{d}\mu\Big| \leq \int |f||g| \,\mathrm{d}\mu \leq \int \big(\frac{1}{p}|f|^p + \frac{1}{q}|g|^q\big) \,\mathrm{d}\mu = \frac{1}{p}\int |f|^p \,\mathrm{d}\mu + \frac{1}{q}\int |g|^q \,\mathrm{d}\mu \\ = \frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.\end{aligned}$$

□

One consequence of this (though not an immediate consequence) is that the dual space of L^p is isometric to L^q, where $\frac{1}{p} + \frac{1}{q} = 1$.

Definition 13.6. If $1 \leq p, q \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, we say that p and q are **Hölder conjugate exponents**. (A special case is $p = q = 2$.)

Given $p > 1$, its conjugate exponent q is given in terms of p by $q = 1/(1 - (1/p))$, i.e. $q = p/(p-1)$.

Hölder's inequality is the tool we needed to prove the triangle inequality in $L^p(\mu)$. This inequality often goes by the name of Minkowski's inequality, but should not be confused with a different inequality, also called Minkowski's inequality and related to $L^p(\mu)$ spaces.

Proposition 13.7 (Minkowski Inequality)**.** *Let* $1 \leq p \leq \infty$ *and suppose* $f, g \in L^p$. *Then*

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

Proof. This is trivial in the case $p = 1$ by the usual triangle inequality in $\mathbb{R}$, and we have already seen a proof in the case $p = \infty$. It therefore remains to prove it in the case $1 < p < \infty$.

We assume moreover that $\|f + g\|_p > 0$, else the inequality is trivial. Note first that, setting $q = 1 - \frac{1}{p} = \frac{p-1}{p}$, we have

$$\Big(\int |f + g|^{(p-1)q} \,\mathrm{d}\mu\Big)^{1-\frac{1}{p}} = \Big(\int |f + g|^p \,\mathrm{d}\mu\Big)^{1-\frac{1}{p}} = \frac{\|f + g\|_p^p}{\|f + g\|_p}.$$

So we have checked that

$$\||f+g|^{p-1}\|_q = \frac{\|f+g\|_p^p}{\|f+g\|_p}.$$

Now we write

$$\begin{aligned}\|f+g\|_p^p = \int |f+g|^p \,\mathrm{d}\mu = \int |f+g||f+g|^{p-1}\,\mathrm{d}\mu \le \int |f||f+g|^{p-1}\,\mathrm{d}\mu + \int |g||f+g|^{p-1}\,\mathrm{d}\mu \\ \le \|f\|_p\||f+g|^{p-1}\|_q + \|g\|_p\||f+g|^{p-1}\|_q = \big(\|f\|_p + \|g\|_p\big)\frac{\|f+g\|_p^p}{\|f+g\|_p},\end{aligned}$$

where we have applied Hölder's inequality in the first inequality of the second line and the previous identity in the last equality. This now rearranges to the desired inequality. □

Having shown that $L^p(\mu)$ is a normed vector space, we may ask whether it has any of the nice properties that make it amenable to Functional Analytic techniques. In general, we cannot expect that $L^p(\mu)$ will be a Hilbert space (though $L^2(\mu)$ is an extremely important example of a Hilbert space), but we can show that $L^p(\mu)$ is a Banach space (i.e., it is complete).

Theorem 13.8. *Let $1 \le p \le \infty$. Then $L^p(\mu)$ is a Banach space.*

Proof. Let $1 \le p < \infty$. We use the characterisation of a Banach space that states that a normed vector space is complete if and only if every absolutely convergent series in the space converges. We therefore let $\{f_k\} \subset L^p(\mu)$ be such that $\sum_{k=1}^\infty \|f_k\|_p = A < \infty$. Denoting by G_n the partial sum function $\sum_{k=1}^n |f_k|$ and $G = \sum_{k=1}^\infty |f_k|$. By Minkowski's inequality, we have

$$\|G_n\|_p = \Big\|\sum_{k=1}^n |f_k|\Big\|_p \le \sum_{k=1}^n \|f_k\|_p \le A \text{ for all } n.$$

Then, by the Monotone Convergence Theorem, we take the limit to see

$$\|G\|_p^p = \int G^p \,\mathrm{d}\mu = \lim_{n\to\infty}\int G_n^p \,\mathrm{d}\mu \le A^p.$$

Hence $G \in L^p$ and, in particular, $G(x) < \infty$ for a.e. x. Therefore, for a.e. x, the series $\sum_{k=1}^\infty f_k(x)$ converges to a value $F(x)$. Clearly $|F| \le G$, and therefore $F \in L^p(\mu)$.

