MATH 303 — Measure Theory
Lecture Notes, Fall 2025

Ethan Ackelsberg

Contents

Part 1

Motivation and Basics of Abstract Measure Theory

CHAPTER 1

Motivating Problems of Measure Theory

Learning Objectives

At the end of this chapter, you will be able to:

- Compare and contrast different approaches to the "problem of measurement" in Euclidean space and identify the advantages and disadvantages of different methods
- Describe applications of measure theory to other areas of mathematics

1. The Problem of Measurement

A basic (and very old) problem in mathematics is to compute the size (length, area, volume) of geometric objects. In this chapter, we will trace the history (in a highly abbreviated form) of the mathematical developments related to the measurement of the size of geometric objects from ancient times up to the 20th century. The guide throughout will be the following open-ended questions:

- What are the "geometric objects" to which we want to (and are able to) assign a notion of size?
- What properties should size (length, area, volume) satisfy?
- How do we compute sizes of geometric objects?

This loosely-defined problem is what we will call the "problem of measurement" in Euclidean space.

2. Ancient Mathematics - Polygons, Polygons, Polygons!

In the Greek school of mathematics of antiquity[1], the computations of areas and volumes of regions was carried out by reducing general regions for which the area or volume was unknown to polygon or polyhedral regions for which the area or volume was easily computed. This consisted of two primary methods: *quadrature* (or *squaring*) and the *method of exhaustion*.

2.1. Quadrature. Quadrature (or squaring) is the process of constructing, from a given two-dimensional region, a square of equal area. This is easily carried out for simple regions such as rectangles (see Example 1.1), parallelograms, and triangles, but quickly becomes much more difficult for curved regions. The problem of "squaring the circle," i.e. carrying out this procedure for a circle in a finite number of steps using only straightedge and compass, stumped ancient mathematicians and for good reason: the fact that π is a transcendental number (proved by Lindemann in 1882) makes a solution to the problem impossible.

Example 1.1: Quadrature of a rectangle

Given a rectangle with sides a, b, we can square the rectangle as follows. Place segments of length a and b end to end and form a (semi)circle with diameter given by the two segments

[1] A very enlightening discussion of the history of "Greek mathematics" recently took place in the pages of the *Notices of the American Mathematical Society*; see [**Kim25**, **Net25**].

(of total length $a + b$), and draw a segment perpendicular to the diameter at the meeting point of the two segments (see Figure 1.1).

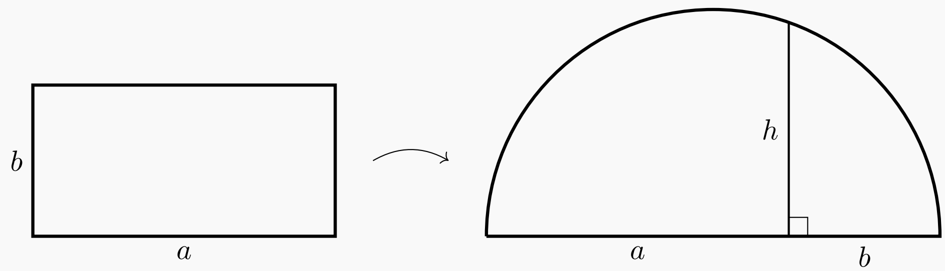


Figure 1.1.

To compute the height h, form two right triangles over bases a and b (see Figure 1.2). The two hypotenuses meet at a right angle by Thales's theorem. Then by three applications of the Pythagorean theorem,

$$\underbrace{(a+b)^2}_{a^2+b^2+2ab} = \underbrace{(a^2+h^2)+(b^2+h^2)}_{a^2+b^2+2h^2},$$

whence $h^2 = ab$. The square with base h shown in red thus has the same area as the original rectangle, so we have successfully squared the rectangle.

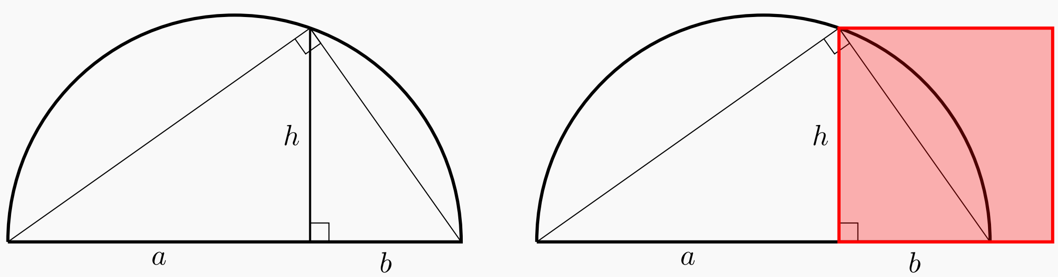


Figure 1.2.

2.2. Method of Exhaustion. Another method utilized in antiquity and containing the seeds of later developments in analysis was the method of exhaustion. Credited to Eudoxus for establishing the method rigorously, the method of exhaustion consists of inscribing and circumscribing sequences of polygons that converge to a given shape (see Figure 1.3). When used in conjunction with the method of squaring, which can be used to compute polygonal areas, the method of exhaustion was a powerful method for measurement in Greek mathematics. Polygonal approximations (rediscovered and improved in various locations and times) continued to be the best-known method for computing π until the end of the 17th century.

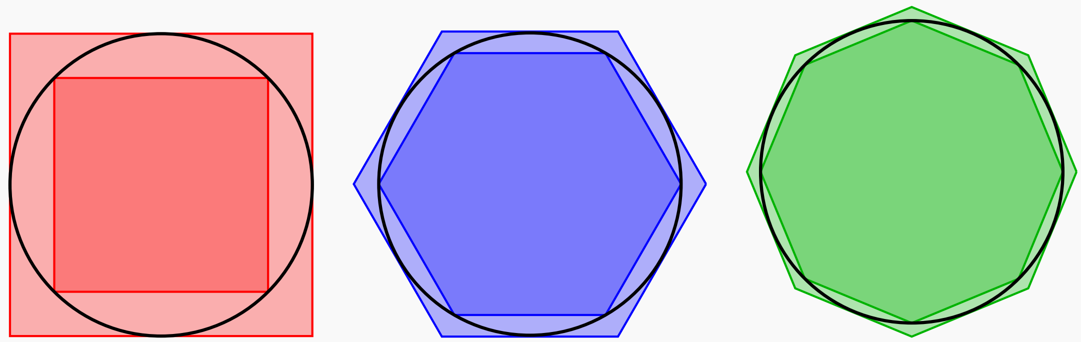


FIGURE 1.3. The method of exhaustion applied to a circle. Archimedes used this method (with regular 96-gons) to show $3 + \frac{10}{71} < \pi < 3 + \frac{1}{7}$.

3. Indivisibles and Infinitesimals

A significant breakthrough in the computation of areas and volumes was formalized in the 17th century by Bonaventura Cavalieri. (Similar methods had been used in antiquity by Archimedes and in the 5th century CE by Chinese mathematicians Zu Chongzhi and Zu Gengzhi, but the method was not in common usage in early modern Europe.) Cavalieri's principle can be expressed as follows:

> *Suppose two regions in the plane are bounded between two parallel lines. If the two regions have cross-sections of equal length, then they have equal area.*

A corresponding statement also holds in three dimensions: if two regions have planar cross-sections of equal area, then they have equal volume.

EXAMPLE 1.2

Cavalieri's principle can be used to compute the volume of a cone. First, by slicing parallel to the base, Cavalieri's principle shows that the volume of a pyramidal region depends only on the area of the base and the height. In particular, the computation of the volume of a cone (or any other starting pyramid) can be reduced to computing the volume of a square pyramid (Figure 1.4).

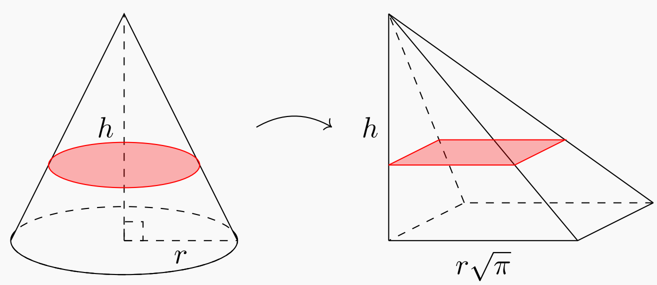


FIGURE 1.4. A demonstration of Cavalieri's principle. The two regions are built out of cross-section of equal area so have equal volume.

The volume of the resulting pyramid can be computed by observing that three pyramids for which the height is equal to the side length of the base can be combined into a cube (Figure 1.5).

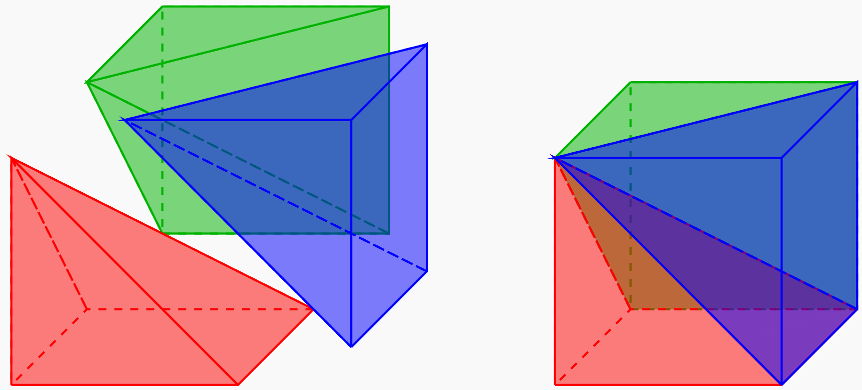


FIGURE 1.5. Three square pyramids combine to form a cube, so the volume of each pyramid is one third the volume of the cube.

Since the volume of the pyramid scales proportionally to its height, we conclude that the pyramid (and hence the cone) has volume $\frac{1}{3}\pi r^2 h$.