We still need to show that $\sum_{k=1}^n f_k \to F$ in $L^p(\mu)$ (i.e. in the topology induced by the $L^p(\mu)$ norm), which we do by observing that

$$\Big|F - \sum_{k=1}^n f_k\Big|^p \le (2G)^p \in L^1(\mu)$$

so that we may apply the Dominated Convergence Theorem and observe

$$\Big\|F - \sum_{k=1}^n f_k\Big\|_p^p = \int \Big|F - \sum_{k=1}^n f_k\Big|^p \,\mathrm{d}\mu \to 0,$$

as required.

The case $p = \infty$ is left as an exercise. □

Lemma 13.9. *Let $1 \le p < \infty$. Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be such that $f_n \to f$ in $L^p(\mu)$. Then there exists a subsequence $(f_{n_k})_{k\ge1}$ such that $f_{n_k} \to f$ a.e.*

The proof of this lemma is non-examinable.

14 Convex functions in measure theory

The Hölder and Minkowski inequalities were key results in our discussion of L^p spaces in Section 13. We now provide a final inequality, Jensen's inequality, which is especially important in probability theory, and concerns *convex* functions.

Notation. Assume throughout this section that $(X, \mathcal{A}, \mu)$ is a σ-finite measure space. Assume also that a, b are given with $-\infty \leq a < b \leq \infty$, and let $I := (a, b)$. That is, $I \subset \mathbb{R}$ is an open interval (possibly the whole of $\mathbb{R}$).

Given $f \in \mathbb{R}(X)$, in this section we shall often write $\mu(f)$ for $\int_X f d\mu$.

We say that $h : \mathbb{R} \to \mathbb{R}$ is a **linear function** (also known as an **affine function**) if there are constants $c, d \in \mathbb{R}$ such that $h(x) = cx + d$ for all $x \in \mathbb{R}$. (This is not the same as a 'linear map' in linear algebra, for which the second constant d would have to be zero.)

Definition 14.1. A function $\varphi : I \to \mathbb{R}$ is said to be **convex** if

$$\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y) \tag{1}$$

for all $x, y \in I$ and for all $0 \leq \alpha \leq 1$.

In other words, φ is convex if for all $x, y \in I$, the linear function h that satisfies $h(x) = \varphi(x)$ and $h(y) = \varphi(y)$ satisfies $h(z) \geq \varphi(z)$ for all $z \in [x, y]$.

How to check convexity? A continuously differentiable function φ is convex on $I = (a, b)$, if $\varphi'(s) \leq \varphi'(t)$ for all $a < s < t < b$.

Indeed, if $a < x < y < b$ and $0 < \alpha < 1$, then setting $z = \alpha x + (1-\alpha)y$, by the Intermediate Value theorem we have for some $u \in (x, z)$ and $v \in (z, y)$ that

$$\begin{aligned}
\frac{\varphi(y)-\varphi(x)}{y-x} &= \Big(\frac{z-x}{y-x}\Big)\Big(\frac{\varphi(z)-\varphi(x)}{z-x}\Big) + \Big(\frac{y-z}{y-x}\Big)\Big(\frac{\varphi(y)-\varphi(z)}{y-z}\Big) \\
&= \Big(\frac{z-x}{y-x}\Big)\varphi'(u) + \Big(\frac{y-z}{y-x}\Big)\varphi'(v) \\
&\geq \Big(\frac{z-x}{y-x}\Big)\varphi'(u) + \Big(\frac{y-z}{y-x}\Big)\varphi'(u) = \varphi'(u) \\
&= \frac{\varphi(z)-\varphi(x)}{z-x},
\end{aligned}$$

where for the inequality in the third line we used that $f'(u) \leq f'(v)$ since $u < v$. The convexity follows.

Therefore if φ is twice differentiable on I, it is convex if $\varphi''(x) \geq 0$ for all $x \in I$.

Examples. (i) Suppose $c \in \mathbb{R}$, and $I = (-\infty, \infty)$ and $\varphi(x) = e^{cx}$ for all $x \in I$. Then $\varphi''(x) = c^2 e^x > 0$ for all $x \in I$ so φ is convex on I.

(ii) Suppose $p \in (1, \infty)$, and $I = (0, \infty)$, and we set $\varphi(x) = x^p$ for all $x \in I$. Then $\varphi''(x) = p(p-1)x^{p-2} > 0$ for all $x \in I$ so φ is convex on I.

As preparation for the last result of this section (Jensen's inequality), we shall use the following result from elementary analysis.

Lemma 14.2. *Suppose $\varphi : I \to \mathbb{R}$ is convex on I. then:*

(i) φ is continuous on I.

(ii) Given $t \in I$, there exists a linear function $h : \mathbb{R} \to \mathbb{R}$ such that $h(t) = \varphi(t)$ and $h(x) \leq \varphi(x)$ for all $x \in I$.

Proof. (Non-examinable) (i) Fix $t \in (a,b)$. We show that φ is continuous at t. Fix any $a < u < t < v < b$. Then for $u < s < v$ we have:

$$\frac{\varphi(t)-\varphi(u)}{t-u} \le \frac{\varphi(s)-\varphi(t)}{s-t} \le \frac{\varphi(v)-\varphi(t)}{v-t}.$$

Hence

$$|\varphi(s)-\varphi(t)| \le |s-t| \max\left\{ \left|\frac{\varphi(t)-\varphi(u)}{t-u}\right|, \left|\frac{\varphi(v)-\varphi(t)}{v-t}\right| \right\}.$$

This shows that as $s \to t$, $\varphi(s) \to \varphi(t)$. (We even get the stronger statement that φ is Lipschitz on any compact subinterval of (a,b).)

(ii) Let

$$\beta := \sup_{s: a<s<t} \frac{\varphi(t)-\varphi(s)}{t-s}.$$

Then we have

$$\varphi(s) \ge \varphi(t) + \beta(s-t), \quad a < s < t. \tag{2}$$

On the other hand, it follows from (1) that for all $a < s < t < u < b$ we have

$$\frac{\varphi(t)-\varphi(s)}{t-s} \le \frac{\varphi(u)-\varphi(t)}{u-t} \tag{3}$$

and therefore

$$\beta \le \frac{\varphi(u)-\varphi(t)}{u-t}, \quad t < u < b,$$

so we also have

$$\varphi(u) \ge \varphi(t) + \beta(u-t), \quad t < u < b. \tag{4}$$

Renaming u to be s in (4), together with (2), we get:

$$\varphi(s) \ge \varphi(t) + \beta(s-t), \quad a < s < b.$$

Therefore taking $h(x) := \varphi(t) + \beta(x-t)$ for $x \in I$, gives us a linear function h having the stated properties. □

Theorem 14.3. (Jensen's Inequality) *Suppose $\mu(X) = 1$ (so $(X, \mathcal{A}, \mu)$ is a probability space), and $f \in L^1(\mu)$ with $f(X) \subset I$, and $\varphi : I \to \mathbb{R}$ is convex on I. then*

$$\varphi\left(\int_X f \, \mathrm{d}\mu\right) \le \int_X (\varphi \circ f) \, \mathrm{d}\mu, \tag{5}$$

that is, $\varphi(\mu(f)) \le \mu(\varphi \circ f)$, where we set $\varphi \circ f(x) = \varphi(f(x))$ for all $x \in X$.

Remark. By Lemma 14.2(i), φ is continuous so $\varphi \circ f$ (defined by $\varphi \circ f(x) = \varphi(f(x))$ is measurable. It may happen that $(\varphi \circ f) \notin L^1(\mu)$. In this case the proof below will show that the right hand side of (5) is $+\infty$.

Proof of Theorem. (Non-examinable) Denote $t := \mu(f)$. Then $t \in I$; this can be proved using linearity (Theorem **??**) and Lemma **??** (HW).

Using Lemma 14.2(ii), choose constants $c, c' \in \mathbb{R}$ such that the linear function $h(x) := cx + c'$ satisfies $h(t) = \varphi(t)$ and $h \le \varphi$ pointwise on I. Therefore integrating, and using linearity (Theorem **??**), and also the fact that μ is a probability measure so $\int_X c' \mathrm{d}\mu = c'$, we obtain that

$$\mu(\varphi \circ f) \ge \mu(h \circ f) = \mu(cf + c') = c\mu(f) + c' = h(\mu(f)) = \varphi(\mu(f)),$$

as required. □

Remark. As mentioned earlier, when working on a probabililty space (as is the case here) the integral $\int_X f d\mu$ is known as the *expectation* of f, often denoted $\mathbb{E}[f]$. Jensen's inequality says that $\varphi(\mathbb{E}[f]) \le \mathbb{E}[\varphi \circ f]$ for convex φ.