While the volume of a cone was known in antiquity using the method of exhaustion (it appeared, for example, in the 12th book of Euclid's *Elements*), Cavalieri's principle provides a much simpler proof. As a result of the negative resolution of Hilbert's third problem, it is also now known that the computation of the volume of a cone or even certain pyramids requires some form of infinitary argument (using Cavalieri's indivisibles, integral calculus, or a limiting process as in the method of exhaustion), since polyhedra of equal volume cannot always be transformed into each other via finitely many polyhedral cuts and rearrangements.

4. Integral Calculus

Notwithstanding earlier developments from Menaechmus and Apollonius in ancient Greek mathematics and Omar Khayyam in 11th century Persian mathematics, the introduction of coordinate systems by Descartes in the 17th century set forth the discipline of *analytic geometry* and revolutionized geometric calculations by uniting geometry with algebra. The variety of "geometric objects" was no longer limited to polygons, polyhedra, conic sections, and other classical objects; mathematicians had been unleashed to describe an endless assortment of new shapes by means of algebraic formulae. But how was one to compute their sizes?

Following earlier contributions by Cavalieri (who computed the area under $y = x^n$ for $n \leq 9$), Wallis (who extended Cavalieri's work to general $n \in \mathbb{Z}$), and many others, Newton and Leibniz discovered an astonishing link between the computation of areas (integration) and differentiation, namely the *fundamental theorem of calculus*.

5. Introducing ε and δ

Early work in calculus was based on infinitesimals and does not meet our present-day standards for mathematical rigor. Though a rigorous foundation for the theory of infinitesimals was eventually established by Abraham Robinson in the 1960s (dubbed "nonstandard analysis" as a result of its later historical development), calculus was first put on firm foundations by the "ε-δ" formalism established in the 19th century by Cauchy, Weierstrass, and others. Using the newly rigorous notions of limits, the ancient method of exhaustion could finally reach its full potential with the integration theory developed by Riemann and Darboux. For purposes of exposition, we will focus on Darboux's approach to integration, which is very similar to Riemann's but with some simplifications.

Definition 1.3: Darboux Integration

Let $B = \prod_{i=1}^{d} [a_i, b_i]$ be a closed box in $\mathbb{R}^d$, and let $f : B \to \mathbb{R}$ be a bounded function.

- A *Darboux partition* of B is a family of finite sequences $(x_{i,j})_{1 \le i \le d, 0 \le j \le n_i}$ such that $a_i = x_{i,0} < x_{i,1} < \cdots < x_{i,n_i} = b_i$ for each $i \in \{1, \ldots, d\}$.

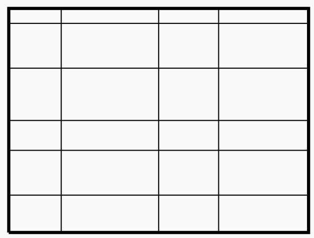


FIGURE 1.6. A Darboux partition in dimension $d = 2$ with $n_1 = 4$ and $n_2 = 6$.

- Given a Darboux partition $P = (x_{i,j})_{1 \le i \le d, 0 \le j \le n_i}$ of B, the *upper* and *lower Darboux sums of f over B* are given by

$$U_B(f, P) = \sum_{\mathbf{j} \in \prod_{i=1}^{d} \{1, \ldots, n_i\}} \sup_{\boldsymbol{x} \in B_{\mathbf{j}}} f(\boldsymbol{x}) \cdot \mathrm{Vol}(B_{\mathbf{j}})$$

and

$$L_B(f, P) = \sum_{\mathbf{j} \in \prod_{i=1}^{d} \{1, \ldots, n_i\}} \inf_{\boldsymbol{x} \in B_{\mathbf{j}}} f(\boldsymbol{x}) \cdot \mathrm{Vol}(B_{\mathbf{j}}),$$

where $B_{\mathbf{j}}$ is the box $\prod_{i=1}^{d} [x_{i,j_i - 1}, x_{i,j_i}]$, and $\mathrm{Vol}(B_{\mathbf{j}}) = \prod_{i=1}^{d} (x_{i,j_i} - x_{i,j_i - 1})$ is the volume of $B_{\mathbf{j}}$.

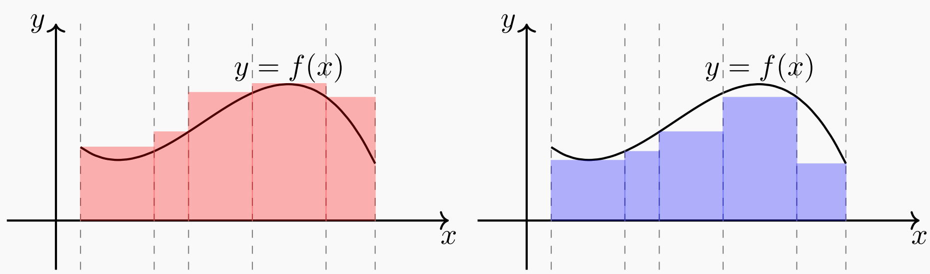


FIGURE 1.7. Upper (red) and lower (blue) Darboux sums of a function f over an interval ($d = 1$).

- The *upper* and *lower Darboux integral of f over B* are

$$U_B(f) = \inf\{U_B(f, P) : P \text{ is a Darboux partition of } B\}$$

and

$$L_B(f) = \sup\{L_B(f,P) : P \text{ is a Darboux partition of } B\}.$$

- The function f is *Darboux integrable over B* if $U_B(f) = L_B(f)$, and their common value is called the *Darboux integral of f over B* and is denoted by $\int_B f(\boldsymbol{x})\, d\boldsymbol{x}$.

Since Riemann integration is more commonly taught, we mention that the Darboux integral and the Rieman integral define the same quantity.

Proposition 1.4

A function f is Darboux integrable if and only if it is Riemann integrable. Moreover, the value of the Darboux integral and the Riemann integral (for a Riemann–Darboux integrable function) are the same.

Because of its flexibility in terms of the dimension of the ambient Euclidean space, the Riemann–Darboux integral comes with an attendant notion of size or "hyper-volume" for regions in Euclidean space: the *Jordan content*.

Definition 1.5

A bounded set $E \subseteq \mathbb{R}^d$ is a *Jordan measurable set* if $\mathbb{1}_E$ is Riemann–Darboux integrable over a box containing E. The *Jordan content* of a Jordan measurable set E is the value $J(E) = \int_B \mathbb{1}_E(\boldsymbol{x})\, d\boldsymbol{x}$, where B is any closed box containing E.

Jordan measurable sets include basic geometric objects such as polyhedra, conic sections, regions bounded by finitely many smooth curves/surfaces, etc. The basic building blocks for Jordan measurable sets are what are called *simple sets* (or *elementary sets*).

Definition 1.6

An *interval* in $\mathbb{R}$ is a set of the form (a,b), $[a,b)$, $(a,b]$, or $[a,b]$ for some real numbers $a \le b$. A *box* in $\mathbb{R}^d$ is a set of the form $B = \prod_{i=1}^d I_i$, where $I_1, \ldots, I_d$ are intervals. A set $S \subseteq \mathbb{R}^d$ is a *simple set* (or *elementary set*) if it is a finite union of boxes $S = \bigcup_{j=1}^k B_j$.

If the boxes $B_1, \ldots, B_k$ are disjoint, then the volume of the simple set $S = \bigcup_{j=1}^k B_j$ is $\mathrm{Vol}(S) = \sum_{j=1}^k \mathrm{Vol}(B_j)$. If some of the boxes intersect, then the volume of $S = \bigcup_{j=1}^k B_j$ can be computed using inclusion-exclusion:

$$\mathrm{Vol}(S) = \sum_{j=1}^k \mathrm{Vol}(B_j) - \sum_{1 \le j_1 < j_2 \le k} \mathrm{Vol}(B_{j_1} \cap B_{j_2}) + \sum_{1 \le j_1 < j_2 < j_3 \le k} \mathrm{Vol}(B_{j_1} \cap B_{j_2} \cap B_{j_3}) - \ldots$$

This expression is well-defined, since the intersection of two boxes is again a box. A Jordan measurable set is a set that is "well-approximated" by simple sets, as we will make precise now.

Definition 1.7

For a bounded set $E \subseteq \mathbb{R}^d$, define the *inner* and *outer Jordan content* by

$$J_*(E) = \sup\{\mathrm{Vol}(S) : S \subseteq E \text{ is a simple set}\}.$$

and

$$J^*(E) = \inf \{\mathrm{Vol}(S) : S \supseteq E \text{ is a simple set}\}.$$

The inner Jordan content can be viewed as a generalization of the method of approximation by inscribed polygons and the outer Jordan content as a generalization of the method of approximation by circumscribed polygons. In order to make sense of the size of an object using the Jordan content, the inscribed and circumscribed regions must approach the same size. In other words, Jordan measurable sets are those for which this extended method of exhaustion successfully converges. This is made precise by the following theorem.

Theorem 1.8

Let $E \subseteq \mathbb{R}^d$ be a bounded set. The following are equivalent:

(i) E is Jordan measurable;
(ii) $J_*(E) = J^*(E)$ (in which case $J(E)$ is equal to this same value);
(iii) $J^*(\partial E) = 0$.

We do not include a proof of Theorem 1.8 but indicate its core content in Figure 1.8.

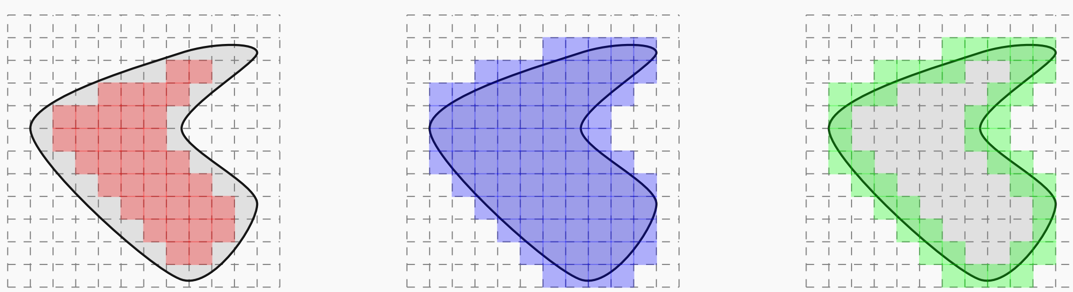


Figure 1.8. Simple sets approximating the inner (red) and outer Jordan content (blue) of a region in dimension $d = 2$. With the red boxes removed from the blue, we get a simple set covering the boundary (in green).