Example 14.4. In all of the following examples we are assuming μ is a probability measure.

(i) Take $\varphi(x) = x^2$. Then $\varphi'(x) = 2x$ and $\varphi'(x) = 2$ so φ is convex on $\mathbb{R}$. In probabilistic notation, Jensen's inequality tells us that $(\mathbb{E}[f])^2 \le \mathbb{E}[f^2]$, for any $f \in L^1(\mu)$ which is well known from probability theory.

(ii) Take $\varphi(x) = e^x$. Then $\varphi''(x) = e^x$ so φ is convex on $\mathbb{R}$. So by Jensen's inequality, $\exp(\mu(f)) \le \mu(e^f)$, whenever $f \in L^1(\mu)$.

(iii) Suppose X is finite, $X = \{x_1, \dots, x_n\}$, with $\mu(\{x_i\}) = \frac{1}{n}$, $f(x_i) = a_i$. Then the previous example specializes to:

$$\exp\left\{\frac{1}{n}(a_1 + \cdots + a_n)\right\} \le \frac{1}{n}(e^{a_1} + \cdots + e^{a_n}).$$

Now putting $b_i = e^{a_i}$, we get the familiar inequality between the geometric and arithmetic mean:

$$(b_1 \cdots b_n)^{1/n} \le \frac{1}{n}(b_1 + \cdots + b_n).$$

15 Applications to probability theory

(Non-examinable)
This section summarizes some applications of measure theory to probability. This distinction is of course a bit artificial, because all of measure theory applies to probability spaces.

Throughout this section we consider a probability space (that is, a measure space with total measure 1) denoted $(\Omega, \mathcal{F}, \mathbb{P})$ instead of $(X, \mathcal{A}, \mu)$.

As mentioned before, a measurable function from Ω to $\mathbb{R}$ is called a **random variable** (RV). Following the usual conventions of probability theory, in this section we use capital letters such as X, Y etc. (rather than f, g etc.) to denote random variables.

As mentioned at the start of Section 9, the **Expectation** of a random variable X is defined to be the integral $\int_\Omega X d\mathbb{P}$, and denoted by $\mathbb{E}[X]$.

Also, in probability theory a measurable subset of Ω (that is, a set that is an element of $\mathcal{F}$) is called an **event**.

Definition 15.1. Suppose $X : \Omega \to \mathbb{R}$ is a random variable (RV) (that is: a measurable function). We define **the σ-algebra generated by** X as

$$\sigma(X) := \{X^{-1}(B) : B \subset \mathbb{R},\ B \in \mathcal{B}\},$$

where we recall that $\mathcal{B}$ denotes the Borel sets in $\mathbb{R}$.

Remarks. We claim that $\sigma(X)$ is a **sub-σ-algebra** of $\mathcal{F}$. That is, (i) $\sigma(X)$ is a σ-algebra, and (ii) $\sigma(X) \subset \mathcal{F}$. We leave it as HW to check (i), while (ii) is a consequence of Theorem **??**.

Moreover X is measurable with respect to $\sigma(X)$ since for any $\alpha \in \mathbb{R}$ we have $(\alpha, \infty) \in \mathcal{B}$ so $X^{-1}((\alpha, \infty]) \in \sigma(X)$.

In fact, $\sigma(X)$ is the smallest σ-algebra with respect to which X is measurable (HW).

Definition 15.2. Events $A, B \in \mathcal{F}$ are said to be **independent**, if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$.

We say that σ-algebras $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$ are **independent**, if for any $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ the events A and B are independent.

Suppose $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are RVs. The RVs X and Y are said to be **independent**, if $\sigma(X)$ and $\sigma(Y)$ are independent σ-algebras. In other words, X and Y are independent if for all $A, B \in \mathcal{B}$,

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B],$$

where $\{X \in A\} := X^{-1}(A)$ and $\{X \in A, Y \in B\} := X^{-1}(A) \cap Y^{-1}(B)$.