While the family of Jordan measurable sets is quite vast, a consequence of Theorem 1.8 is that not all sets are Jordan measurable.

Example 1.9

The sets $\mathbb{Q} \cap [0, 1]$ and $[0, 1] \setminus \mathbb{Q}$ are not Jordan measurable.

In addition to the above example, there are many other "nice" sets that are not Jordan measurable. There are, for instance, bounded open sets in $\mathbb{R}$ that are not Jordan measurable. We will work out one such example in detail.

Example 1.10

The complement U of the fat Cantor set (also known as the Smith–Volterra–Cantor set) $K \subseteq [0, 1]$ is Jordan non-measurable. We construct K iteratively, starting from $[0, 1]$, by removing intervals of length 4^{-n} at step n. In other words, at step n, we remove an interval of length 4^{-n} around each rational point with denominator 2^n.

Figure 1.9. Iterative construction of the fat Cantor set.

Let

$$U = \bigcup_{n=0}^{\infty} \bigcup_{j=1}^{2^n} \left(\frac{2j+1}{2^{n+1}} - \frac{1}{2 \cdot 4^{n+1}}, \frac{2j+1}{2^{n+1}} + \frac{1}{2 \cdot 4^{n+1}} \right).$$

Then $K = [0,1] \setminus U$. The inner Jordan content of U is

$$J_*(U) = \sum_{n=0}^{\infty} \sum_{j=1}^{2^n} \operatorname{Len} \left(\frac{2j+1}{2^{n+1}} - \frac{1}{2 \cdot 4^{n+1}}, \frac{2j+1}{2^{n+1}} + \frac{1}{2 \cdot 4^{n+1}} \right) = \sum_{n=0}^{\infty} 2^n \cdot \frac{1}{4^{n+1}} = \frac{1}{4} \sum_{n=0}^{\infty} 2^{-n} = \frac{1}{2}.$$

However, $\overline{U} = [0,1]$ (since U contains every rational number whose denominator is a power of 2), so the outer Jordan content of U is $J^*(U) = J^*([0,1]) = 1$.

6. Set Theory, Choice, and the Impossibility of Measuring Everything

As mathematicians continued the quest for formalization, the notion of "geometric object" continued to expand in possible meaning. With set theory taking its place at the foundations of mathematics, one could now dream of perhaps assigning a size to arbitrary subsets of Euclidean space. Giuseppe Vitali dashed such hopes with a clever construction in 1905.

Thus far, our discussion of the notion of size has been largely based on geometric intuition. In order to say that there exists a set incapable of being assigned a sensible notion of size, we now take an axiomatic approach and reformulate (a version of) the problem of measurement as a concrete mathematical statement.

Problem 1.11: The Problem of Measurement (Strong Form)

Let $d \in \mathbb{N}$. Does there exists a notion of d-dimensional volume Vol, defined for all subsets of $\mathbb{R}^d$, such that

- normalized: $\operatorname{Vol}([0,1)^d) = 1$;
- isometry-invariant: if A and B are isometric, then $\operatorname{Vol}(A) = \operatorname{Vol}(B)$;
- countably additive: if $E_1, E_2, \ldots$ are pairwise disjoint, then $\operatorname{Vol}\left(\bigsqcup_{n \in \mathbb{N}} E_n\right) = \sum_{n=1}^{\infty} \operatorname{Vol}(E_n)$.

Vitali showed that Problem 1.13 has a negative answer even for $d = 1$.

Theorem 1.12

There is no normalized, translation-invariant, countably additive function defined for all subsets of $\mathbb{R}$.

Proof. Define an equivalence relation on $[0,1)$ by $x \sim y$ if $y - x \in \mathbb{Q}$. By the axiom of choice, let $E \subseteq [0,1)$ be a set containing exactly one representative of each equivalence class. For each $t \in \mathbb{Q} \cap [0,1)$, let $E_t = \{x + t \bmod 1 : x \in E\} \subseteq [0,1)$.

Claim 1. The sets $(E_t)_{t \in \mathbb{Q} \cap [0,1)}$ are pairwise disjoint.

For $t, s \in \mathbb{Q} \cap [0,1)$ and $x, y \in E$, if $x + t \equiv y + s \pmod 1$, then

$$y - x \equiv t - s \pmod 1,$$

so $x \sim y$. But E contains only one element from each equivalence class, so $x = y$ and $t = s$.

Claim 2. $\bigsqcup_{t \in \mathbb{Q} \cap [0,1)} E_t = [0,1)$

Let $x \in [0,1)$. Then there exists $y \in E$ with $y \sim x$, since E has a representative of each equivalence class. Let $t = x - y \bmod 1 \in \mathbb{Q} \cap [0,1)$. Then

$$y + t \equiv x \pmod 1.$$

so $x \in E_t$.

Suppose for contradiction that L is a normalized, translation-invariant, countably additive function defined for all subsets of $\mathbb{R}$.

Claim 3. For every $t \in \mathbb{Q} \cap [0,1)$, $L(E_t) = L(E)$.

We can write

$$E_t = ((E + t) \cap [0,1)) \sqcup ((E + t) \cap [1,2) - 1).$$

Therefore, by translation-invariance,

$$L(E_t) = L(E + t) = L(E).$$

Combining Claims 1–3 and using countable additivity of L,

$$1 = L([0,1)) = \sum_{t \in \mathbb{Q} \cap [0,1)} L(E_t) = \sum_{t \in \mathbb{Q} \cap [0,1)} L(E) = \infty \cdot L(E).$$

But there is no value of $L(E)$ that can satisfy this equation, so we have reached a contradiction. □

Confronted with Vitali's example, one must make some compromise. In order to comport with an intuitive meaning of "size," normalization and isometry-invariance appear absolutely essential. This leaves two options: (1) restrict the domain of the volume function to only assign size to a certain subclass of "nice" sets and hope to avoid the pathologies of the Vitali sets, or (2) sacrifice countable additivity for the weaker notion of finite additivity. We address the two possibilities in turn, starting with the latter. Relaxing our additivity assumption to *finite additivity*, we arrive at a new form of the problem of measurement.

Problem 1.13: The Problem of Measurement (Weak Form)

Let $d \in \mathbb{N}$. Does there exists a notion of d-dimensional volume Vol, defined for all subsets of $\mathbb{R}^d$, such that

- NORMALIZED: $\mathrm{Vol}([0,1)^d) = 1$;
- ISOMETRY-INVARIANT: if A and B are isometric, then $\mathrm{Vol}(A) = \mathrm{Vol}(B)$;
- FINITELY ADDITIVE: if A and B are disjoint, then $\mathrm{Vol}(A \sqcup B) = \mathrm{Vol}(A) + \mathrm{Vol}(B)$.

Surprisingly, the solvability of this weak form of the problem of measurement depends on the dimension d. The problem was solved by Banach in dimensions $d = 1$ and $d = 2$ using a version of the Hahn–Banach theorem from functional analysis. However, in dimensions 3 and higher, a paradoxical situation emerges.

Theorem 1.14: Banach–Tarski Theorem

Let $d \geq 3$. Given any two bounded regions $A, B \subseteq \mathbb{R}^d$, both with nonempty interior, there exist partitions $A = A_1 \sqcup \ldots \sqcup A_k$ and $B = B_1 \sqcup \ldots B_k$ for some $k \in \mathbb{N}$ such that A_i and B_i are congruent for each $i \in \{1, \ldots, k\}$. In particular, the unit ball can be decomposed into finitely many pieces and reassembled into two congruent copies of the unit ball.

Thus, at least in high dimensional situations, even with weaken the notion of "size" to only be finitely additive, there is no consistent way to measure every subset of $\mathbb{R}^d$. There is also good reason to insist on the property of countable additivity. For example, finitely-additive notions of measure produce integrals that do not interact with limits in the way that one might hope. For ease of exposition, we give an example with the Riemann integral, but similar examples can be constructed for any notion of integration that fails to be countably additive (including Banach's notion of integration in dimensions 1 and 2).

Example 1.15

Enumerate the set $\mathbb{Q} \cap [0,1] = \{q_1, q_2, \ldots\}$. Let $f_n : [0,1] \to [0,1]$ be the function

$$f_n(x) = \begin{cases} 1, & \text{if } x \in \{q_1, \ldots, q_n\} \\ 0, & \text{otherwise.} \end{cases}$$

Then f_n is Riemann integrable and $f_n \to \mathbb{1}_{\mathbb{Q}\cap[0,1]}$ pointwise, but $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$ is not Riemann integrable.

Gathering all of our observations thus far, the best we can hope for in addressing the problem of measurement is to exhibit a rich class of "measurable sets" (including, for example, polygons, circles, open sets[2], and general Jordan-measurable sets) for which we can define a normalized, translation-invariant, and countably additive notion of measure.

7. The Solution of Lebesgue

The Jordan non-measurable set in Example 1.10 appears to have a sensible notion of "length." Indeed, the complement U, being a disjoint union of intervals, could be reasonably assigned as a "length" the sum of the lengths of the (countably many) intervals of which it is made. This produces a value of $\frac{1}{2}$ for the length of U, and so we should take K to also have length $\frac{1}{2}$, since $K \sqcup U = [0,1]$ is an interval of length 1. The feature that U is a disjoint union of intervals turns out to not be any special feature of U at all but instead a general feature of open sets in $\mathbb{R}$.

[2] Example 1.10 shows that there are open sets that are not Jordan-measurable, so we need a more general construction to handle arbitrary open sets.

Proposition 1.16

Let $U \subseteq \mathbb{R}$ be an open set. Then U can be expressed as a countable disjoint union of open intervals.

By Proposition 1.16, it seems reasonable to define the length of an open set $U \subseteq \mathbb{R}$ as follows. Write $U = (a_1, b_1) \sqcup (a_2, b_2) \sqcup \dots$ as a disjoint union of open intervals, and define its length as $(b_1 - a_1) + (b_2 - a_2) + \dots$. Then open sets may play the role that simple sets played in the definition of the Jordan content, and this leads to the Lebesgue measure.

Remark. In higher dimensions, Proposition 1.16 needs to be modified, but one can still reasonably talk about the d-dimensional volume of open sets in $\mathbb{R}^d$. The key is to replace open intervals with half-open boxes $\prod_{i=1}^{d} [a_i, b_i)$.

Definition 1.17

Let $E \subseteq \mathbb{R}^d$.

- The *outer Lebesgue measure of E* is the quantity
$$\begin{aligned}\lambda^*(E) &= \inf \{\mathrm{Vol}(U) : U \supseteq E \text{ is open}\} \\ &= \inf \left\{ \sum_{j=1}^{\infty} \mathrm{Vol}(B_j) : B_1, B_2, \dots \text{ are boxes, and } E \subseteq \bigcup_{j=1}^{\infty} B_j \right\}.\end{aligned}$$
- The set E is *Lebesgue measurable* (with *Lebesgue measure* $\lambda(E) = \lambda^*(E)$) if for every $\varepsilon > 0$, there exists an open set $U \subseteq \mathbb{R}^d$ such that $E \subseteq U$ and $\lambda^*(U \setminus E) < \varepsilon$.

Proposition 1.18

If $E \subseteq \mathbb{R}^d$ is Jordan measurable, then E is Lebesgue measurable and $J(E) = \lambda(E)$.

The family of Lebesgue measurable sets is much larger than the family of Jordan measurable sets. Among the several nice properties of the Lebesgue measure (and abstract measures) that we will see later in the course are:

Proposition 1.19

(1) If $(E_n)_{n \in \mathbb{N}}$ are Lebesgue measurable sets, then $\bigcup_{n=1}^{\infty} E_n$ and $\bigcap_{n=1}^{\infty} E_n$ are Lebesgue measurable.
(2) If $(E_n)_{n \in \mathbb{N}}$ are pairwise disjoint and Lebesgue measurable, then $\lambda\left(\bigsqcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \lambda(E_n)$.
(3) if A and B are congruent, then $\lambda(A) = \lambda(B)$.
(4) If $E_1 \subseteq E_2 \subseteq \dots \subseteq \mathbb{R}^d$ are Lebesgue measurable sets, then $\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \lambda(E_n)$.
(5) If $E_1 \supseteq E_2 \supseteq \dots$ are Lebesgue measurable subsets of $\mathbb{R}^d$ and $\lambda(E_1) < \infty$, then $\lambda\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \lambda(E_n)$.

The "continuity" properties expressed in items (4) and (5) result in a corresponding notion of integration that is able to interact desirably with pointwise limits, overcoming the shortcomings of the Riemann integral illustrated in Example 1.15.

8. Applications of Abstract Measure Theory

The mathematical language and tools encompassed in measure theory play a foundational role in many other areas of mathematics. A highly abbreviated sampling follows.

Probability Theory. Measure theory provides the axiomatic foundations of probability theory, providing rigorous notions of *random variables* and *probabilities* of *events*. Important limit laws (the law of large numbers and central limit theorem, for example) are phrased mathematically using measure-theoretic notions of convergence.

Fourier Analysis. Periodic (say, continuous or Riemann-integrable) functions on the real line have corresponding Fourier series representations $f(x) \sim \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}$. The functions $e^{2\pi inx}$ are orthonormal, and Parseval's identity gives $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2 = \int_0^1 |f(x)|^2 \, dx$. Given a sequence $(a_n)_{n\in\mathbb{N}}$, one may ask whether $\sum_{n\in\mathbb{Z}} a_n e^{2\pi inx}$ is the Fourier expansion of some function f, and if so, what properties does f have? Another natural question is whether the series $\sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}$ actually converges to the function f, and if so, in which sense? Both of these questions are properly answered in a measure-theoretic framework. If one is interested in decomposing functions defined on other groups (for instance, on compact abelian groups) into their Fourier series, then one also needs to develop a method of integrating functions on groups in order to compute Fourier coefficients and make sense of Parseval's identity.

Functional Analysis and Operator Theory. When one studies familiar concepts from linear algebra in infinite-dimensional spaces, measures become unavoidable for many tasks. For example, versions of the spectral theorem (generalizing the representation of suitable matrices in terms of their eigenvalues and eigenvectors) for operators on infinite-dimensional spaces require the abstract notion of a measure.

Ergodic Theory. Ergodic theory was developed to study the long-term statistical behavior of dynamical (time-dependent) systems, providing a framework to resolve important problems in physics related to the "ergodic hypothesis" in thermodynamics and the "stability" of the solar system. It turns out that the appropriate mathematical formalism for understanding these problems comes from abstract measure theory.

Fractal Geometry. Self-similar geometric objects such as the Koch snowflake, Sierpiński carpet, and the middle-thirds Cantor set (see Figure 1.10) can be meaningfully assigned a notion of "dimension" that can take a non-integer value. How does one determine the dimension of a fractal object? There are several different approaches to dimension, but one of the most popular is the *Hausdorff dimension*, which relies on a family of measures that interpolate between the integer-dimensional Lebesgue measures.

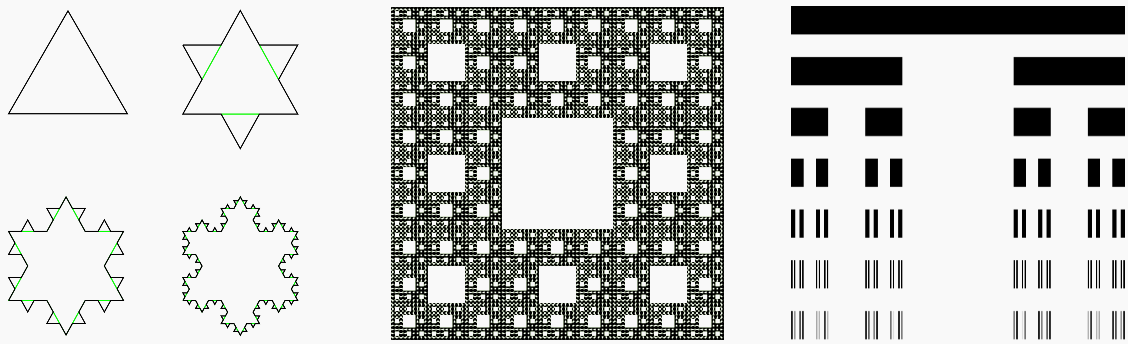


FIGURE 1.10. Fractal shapes: the Koch snowflake (left) of Hausdorff dimension $\frac{\log 4}{\log 3} \approx 1.26$, Sierpiński carpet (middle) of dimension $\frac{\log 8}{\log 3} \approx 1.89$, and middle-thirds Cantor set (right) of dimension $\frac{\log 2}{\log 3} \approx 0.63$.

CHAPTER 2

Measure Spaces

Learning Objectives

At the end of this chapter, you will be able to:

- Define the fundamental objects in measure theory (measurable sets, measurable functions, and measures)
- Identify and utilize tools for proving measurability of functions
- Prove basic properties of measures

1. σ-Algebras

Before defining measures, we must determine which subsets of a given set X we would like to be able to measure. Of course, in the best case scenario, we may hope to measure every subset of X. However, as demonstrated by Theorem 1.12, attempting to measure every set is often incompatible with other desirable properties for a measure. Thus, instead of insisting that a measure be defined for arbitrary subsets, we will be satisfied with having a sufficiently rich class of "measurable" subsets. What properties should we impose on this class? Certainly, we want the full space X to be measurable, and we should allow ourselves to perform the basic set-theoretic operations (complements, unions, and intersections). Allowing *finite* unions and intersections leads to the concept of an *algebra* of sets. Algebras are a very useful notion, but (as we saw with the Jordan content in the previous chapter) they are insufficient for appropriately handling limits. We will therefore upgrade from algebras to *σ-algebras*:

Definition 2.1

Let X be a set. A *σ-algebra* on X is a family $\mathcal{B} \subseteq \mathscr{P}(X)$ of subsets of X with the following properties:

- $X \in \mathcal{B}$;
- If $B \in \mathcal{B}$, then $X \setminus B \in \mathcal{B}$;
- If $(B_n)_{n\in\mathbb{N}}$ is a countable family of elements of $\mathcal{B}$, then $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{B}$.

Remark. In the definition of a σ-algebra, we have made no explicit mention of intersections. However, by De Morgan's laws, we can also generate countable intersection of sets:

$$\bigcap_{n\in\mathbb{N}} B_n = X \setminus \left(\bigcup_{n\in\mathbb{N}} (X \setminus B_n) \right).$$

Example 2.2

Some examples of σ-algebras include the following:

- For any set X, the power set $\mathscr{P}(X)$ is a σ-algebra, as is the pair $\{\emptyset, X\}$.

- The family $\mathcal{B} = \{B \subseteq \mathbb{R} : \text{either } B \text{ or } \mathbb{R} \setminus B \text{ is countable}\}$ of countable and co-countable subsets of $\mathbb{R}$ is a σ-algebra.
- Unions of unit-length intervals in $\mathbb{R}$ form a σ-algebra $\mathcal{B} = \left\{\bigcup_{n \in S}[n, n+1) : S \subseteq \mathbb{Z}\right\}$.

The basic object of study in abstract measure theory is a *measurable space*, which is a set for which we have designated a σ-algebra of measurable sets. More formally, we have the following definition.

Definition 2.3

A *measurable space* is a pair $(X, \mathcal{B})$, where X is a set and $\mathcal{B}$ is a σ-algebra on X. Elements of the σ-algebra $\mathcal{B}$ are called *measurable sets*.

In order to produce a wider variety of examples of σ-algebras than what appears in Example 2.2, it is helpful to have a general construction for producing a σ-algebra from a given family of sets. For example, in a topological space, it is natural to insist that all open sets be measurable. But then we need a method for producing a σ-algebra that contains all of the open sets (and is not the power set $\mathscr{P}(X)$, as this may contain pathological examples like the Vitali sets that make it impossible to define interesting measures). The following property of σ-algebras enables the desired general constructing of σ-algebras.