Suppose $Z : \Omega \to \mathbb{R}$ is a further RV. The RVs X, Y and Z are said to be **mutually independent**, if for all $A, B, C \in \mathcal{B}$ we have

$$\mathbb{P}[X \in A, Y \in B, Z \in C] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B], \mathbb{P}[Z \in C].$$

Theorem 15.3 (independence criterion). *Suppose $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are RVs. If $\mathbb{P}[X \le a, Y \le b] = \mathbb{P}[X \le a]\mathbb{P}[Y \le b]$ for all $a \in \mathbb{R}, b \in \mathbb{R}$, then X and Y are independent.*

Proof. (Outline). Let $a \in \mathbb{R}$. Define the collections of sets

$$\mathcal{L}_a := \{B \in \mathcal{B} : \mathbb{P}[X \leq a, Y \in B] = \mathbb{P}[X \leq a]\mathbb{P}[Y \in B]\};$$
$$\mathcal{D} := \{(-\infty, b] : b \in \mathbb{R}\}.$$

Then one can show that $\mathcal{L}$ is a λ-system. Also, by assumption, $\mathcal{D} \subset \mathcal{L}$, and $\mathcal{D}$ is a π-system which generates $\mathcal{B}$. Therefore by Dynkin's π-λ theorem (Theorem **??**), $\mathcal{L} \supset \sigma(\mathcal{D}) = \mathcal{B}$.

Now, for $B \in \mathcal{B}$, set

$$\mathcal{L}_B := \{A \in \mathcal{B} : \mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]\}$$

Again, this is a λ-system and by the previous paragraph, it contains $\mathcal{D}$, so by Dynkin's π-λ theorem, $\mathcal{L}_B \supset \sigma(\mathcal{D}) = \mathcal{B}$. The conclusion follows. □

Theorem 15.4 (Independence and expectation). *If X and Y are integrable RVs on $(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*

Proof. First suppose X, Y are nonnegative and simple, i.e. $X(\Omega)$ and $Y(\Omega)$ are finite and contained in $[0, \infty)$. Enumerate the possible values of X (i.e., the elements of $X(\Omega)$) as $a_1, \ldots, a_n$ (taken to be distinct) and the possible values of Y as $b_1, \ldots, b_m$ (taken to be distinct). Set $A_i = X^{-1}(\{a_i\})$ for $1 \leq i \leq n$ and set $B_j = Y^{-1}(\{b_j\})$ for $1 \leq j \leq m$. Then $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ and $Y = \sum_{j=1}^m b_j \mathbf{1}_{B_j}$ so

$$XY = \left(\sum_{i=1}^n a_i \mathbf{1}_{A_i}\right)\left(\sum_{j=1}^m b_j \mathbf{1}_{B_j}\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbf{1}_{A_i} \mathbf{1}_{B_j} = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbf{1}_{A_i \cap B_j}.$$

Then $A_i \in \sigma(X)$ and $B_j \in \sigma(Y)$ for all i, j so by independence, and Lemma 9.8,

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{P}[A_i \cap B_j] \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{P}[A_i]\mathbb{P}[B_j] \\ &= \left(\sum_{i=1}^n a_i \mathbb{P}[A_i]\right)\left(\sum_{j=1}^m b_j \mathbb{P}[B_j]\right) = \mathbb{E}[X]\mathbb{E}[Y], \end{aligned}$$

so the result holds in this case.

Next, suppose only that $X \geq 0$ (i.e., $X(\Omega) \subset [0, \infty)$) and $Y \geq 0$. Then by Theorem 8.10, there exist nonnegative simple random variables X_n and Y_n (for $n \in \mathbb{N}$) such that $X_n \uparrow X$ and $Y_n \uparrow Y$ (pointwise) as $n \to \infty$. Then by the algebra of limits we also have $X_n Y_n \uparrow XY$ pointwise, and hence by MON (Theorem **??**) and the case considered already,

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n].$$

Hence by the algebra of limits and MON again,

$$\mathbb{E}[XY] = (\lim_{n\to\infty} \mathbb{E}[X_n])(\lim_{n\to\infty} \mathbb{E}[Y_n]) = \mathbb{E}[X]\mathbb{E}[Y],$$

so the result holds in this case.