Proposition 2.4

Suppose $(\mathcal{B}_i)_{i \in I}$ is a family of σ-algebras on X. Then $\bigcap_{i \in I} \mathcal{B}_i$ is a σ-algebra.

Proof. Let $\mathcal{B} = \bigcap_{i \in I} \mathcal{B}_i$.

For every $i \in I$, we have $X \in \mathcal{B}_i$, so $X \in \mathcal{B}$.

Suppose $B \in \mathcal{B}$. Then $B \in \mathcal{B}_i$ for every $i \in I$, so $X \setminus B \in \mathcal{B}_i$ for every $i \in I$. Hence, $X \setminus B \in \mathcal{B}$.

Let $(B_n)_{n \in \mathbb{N}}$ be a countable family of sets in $\mathcal{B}$. For each $i \in I$, the sets $(B_n)_{n \in \mathbb{N}}$ belong to $\mathcal{B}_i$, so $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{B}_i$. Therefore, $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{B}$. □

Definition 2.5

The *σ-algebra generated by a family* $\mathcal{S} \subseteq \mathscr{P}(X)$ is the smallest σ-algebra containing $\mathcal{S}$, denoted by $\sigma(\mathcal{S})$.

Remark. Note that $\sigma(\mathcal{S})$ is well-defined by Proposition 2.4:

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

In topological spaces (such as the real line), we will often consider the σ-algebra generated by the topology.

Definition 2.6

Let (X, τ) be a topological space. The *Borel σ-algebra* is the σ-algebra generated by the open subsets of X, i.e. $\text{Borel}(X) = \sigma(\tau)$.

Borel sets can be placed in a hierarchy in terms of their level of complexity. At the simplest level are the open (G) and closed (F) sets. Next come countable intersections of open sets (G_δ sets) and countable unions of closed sets (F_σ sets) and so on.

$$\begin{array}{ccccccccccccc}
\Sigma^0_1 = G & \xrightarrow{\bigcap} & \Pi^0_2 = G_\delta & \xrightarrow{\bigcup} & \Sigma^0_3 = G_{\delta\sigma} & \xrightarrow{\bigcap} & \Pi^0_4 = G_{\delta\sigma\delta} & \xrightarrow{\bigcup} & \dots & \xrightarrow{\bigcup} & \Sigma^0_\omega & \xrightarrow{\bigcap} & \dots \\
\updownarrow c & & \updownarrow c & & \updownarrow c & & \updownarrow c & & & & \updownarrow c & & \\
\Pi^0_1 = F & \xrightarrow{\bigcup} & \Sigma^0_2 = F_\sigma & \xrightarrow{\bigcap} & \Pi^0_3 = F_{\sigma\delta} & \xrightarrow{\bigcup} & \Sigma^0_4 = F_{\sigma\delta\sigma} & \xrightarrow{\bigcap} & \dots & \xrightarrow{\bigcap} & \Pi^0_\omega & \xrightarrow{\bigcup} & \dots
\end{array}$$

FIGURE 2.1. The Borel hierarchy for subsets of a topological space.

The placement of a (Borel) set within the Borel hierarchy is a useful notion of "complexity" for sets. Intuitively speaking, if a set is lower down in the Borel hierarchy, then it is in some sense easier to define than a set higher up the hierarchy. Determining where sets occur in the Borel hierarchy (or if they are Borel at all) is a common theme in an area of mathematical logic known as *descriptive set theory*. We will largely not concern ourselves with such problems in this course, but some suggested additional reading appears at the end of this chapter for those who are interested.

2. Measurable Functions

Recall that a function $f : X \to Y$ from one topological space to another is continuous if the preimage of every open set in Y is open in X. Measurable functions are defined analogously, but with "open" replaced by "measurable."

DEFINITION 2.7

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

Some basic properties of measurable functions that will be used frequently are as follows:

PROPOSITION 2.8

(1) Let $(X, \mathcal{B})$, $(Y, \mathcal{C})$, and $(Z, \mathcal{D})$ be measurable spaces. Let $f : X \to Y$ and $g : Y \to Z$ be measurable functions. Then $g \circ f : X \to Z$ is measurable.
(2) Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces, and let $f : X \to Y$. Suppose $\mathcal{S} \subseteq \mathscr{P}(Y)$ is a family of sets such that $\sigma(\mathcal{S}) = \mathcal{C}$. If $f^{-1}(S) \in \mathcal{B}$ for every $S \in \mathcal{S}$, then f is a measurable function.
(3) Suppose X and Y are topological spaces and $\mathcal{B} = \mathrm{Borel}(X)$ and $\mathcal{C} = \mathrm{Borel}(Y)$ are the Borel σ-algebras on X and Y respectively. Then every continuous function $f : X \to Y$ is measurable.

PROOF. **(1)** Let $D \in \mathcal{D}$. Since g is measurable, we have $C = g^{-1}(D) \in \mathcal{C}$. Then since f is measurable, $B = f^{-1}(C) \in \mathcal{B}$. But $B = f^{-1}(g^{-1}(D)) = (g \circ f)^{-1}(D)$, so $g \circ f$ is measurable.

(2) Let $\mathcal{F} = \{E \subseteq Y : f^{-1}(E) \in \mathcal{B}\}$. We claim that $\mathcal{F}$ is a σ-algebra. Then since $\mathcal{S} \subseteq \mathcal{F}$, we conclude that $\mathcal{C} = \sigma(\mathcal{S}) \subseteq \mathcal{F}$, so f is measurable. Let us now prove the claim:

- $f^{-1}(Y) = X \in \mathcal{B}$, so $Y \in \mathcal{F}$.

- Suppose $E \in \mathcal{F}$. Then $f^{-1}(Y \setminus E) = X \setminus \underbrace{f^{-1}(E)}_{\in \mathcal{B}} \in \mathcal{B}$, so $Y \setminus E \in \mathcal{F}$.
- Suppose $E_1, E_2, \dots \in \mathcal{F}$, and let $E = \bigcup_{n\in\mathbb{N}} E_n$. Then
$$f^{-1}(E) = \bigcup_{n\in\mathbb{N}} \underbrace{f^{-1}(E_n)}_{\in \mathcal{B}} \in \mathcal{B},$$
so $E \in \mathcal{F}$.

This proves that $\mathcal{F}$ is a σ-algebra on Y.

(3) This follows from (1) by taking $\mathcal{S}$ to be the collection of open sets in Y. □

3. The Extended Real Numbers and Extended Real-Valued Functions

One obtains an important class of measurable functions when one considers functions defined on a measurable space taking real values. For many applications and in order to account more fully for limits of functions, it is often convenient to work with the slightly more general concept of *extended* real-valued functions.

Definition 2.9

The *extended real numbers* are the set $[-\infty, \infty] = \mathbb{R} \cup \{\infty, -\infty\}$ with the following topological and algebraic properties:

- The topology on $[-\infty, \infty]$ is generated by open intervals (a, b) with $a, b \in \mathbb{R}$ and sets of the form $(a, \infty] = (a, \infty) \cup \{\infty\}$ and $[-\infty, b) = (-\infty, b) \cup \{-\infty\}$ for $a, b \in \mathbb{R}$.
- Addition is extended as a commutative operation with $\infty + x = \infty$ and $-\infty + x = -\infty$ for real numbers $x \in \mathbb{R}$. For addition of two infinite quantities, we define $\infty + \infty = \infty$ and $-\infty + (-\infty) = -\infty$. However, $-\infty + \infty$ is undefined.
- Multiplication is also extended as a commutative operation with the properties
$$x \in (0, \infty) \implies \infty \cdot x = \infty \quad \text{and} \quad -\infty \cdot x = -\infty;$$
$$x \in (-\infty, 0) \implies \infty \cdot x = -\infty \quad \text{and} \quad -\infty \cdot x = \infty.$$
By convention, we define $\infty \cdot 0 = -\infty \cdot 0 = 0$. Multiplication of infinities is defined by $\infty \cdot \infty = (-\infty) \cdot (-\infty) = \infty$, and $-\infty \cdot \infty = -\infty$.

The topology we have defined on $[-\infty, \infty]$ is the *two-point compactification* of $\mathbb{R}$. You will check in the exercises that $[-\infty, \infty]$ is indeed a compact space (that is homeomorphic to a closed interval, say $[0, 1]$). The algebraic operations on $[-\infty, \infty]$ are all as one would expect, with one exception: $\infty \cdot 0$ is often considered as an "indeterminate form", but here we have given it a definite value of 0. The reason for this convention is the following proposition, which you will also prove in the exercises:

Proposition 2.10

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $[-\infty, \infty]$, and let $c \in \mathbb{R}$. If $(x_n)_{n\in\mathbb{N}}$ converges to an extended real number, then the sequence $(cx_n)_{n\in\mathbb{N}}$ also converges, and
$$\lim_{n\to\infty} (cx_n) = c \cdot \lim_{n\to\infty} x_n. \tag{2.1}$$

Proof. Exercise. □

In order to have the desirable property (2.1), one has no choice but to define $\infty \cdot 0 = 0$: by taking the sequence $x_n = n$, we have

$$0 \cdot \infty = 0 \cdot \lim_{n\to\infty} n = \lim_{n\to\infty} (0 \cdot n) = 0.$$

Warning: Property (2.1) does not hold for $c \in \{\infty, -\infty\}$, as can be seen by taking a sequence $(x_n)_{n\in\mathbb{N}}$ that converges to 0.

We say that an extended real-valued function $f : X \to [-\infty, \infty]$ defined on a measurable space $(X, \mathcal{B})$ is *$\mathcal{B}$-measurable* (or simply *measurable*) if it is measurable as a function between the measurable spaces $(X, \mathcal{B})$ and $([-\infty, \infty], \mathrm{Borel}([-\infty, \infty]))$. Since we will always take the same σ-algebra on $[-\infty, \infty]$, we omit explicit reference to the Borel σ-algebra when discussing measurable extended real-valued functions.

Proposition 2.11

Let $(X, \mathcal{B})$ be a measurable space.

(1) Let $f : X \to [-\infty, \infty]$. The following are equivalent:
  - (a) f is measurable;
  - (b) for every $c \in \mathbb{R}$, $f^{-1}((c, \infty]) \in \mathcal{B}$;
  - (c) for every $c \in \mathbb{R}$, $f^{-1}([c, \infty]) \in \mathcal{B}$;
  - (d) for every $c \in \mathbb{R}$, $f^{-1}([-\infty, c)) \in \mathcal{B}$;
  - (e) for every $c \in \mathbb{R}$, $f^{-1}([-\infty, c]) \in \mathcal{B}$.

(2) Suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions from X to $[-\infty, \infty]$. The following functions are also measurable:
  - (a) $\sup_{n\in\mathbb{N}} f_n$;
  - (b) $\inf_{n\in\mathbb{N}} f_n$;
  - (c) $\limsup_{n\to\infty} f_n$;
  - (d) $\liminf_{n\to\infty} f_n$.

(3) Suppose $f, g : X \to \mathbb{R}$ are measurable functions. Then $f + g$ and $f \cdot g$ are measurable.

Notation. For convenience, we will often write sets of the form $f^{-1}((c, \infty])$ as $\{f > c\}$ and similarly for $\{f \geq c\}$, $\{f < c\}$, and $\{f \leq c\}$.

Proof of Proposition 2.11. **(1)** By Proposition 2.8(2), it suffices to check that each of the relevant collections of intervals generates the Borel σ-algebra on $[-\infty, \infty]$. Let us show that the collection of intervals $(c, \infty]$ for $c \in \mathbb{R}$ generates the Borel σ-algebra. All of the other proofs are similar, so we omit them.

Let $\mathcal{S} = \{(c, \infty] : c \in \mathbb{R}\}$. Note that every element of $\mathcal{S}$ is open in $[-\infty, \infty]$, so $\sigma(\mathcal{S}) \subseteq \mathrm{Borel}([-\infty, \infty])$. On the other hand, we can write $(a, b] = (a, \infty] \setminus (b, \infty]$ for $a, b \in \mathbb{R}, a < b$. Every open set in $\mathbb{R}$ is a countable (disjoint) union of such intervals, so every open subset of $\mathbb{R}$ is contained in $\sigma(\mathcal{S})$. We obtain the additional open sets in $[-\infty, \infty]$ from the rays $(c, \infty] \in \mathcal{S}$ and

$$[-\infty, c) = \bigcup_{n\in\mathbb{N}} \left[-\infty, c - \frac{1}{n}\right] = \bigcup_{n\in\mathbb{N}} \left([-\infty, \infty] \setminus \left(c - \frac{1}{n}, \infty\right]\right) \in \sigma(\mathcal{S}).$$

Thus, $\mathrm{Borel}([-\infty, \infty]) \subseteq \sigma(\mathcal{S})$.

(2) We will use (1).

(a) Let $f = \sup_{n\in\mathbb{N}} f_n$. Note that $\{f > c\} = \bigcup_{n\in\mathbb{N}} \{f_n > c\}$. Each of the sets $\{f_n > c\}$ belongs to $\mathcal{B}$, so $\{f > c\} \in \mathcal{B}$.

(b) Similarly to (a), letting $f = \inf_{n\in\mathbb{N}} f_n$, we may express $\{f < c\} = \bigcup_{n\in\mathbb{N}} \underbrace{\{f_n < c\}}_{\in\mathcal{B}} \in \mathcal{B}$.

(c) Recall that $\limsup_{n\to\infty} f_n = \inf_{k\in\mathbb{N}} \sup_{n\geq k} f_n$, so measurability of $\limsup_{n\to\infty} f_n$ follows from (a) and (b).

(d) Similar to (c): $\liminf_{n\to\infty} f_n = \sup_{k\in\mathbb{N}} \inf_{n\geq k} f_n$.

(3) Let $A : \mathbb{R}^2 \to \mathbb{R}$ and $M : \mathbb{R}^2 \to \mathbb{R}$ be the maps $A(x,y) = x+y$ and $M(x,y) = xy$. Both of the maps A and M are continuous and therefore (Borel) measurable. Moreover, $(f+g)(x) = A(f(x), g(x))$ and $(f\cdot g)(x) = M(f(x), g(x))$. Since the composition of measurable maps is measurable (see Proposition 2.8(1)), it suffices to prove $h : x \mapsto (f(x), g(x))$ is a measurable function from X to $\mathbb{R}^2$. By Proposition 2.8(2), we only need to check preimages of sets generating the Borel σ-algebra on $\mathbb{R}^2$. For convenience, we will take the boxes $[a,b)\times[c,d)$ (the first homework problem was to show that every open set in $\mathbb{R}^2$ is a countable (disjoint) union of such boxes, so they generate the Borel σ-algebra). Observe that

$$h^{-1}([a,b)\times[c,d)) = f^{-1}([a,b)) \cap g^{-1}([c,d)) \in \mathcal{B},$$

since f and g are measurable, so h is indeed a measurable function. □

Example 2.12

Let $(X,\mathcal{B})$ be a measurable space and $E \subseteq X$. The function $\mathbb{1}_E$ is measurable if and only if $E \in \mathcal{B}$.

4. Measures

We are now prepared to define measures on abstract measurable spaces.

Definition 2.13

Let $(X,\mathcal{B})$ be a measurable space. A *measure* on $(X,\mathcal{B})$ is a function $\mu : \mathcal{B} \to [0,\infty]$ such that

- $\mu(\emptyset) = 0$;
- countable additivity: for any sequence $(E_n)_{n\in\mathbb{N}}$ of pairwise disjoint elements of $\mathcal{B}$, one has $\mu\left(\bigsqcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \mu(E_n)$.

The triple $(X,\mathcal{B},\mu)$ is called a *measure space*.

Nontrivial examples of measures take some effort to construct, and we will spend significant portions of the course discussing different methods for constructing interesting measures. However, there are a few immediate examples that do not require complicated constructions.

Example 2.14

Examples of measures include:

- For any set X, the *counting measure* is a measure defined on the σ-algebra $\mathscr{P}(X)$ by $\mu(E) = |E|$ if E is a finite set and $\mu(E) = \infty$ if E is an infinite set.
- Given a point $x \in X$, the *Dirac measure* defined on $\mathscr{P}(X)$ is the measure $\delta_x(E) = 1$ if $x \in E$ and $\delta_x(E) = 0$ if $x \notin E$.

We will use the following basic properties of measures frequently throughout this course:

Proposition 2.15

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

(1) MONOTONICITY: For any $A, B \in \mathcal{B}$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

(2) COUNTABLE SUB-ADDITIVITY: For any sequence $(E_n)_{n\in\mathbb{N}}$ in $\mathcal{B}$,

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) \leq \sum_{n\in\mathbb{N}} \mu(E_n).$$

(3) CONTINUITY FROM BELOW: If $E_1 \subseteq E_2 \subseteq \cdots \in \mathcal{B}$, then

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

(4) CONTINUITY FROM ABOVE: If $E_1 \supseteq E_2 \supseteq \cdots \in \mathcal{B}$ and $\mu(E_1) < \infty$, then

$$\mu\left(\bigcap_{n\in\mathbb{N}} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

PROOF. **(1)** Write $B = A \sqcup (B \setminus A)$. Then $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$, since μ takes nonnegative values.

(2) Define a new sequence of sets E'_n by $E'_1 = E_1$ and $E'_n = E_n \setminus \bigcup_{j=1}^{n-1} E_j$ for $n \geq 2$. Then the sets $(E'_n)_{n\in\mathbb{N}}$ are pairwise disjoint and satisfy $E'_n \subseteq E_n$ and $\bigsqcup_{n\in\mathbb{N}} E'_n = \bigcup_{n\in\mathbb{N}} E_n$. Therefore,

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \mu\left(\bigsqcup_{n\in\mathbb{N}} E'_n\right) = \sum_{n\in\mathbb{N}} \mu(E'_n) \leq \sum_{n\in\mathbb{N}} \mu(E_n),$$

where in the last step we have applied monotonicity of μ (property (1)).

(3) Let $E'_1 = E_1$ and $E'_n = E_n \setminus E_{n-1}$ for $n \geq 2$. For convenience, we will set $E_0 = \emptyset$ so that we also have $E'_1 = E_1 \setminus E_0$. Then

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \mu\left(\bigsqcup_{n\in\mathbb{N}} E'_n\right) = \sum_{n\in\mathbb{N}} \mu(E'_n) \overset{(*)}{=} \sum_{n\in\mathbb{N}} (\mu(E_n) - \mu(E_{n-1})) \overset{(**)}{=} \lim_{n\to\infty} \mu(E_n).$$

The step $(*)$ uses additivity of μ, and $(**)$ comes from the telescoping of the sum.

(4) Define a new sequence $A_n = E_1 \setminus E_n$. Then $\emptyset = A_1 \subseteq A_2 \subseteq \ldots$, so

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

by (3). But $\bigcup_{n\in\mathbb{N}} A_n = E_1 \setminus \bigcap_{n\in\mathbb{N}} E_n$, so

$$\mu(E_1) - \mu\left(\bigcap_{n\in\mathbb{N}} E_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \mu(E_1) - \lim_{n\to\infty} \mu(E_n),$$

whence we deduce that (4) holds, since $\mu(E_1) < \infty$. □

Example 2.16

Property (4) may fail if $\mu(E_1) = \infty$. Let $X = \mathbb{N}$, $\mathcal{B} = \mathscr{P}(\mathbb{N})$, and let μ be the counting measure. Let $E_n = \{m \in \mathbb{N} : m \geq n\}$. Then $\mu(E_n) = \infty$ for every $n \in \mathbb{N}$, but $\bigcap_{n\in\mathbb{N}} E_n = \emptyset$, so

$$\mu\left(\bigcap_{n\in\mathbb{N}} E_n\right) = 0 \neq \infty = \lim_{n\to\infty} \mu(E_n).$$

Chapter Notes

The content of this chapter is common to every text on abstract measure theory, though the order of presentation differs. We have elected to follow more or less the order of presentation from Rudin's *Real and Complex Analysis* [**Rud87**, Chapter 1]. Alternative presentations can be found in [**Fol99**, Sections 1.2, 1.3, and 2.1], and [**Tao11**, Section 1.4].