In the general case, we have $(XY)^+ = X^+Y^+ + X^-Y^-$, and $(XY)^- = X^+Y^- + X^-Y^+$. Therefore by linearity (Theorem **??**),

$$\mathbb{E}[XY] = \mathbb{E}[(XY)^+] - \mathbb{E}[(XY)^-] = \mathbb{E}[X^+Y^+] + \mathbb{E}[X^-Y^-] - \mathbb{E}[X^+Y^-] - \mathbb{E}[X^-Y^+],$$

provided all of these expectations are finite. Hence by the case considered previously,

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[X^+]\mathbb{E}[Y^+] + \mathbb{E}[X^-]\mathbb{E}[Y^-] - \mathbb{E}[X^+]\mathbb{E}[Y^-] - \mathbb{E}[X^-]\mathbb{E}[Y^+] \\ &= (\mathbb{E}[X^+] - \mathbb{E}[X^-])(\mathbb{E}[Y^+] - \mathbb{E}[Y^-]) = \mathbb{E}[X]\mathbb{E}[Y],\end{aligned}$$

and all of the expectations mentioned are finite because we assumed integrability of X and Y. The proof is complete. □

Theorem 15.5 (**Existence of conditional expectation).** *Let $X : \Omega \to \mathbb{R}$ be a RV, such that $X \in L^1(\mathbb{P})$. Suppose that $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra. There exists a RV $Y : \Omega \to \mathbb{R}$, such that:*
(i) Y is measurable with respect to $\mathcal{G}$;
(ii) $Y \in L^1(\mathbb{P})$;
(iii) For every $B \in \mathcal{G}$ we have:

$$\int_B Y \, d\mathbb{P} = \int_B X \, d\mathbb{P}.$$

Moreover, Y is unique up to a.e.[$\mathbb{P}$] equivalence.

Definition 15.6. The RV Y constructed in Theorem 15.5 is called the **conditional expectation of X given $\mathcal{G}$**, and is denoted $Y =: \mathbb{E}[X \,|\, \mathcal{G}]$.

Proof of Theorem 15.5. We prove the statement when $X \geq 0$. The general case follows by considering the positive and negative parts of X.

Observe that $(\Omega, \mathcal{G})$ is a measurable space itself. Define the measure $\mu : \mathcal{G} \to [0, \infty)$ by the formula:

$$\mu(B) := \int_B X \, d\mathbb{P}, \quad B \in \mathcal{G}.$$

This is a measure by Theorem 11.5.

We have $\mu \ll \mathbb{P}|_{\mathcal{G}}$, where $\mathbb{P}|_{\mathcal{G}}$ denotes the restriction of $\mathbb{P}$ to the σ-algebra $\mathcal{G}$. Indeed, if $B \in \mathcal{G}$ and $\mathbb{P}[B] = 0$, we have $\mu(B) = \int_B X \, d\mathbb{P} = 0$. Also, $\mu(\Omega) = \int_\Omega X \, d\mathbb{P} < \infty$, since $X \in L^1(\mathbb{P})$. By the Radon-Nikodým Theorem (Theorem 11.7), there exists a function $Y : \Omega \to \mathbb{R}$, measurable on the space $(\Omega, \mathcal{G})$, such that $Y \in L^1(\mathbb{P}|_{\mathcal{G}})$ and $\mu(B) = \int_B Y \, d\mathbb{P}$. This proves statements (i)–(iii) of the theorem. If Y' is another $\mathcal{G}$-measurable RV satifying (i)–(iii), then we have $\int_B Y \, d\mathbb{P} = \int_B Y' \, d\mathbb{P}$ for all $B \in \mathcal{G}$, and hence $Y = Y'$ a.e.[$\mathbb{P}|_{\mathcal{G}}$] (see Exercise 44(b)). □

Remarks. Another way to write property (iii) in Theorem 15.5 is to say that for every $A \in \mathcal{G}$ we have $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$.

The **Tower property** of conditional expectations (also called the **law of total probability**) says that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X].$$

This is immediate from the definition above; taking $Y = \mathbb{E}[X|\mathcal{G}]$ and taking $B = \Omega$ in property (iii) of Theorem 15.5 gives $\mathbb{E}[Y] = \mathbb{E}[X]$.

Definition 15.7. In the special case where Z is a further RV on $(\Omega, \mathcal{F}, \mathbb{P})$, we write $\mathbb{E}[X|Z]$ for $\mathbb{E}[X|\sigma(Z)]$.

It can be shown that in this case, any $\sigma(Z)$-measurable random variable Y can be written as a function of Z: $Y = f(Z)$ for some Borel function $f : \mathbb{R} \to \mathbb{R}$. [Tricky exercise: prove this, starting with the case where Y is simple]

Therefore $\mathbb{E}[X|Z]$ can be characterized as an integrable variable of the form $Y = f(Z)$, with $f : \mathbb{R} \to \mathbb{R}$ a Borel function, satisfying (iii) above.