Introductory texts on measure theory tend not to give much treatment to the Borel hierarchy or other topics in descriptive set theory (and we will also not expand on such topics within these lecture notes). Those interested in learning more can take a look at the book of Kechris [**Kec95**] and/or the lecture notes of Tserunyan [**Tse22**], which draw quite heavily on [**Kec95**].

CHAPTER 3

Integration Against a Measure

Learning Objectives

At the end of this chapter, you will be able to:

- Interpret measurability of a function in terms of approximation by simple functions
- Define the integral of a measurable function against a measure
- Prove properties of integration (linearity and fundamental limit theorems)
- Compare and contrast the measure-theoretic approach to integration with the Riemann–Darboux approach
- Apply the definitions and limit theorems to other problems in combinatorics, probability, and analysis

Our next task is to develop an integration theory for integrating measurable functions on abstract measure spaces. In the Riemann–Darboux approach to integration, the strategy is to approximate a general (integrable) function $f : [a, b] \to [0, \infty)$ by step functions, for which the integral is easily defined. This approach has two serious drawbacks that we wish to overcome. One, which we have seen in Example 1.15, is that the Riemann integral does not interact favorably with pointwise limits. The other more severe limitation is that step functions are built as piecewise constant functions on *intervals* (or boxes in higher dimensions), which uses the underlying geometry of Euclidean space. In the setting of abstract measure spaces, there is no geometry on which to rely. In this chapter, we will see that the more general notion of a *simple function* overcomes both of the aforementioned issues with the Riemann integral.

1. Integration of Simple Functions

Definition 3.1

Let $(X, \mathcal{B})$ be a measurable space. A *simple function* is a measurable function $s : X \to \mathbb{C}$ taking only finitely many values.

Partitioning X into finitely many pieces corresponding to the values of a simple function s, we may write simple functions as linear combinations of indicator functions of measurable sets. That is, $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$ for some numbers $c_j \in \mathbb{C}$ and measurable sets $E_j \in \mathcal{B}$. Given a measure μ on $(X, \mathcal{B})$, we define the integral of a simple function in the obvious way. To avoid issues with adding and subtracting infinities, we will deal for now only with nonnegative functions.

Definition 3.2

Let $(X, \mathcal{B}, \mu)$ be a measure space and $s : X \to [0, \infty)$ a simple function. Write $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$ with $c_j \geq 0$ and $E_j \in \mathcal{B}$. The *integral of s with respect to μ* is given by

$$\int_X s \, d\mu = \sum_{j=1}^{n} c_j \mu(E_j).$$

Proposition 3.3

The integral of a nonnegative simple function is well-defined. That is, the value of the integral of a simple function s does not depend on the representation of s as a linear combination of indicator functions of measurable sets.

Proof. Suppose $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$. Let $a_1, \ldots, a_m$ be the finite collection of values taken by s, and let $A_k = \{s = a_k\}$ for $k = 1, \ldots, m$. Then the sets $A_1, \ldots, A_k$ partition X, and $s = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k}$. We will show $\sum_{j=1}^{n} c_j \mu(E_j) = \sum_{k=1}^{m} a_k \mu(A_k)$.

Define a new collection of sets $E'_J = \bigcap_{j \in J} E_j \setminus \bigcup_{i \notin J} E_j$ for $J \subseteq \{1, \ldots, n\}$. In other words, $x \in E'_J$ means that $x \in E_j$ if and only if $j \in J$. This defines a partition of X (see Figure 3.1).

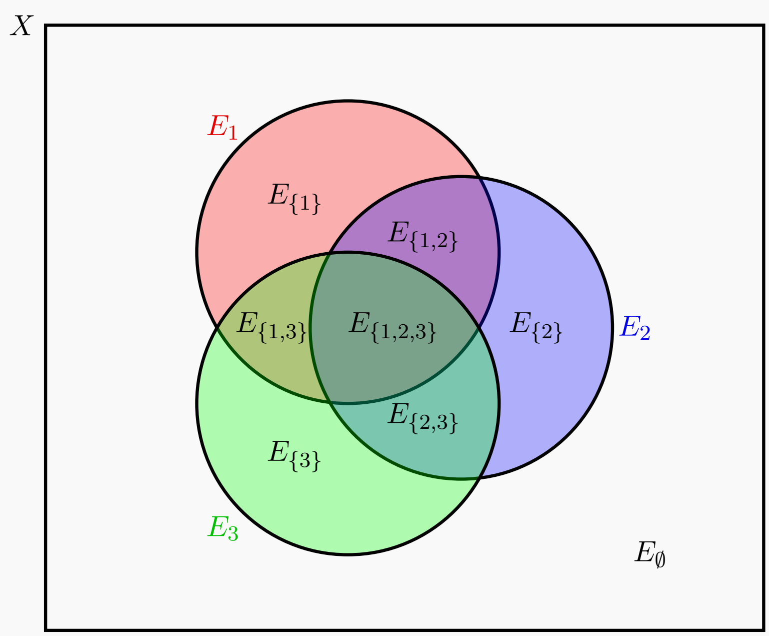


Figure 3.1. Construction of a partition $(E'_J)_{J \subseteq \{1,\ldots,n\}}$ from sets $(E_j)_{j \in \{1,\ldots,n\}}$.

Note that the value of s on the set E'_J is $c'_J = \sum_{j \in J} c_j$. We can therefore relate the sets E'_J to the sets A_k by

$$A_k = \bigsqcup_{J \subseteq \{1,\ldots,n\}, c'_J = a_k} E'_J.$$

Then on the one hand,

$$\sum_{k=1}^{m} a_k\mu(A_k) = \sum_{k=1}^{m} a_k \sum_{J\subseteq\{1,\ldots,n\}, c'_J=a_k} \mu(E'_J) = \sum_{J\subseteq\{1,\ldots,n\}} c'_J\mu(E'_J).$$

On the other hand,

$$\sum_{j=1}^{n} c_j\mu(E_j) = \sum_{j=1}^{n} c_j \sum_{\{j\}\subseteq J\subseteq\{1,\ldots,n\}} \mu(E'_J) = \sum_{J\subseteq\{1,\ldots,n\}} \sum_{j\in J} c_j\mu(E'_J) = \sum_{J\subseteq\{1,\ldots,n\}} c'_J\mu(E'_J).$$

This completes the proof. □

We used a particular representation of a simple function in the previous proof that will continue to be convenient to work with. Say that $\sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$ is the *standard representation* of a simple function s if $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$, and the sets $E_1, \ldots, E_n$ partition X (that is, they are pairwise disjoint and their union is X) and the values $c_1, \ldots, c_n$ are distinct.

Proposition 3.4

Let $(X, \mathcal{B}, \mu)$ be a measure space, let $s, t : X \to [0, \infty)$ be simple functions, and let $c \in \mathbb{R}$, $c \geq 0$. Then
(1) $\int_X cs \; d\mu = c \cdot \int_X s \; d\mu$;
(2) $\int_X (s+t) \; d\mu = \int_X s \; d\mu + \int_X t \; d\mu$;
(3) if $s \leq t$, then $\int_X s \; d\mu \leq \int_X t \; d\mu$.

Proof. **(1)** Let $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$. Then $cs = \sum_{j=1}^{n} (cc_j) \mathbb{1}_{E_j}$, so

$$\int_X cs \; d\mu = \sum_{j=1}^{n} (cc_j)\mu(E_j) = c \cdot \sum_{j=1}^{n} c_j\mu(E_j) = c \cdot \int_X s \; d\mu.$$

For (2) and (3), it will be helpful to work with the standard representation, so let $s = \sum_{j=1}^{n} c_j \mathbb{1}_{E_j}$ and $t = \sum_{k=1}^{m} d_k \mathbb{1}_{F_k}$ be the standard representations. Define sets $A_{j,k} = E_j \cap F_k$ for $j \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, m\}$. Then $E_j = \bigsqcup_{k=1}^{m} A_{j,k}$ and $F_k = \bigsqcup_{j=1}^{n} A_{j,k}$.

(2) The function $s + t$ takes the value $c_j + d_k$ on $A_{j,k}$, so

$$\int_X (s+t) \; d\mu = \sum_{j,k} (c_j + d_k)\mu(A_{j,k}) = \sum_{j=1}^{n} c_j \underbrace{\sum_{k=1}^{m} \mu(A_{j,k})}_{\mu(E_j)} + \sum_{k=1}^{m} d_k \underbrace{\sum_{j=1}^{n} \mu(A_{j,k})}_{\mu(F_k)} = \int_X s \; d\mu + \int_X t \; d\mu.$$

(3) By assumption, if $A_{j,k} \neq \emptyset$, then $c_j \leq d_k$. Thus,

$$\int_X s \; d\mu = \sum_{j=1}^{n} c_j\mu(E_j) = \sum_{j,k} c_j\mu(A_{j,k}) \leq \sum_{j,k} d_k\mu(A_{j,k}) = \sum_{k=1}^{m} d_k\mu(F_k) = \int_X t \; d\mu.$$

□

Definition 3.5

Let $(X, \mathcal{B}, \mu)$ be a measure space, $s : X \to [0, \infty)$ a simple function, and $E \in \mathcal{B}$ a measurable set. The *integral of s with respect to μ over E* is given by

$$\int_E s \; d\mu = \int_X s \cdot \mathbb{1}_E \; d\mu.$$

Note that if s is simple, then $s \cdot \mathbb{1}_E$ is also simple, so the above definition makes sense.

Proposition 3.6

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $s : X \to [0, \infty)$ be a simple function. Then

$$\nu(E) = \int_E s \; d\mu$$

defines a measure on $(X, \mathcal{B})$.