A Lebesgue but not Borel measurable sets

Theorem A.1. *There exists a set $B \subset \mathbb{R}$ such that $B \in \Sigma_1 \setminus \mathcal{B}_{\mathbb{R}}$.*

Non-examinable. We begin by recalling the construction of the middle third Cantor set $C \subset [0,1]$ from Section 3. C has a helpful characterisation in terms of ternary expansions of real numbers (that is, expansions in base 3). Observe that every $x \in [0,1]$ can be expanded as a sum

$$x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}, \qquad \text{where each } a_n \in \{0,1,2\}.$$

This expansion is unique unless $x = m3^{-k}$ for integers m and k (with m not divisible by 3), in which case we can either have $a_j = 0$ for all $j > k$ or $a_j = 2$ for all $j > k$. One of these expansions will have $a_k = 1$ and the other will have either $a_k = 0$ or $a_k = 2$. By convention, we will always choose the latter expansion. Under this convention, we see that $a_1 = 1$ if and only if $x \in (\frac{1}{3}, \frac{2}{3})$, if $a_1 \neq 1$, then $a_2 = 1$ if and only if $x \in (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$, etc. That is, the sets removed in the middle third construction are exactly those for which the ternary expansion of x contains a 1. We saw in Section 3 that the outer measure of C is zero, and hence we know that C is measurable and $\lambda^1(C) = 0$. We label the removed sets $I_j = (a_j, b_j)$ and recall $\lambda^1\big(\bigcup_{j=1}^{\infty} I_j\big) = 1$.

One more fact that will be helpful is to note that if $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$ and $y = \sum_{n=1}^{\infty} \frac{b_n}{3^n}$ then we have $x < y$ if and only if there exists n such that $a_j = b_j$ for $j < n$ and $a_n < b_n$.

With this notation in hand, we return to the construction. We define the Cantor function f as follows: For each $x \in C$ expanded as $\sum_{n=1}^{\infty} \frac{a_n}{3^n}$, we define

$$f(x) = \frac{1}{2} \sum_{n=1}^{\infty} \frac{a_n}{2^n}.$$

It should be apparent that $f(x)$ is the base 2 expansion of a number in $[0,1]$ (as $a_n \in \{0,2\}$, so that $\frac{1}{2}a_n \in \{0,1\}$) and every number is $[0,1]$ is obtained via this procedure, so that f is surjective from C to $[0,1]$. We extend f to a function $f : [0,1] \to [0,1]$ by observing that if $x, y \in C$ and $x < y$, then either $f(x) < f(y)$ or x and y are the two endpoints of one of the intervals I_j removed in the construction of C. In this case, we define $f = f(x) = f(y)$ on I_j. In this way, we arrive at a weakly increasing surjective function from $[0,1]$ to $[0,1]$, which must therefore be continuous (as it has no jump discontinuities). [NB: the Cantor function f is sometimes referred to as the Devil's staircase. f is constant on each interval I_j, the union of which is a set of full measure; but f is continuous and $f(1) - f(0) = 0$!]

We now define a new function, $g : [0,1] \to [0,2]$ as $g(x) = f(x) + x$. As g is a sum of continuous functions, it is continuous, while from the weak monotonicity of f, we see that if $x < y$, then $g(x) = f(x) + x < f(y) + y = g(y)$, so that g is strictly monotone. g is therefore a continuous, strictly increasing bijection with a continuous inverse, $h : [0,2] \to [0,1]$.

Claim: The measure $\lambda^1(g(C)) = 1$.
To prove the claim, we recall that $[0,1] \setminus C = \bigcup_{j=1}^{\infty} I_j$, where each I_j is an open interval, all the I_j are pairwise disjoint, and f is constant on each I_j. By additivity of λ^1 and bijectivity of g, we observe

$$2 = \lambda^1([0,2]) = \lambda^1(g([0,1])) = \lambda^1(g([0,1] \setminus C) \cup g(C)) = \lambda^1(g([0,1] \setminus C)) + \lambda^1(g(C)).$$

To prove the claim, we will therefore show that $\lambda^1(g([0,1] \setminus C)) = 1$. In fact, as g is a bijection, the sets $g(I_j)$ are pairwise disjoint, so that by σ-additivity,

$$\begin{aligned}
\lambda^1(g([0,1] \setminus C)) &= \lambda^1\big(g\big(\bigcup_{j=1}^{\infty} I_j\big)\big) = \sum_{j=1}^{\infty} \lambda^1(g(I_j)) \\
&= \sum_{j=1}^{\infty} \lambda^1\big((f(a_j) + a_j, f(b_j) + b_j)\big) && \text{where we write } I_j = (a_j, b_j) \\
&= \sum_{j=1}^{\infty} \big(f(b_j) + b_j - f(a_j) - a_j\big) \\
&= \sum_{j=1}^{\infty} (b_j - a_j) && \text{as } f \text{ is constant on } I_j \\
&= \sum_{j=1}^{\infty} \lambda^1(I_j) = \lambda^1\big(\bigcup_{j=1}^{\infty} I_j\big) = 1.
\end{aligned}$$

This proves the claim.

As the set $g(C)$ has measure 1, it must contain a non-measurable subset $A \subset g(C)$ (the proof of this is a variant of the construction of the Vitali set). Then the set $B = g^{-1}(A) \subset C$ has outer Lebesgue measure 0 (as $\mu^*(B) \leq \mu^*(C) = 0$) and is therefore Lebesgue measurable (with measure zero). But B cannot be Borel else we would have that $h^{-1}(B) = g(B) = A$ is also Borel due to the continuity of h. But as A is not measurable, it cannot be Borel. □

B π and λ systems and the Monotone Class Theorem

We first present the following result, known as the Monotone Class theorem, This should not be confused with the Monotone Convergence theorem (Theorem ??)! To further confuse matters, there are some other versions of the Monotone Class theorem in the literature, but we just consider this version here.

Theorem B.1 (Monotone Class theorem)**.** *Let $\mathcal{D}$ be a π-system in a non-empty set X with $X \in \mathcal{D}$. Let $\mathcal{H}$ be a collection of functions $X \to \mathbb{R}$, satisfying the following:*
(a) If $A \in \mathcal{D}$, then $\mathbb{1}_A \in \mathcal{H}$.
(b) If $f, g \in \mathcal{H}$ and $\alpha \in \mathbb{R}$, then also $f + g \in \mathcal{H}$ and $\alpha f \in \mathcal{H}$ (that is, $\mathcal{H}$ is a linear space);

(c) If $f_n \in \mathcal{H}$ for $n \in \mathbb{N}$, with $0 \le f_n \uparrow f$ pointwise, and the limit function f is bounded, then also $f \in \mathcal{H}$ (that is, $\mathcal{H}$ is closed under bounded monotone convergence of non-negative functions).
Then $\mathcal{H}$ contains all bounded measurable functions with respect to $\sigma(\mathcal{D})$.

Proof. *(Non-examinable)* First we prove the result for indicator functions. Set

$$\mathcal{L} := \{A \subset X : \mathbb{1}_A \in \mathcal{H}\}.$$

Then $\mathcal{D} \subset \mathcal{L}$ by assumption (a). We show that $\mathcal{L}$ is a λ-system in X:
• Since the constant function $1 = \mathbf{1}_X \in \mathcal{H}$, by (b) all constant functions are in $\mathcal{H}$. In particular the zero function is in $\mathcal{H}$ so $\varnothing \in \mathcal{H}$;
• If $B \in \mathcal{L}$, then $\mathbf{1}_{B^c} = 1 - \mathbf{1}_B \in \mathcal{H}$. Hence $B^c \in \mathcal{L}$.
• If $B_n \in \mathcal{L}$ and $\{B_n\}$ are pairwise disjoint, then setting $U_k = \cup_{n=1}^k B_n$ and $U = \cup_{n=1}^\infty B_n$, we have $\mathbf{1}_{U_k} = \sum_{n=1}^k \mathbf{1}_{B_n} \in \mathcal{H}$, and $\mathbb{1}_{U_k} \uparrow \mathbb{1}_U$, and by (c) we have $\mathbb{1}_U \in \mathcal{H}$. Hence $U \in \mathcal{L}$.
This verifies that $\mathcal{L}$ is a λ-system. By the π-λ theorem, we conclude that $\sigma(\mathcal{D}) \subset \mathcal{L}$.

Summarizing the previous paragraph: the indicator function of any element of $\sigma(\mathcal{D})$ is in $\mathcal{H}$. It follows using (b) that all simple $\sigma(\mathcal{D})$-measurable functions belong to $\mathcal{H}$.

Using (c) and approximation from below by simple functions (Theorem 8.10), we get that all non-negative bounded $\sigma(\mathcal{D})$-measurable functions belong to $\mathcal{H}$. Using $f = f^+ - f^-$, the statement follows for all bounded $\sigma(\mathcal{D})$-measurable functions.

□