Proof. Note that $s \cdot \mathbb{1}_\emptyset = 0$, so $\nu(\emptyset) = 0$. Suppose $(E_n)_{n\in\mathbb{N}}$ is a pairwise disjoint family of measurable sets, and let $E = \bigsqcup_{n\in\mathbb{N}} E_n$. Write $s = \sum_{j=1}^m a_j \mathbb{1}_{A_j}$. Then $s \cdot \mathbb{1}_E = \sum_{j=1}^m a_j \mathbb{1}_{A_j \cap E}$, so

$$\nu(E) = \sum_{j=1}^m a_j \mu(A_j \cap E) = \sum_{j,n} a_j \mu(A_j \cap E_n) = \sum_{n\in\mathbb{N}} \int_X s \cdot \mathbb{1}_{E_n} \; d\mu = \sum_{n\in\mathbb{N}} \nu(E_n).$$

Note that the sum over n is an infinite sum so reordering requires some justification. Fortunately, all of the values $a_j \mu(A_j \cap E_n)$ are nonnegative, so the sum can be computed in any order without changing the value. □

2. Integration of Nonnegative Measurable Functions

We now want to extend the definition of the integral against a measure to all nonnegative measurable functions. The next proposition shows that simple functions are a sufficiently general class to approximate arbitrary measurable functions.

Proposition 3.7

Let $(X, \mathcal{B})$ be a measurable space, and let $f : X \to [0, \infty]$ be measurable. Then there exists a sequence $(s_n)_{n\in\mathbb{N}}$ of simple functions such that $0 \le s_1 \le s_2 \le \cdots \le f$, and $s_n \to f$ pointwise.

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

$$s_n(x) = \begin{cases} \frac{a}{2^n}, & \text{if } \frac{a}{2^n} \le f(x) < \frac{a+1}{2^n} \text{ and } a < n \cdot 2^n. \\ n, & \text{if } f(x) \ge n. \end{cases}$$

The functions s_n are nondecreasing and satisfy $s_n \le f$ by construction. If $f(x) < \infty$, then $f(x) - s_n(x) < 2^{-n}$ for all $n > f(x)$, so $s_n(x) \to f(x)$ as $n \to \infty$. On the other hand, if $f(x) = \infty$, then $s_n(x) = n$ for every $n \in \mathbb{N}$, so $s_n(x) \to \infty = f(x)$ as $n \to \infty$.

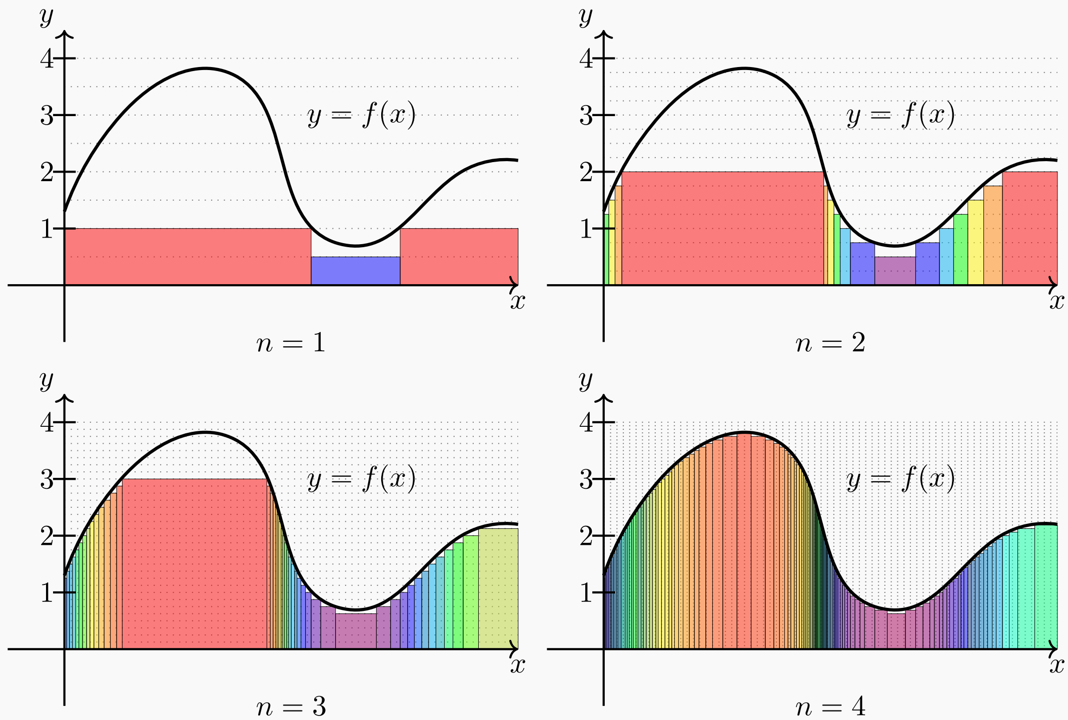


FIGURE 3.2. Successive approximations of a function f by simple functions.

□

It is therefore reasonable to define the integral of an arbitrary nonnegative measurable function as follows.

DEFINITION 3.8

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f : X \to [0, \infty]$ be measurable. We define the *integral of f with respect to μ* as

$$\int_X f \, d\mu = \sup \left\{ \int_X s \, d\mu : s \text{ simple and } 0 \leq s \leq f \right\}.$$

Given a measurable set $E \in \mathcal{B}$, the *integral of f with respect to μ over E* is defined by

$$\int_E f \, d\mu = \int_X f \cdot \mathbb{1}_E \, d\mu.$$

One may object at this point and suggest an alternative definition. Since $f : X \to [0, \infty]$ can be obtained as $f = \lim_{n\to\infty} s_n$ for an increasing sequence of simple functions $0 \leq s_1 \leq s_2 \leq \ldots$, why not define $\int_X f \, d\mu = \lim_{n\to\infty} \int_X s_n \, d\mu$? As we will see shortly, this is in fact an equivalent definition that is extremely useful for many applications. However, *as a definition*, it has two serious defects: why should the limit exist? and why should the value be the same for all possible approximations by simple functions? This is why we prefer Definition 3.8 above (and why this is the standard definition across measure theory textbooks).

Proposition 3.9

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g : X \to [0, \infty]$ be measurable. If $f \leq g$, then

$$\int_X f \, d\mu \leq \int_X g \, d\mu.$$

Proof. It suffices to observe $\{s \text{ simple function} : 0 \leq s \leq f\} \subseteq \{s \text{ simple function} : 0 \leq s \leq g\}$. □

Theorem 3.10: Monotone Convergence Theorem

Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $0 \leq f_1 \leq f_2 \leq \ldots$, and let $f = \lim_{n\to\infty} f_n$. Then

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

Remark. Note that a consequence of the monotone convergence theorem is that $\int_X f \, d\mu$ can be computed by taking a sequence of simple functions $0 \leq s_1 \leq s_2 \leq \cdots \to f$ and computing $\lim_{n\to\infty} \int_X s_n \, d\mu$.

Proof of Monotone Convergence Theorem. First, f is a measurable function by Proposition 2.11. By monotonicity of the integral (Proposition 3.9), the sequence $\int_X f_n \, d\mu$ is increasing, so $\lim_{n\to\infty} \int_X f_n \, d\mu = \sup_{n\in\mathbb{N}} \int_X f_n \, d\mu \in [0, \infty]$ exists as an extended real number. Moreover,

$$\int_X f \, d\mu \geq \lim_{n\to\infty} \int_X f_n \, d\mu,$$

since the inequality holds for each $n \in \mathbb{N}$. Therefore, it suffices to show

$$\int_X f \, d\mu \leq \lim_{n\to\infty} \int_X f_n \, d\mu.$$

If $\lim_{n\to\infty} \int_X f_n \, d\mu = \infty$, there is nothing to prove, so assume $\lim_{n\to\infty} \int_X f_n \, d\mu < \infty$.

Let $c < 1$. Let $s : X \to [0, \infty)$ be a simple function, $0 \leq s \leq f$. For $n \in \mathbb{N}$, let $E_n = \{f_n \geq cs\}$. Then $E_1 \subseteq E_2 \subseteq \ldots$ and $X = \bigcup_{n\in\mathbb{N}} E_n$. By Proposition 3.6, let $\nu : \mathcal{B} \to [0, \infty]$ be the measure $\nu(E) = \int_E s \, d\mu$. We have

$$\begin{aligned}
c \cdot \int_X s \, d\mu &= c \cdot \nu(X) \\
&= c \cdot \lim_{n\to\infty} \nu(E_n) && \text{(continuity from below)} \\
&= \lim_{n\to\infty} c \cdot \nu(E_n) && \text{(Proposition 2.10)} \\
&= \lim_{n\to\infty} \int_{E_n} cs \, d\mu && \text{(Proposition 3.4)} \\
&\leq \lim_{n\to\infty} \int_X f_n \, d\mu && \text{(monotonicity).}
\end{aligned}$$

Taking a supremum over all such simple functions, we conclude

$$c \cdot \int_X f \, d\mu \leq \lim_{n\to\infty} \int_X f_n \, d\mu.$$

Letting $c \to 1$ yields the desired result. □

Proposition 3.11

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g : X \to [0, \infty]$ be measurable functions. Let $c \in [0, \infty)$.
(1) $\int_X cf \; d\mu = c \cdot \int_X f \; d\mu$.
(2) $\int_X (f + g) \; d\mu = \int_X f \; d\mu + \int_X g \; d\mu$.

Proof. **(1)** This follows quickly from the definition of the integral and Proposition 3.4.

(2) We use the monotone convergence theorem. Let $0 \leq s_1 \leq s_n \leq \cdots \leq f$ with $s_n \to f$ and $0 \leq t_1 \leq t_2 \leq \cdots \leq g$ with $t_n \to g$. Then $0 \leq s_1 + t_1 \leq s_2 + t_2 \leq \cdots \leq f + g$ and $s_n + t_n \to f + g$. Thus,

$$\begin{aligned} \int_X (f+g) \; d\mu &= \lim_{n\to\infty} \int_X (s_n + t_n) \; d\mu && \text{(MCT)} \\ &= \lim_{n\to\infty} \int_X s_n \; d\mu + \lim_{n\to\infty} \int_X t_n \; d\mu && \text{(Proposition 3.4)} \\ &= \int_X f \; d\mu + \int_X g \; d\mu && \text{(MCT).} \end{aligned}$$

□

Theorem 3.12

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions, $f_n : X \to [0, \infty]$. Then

$$\int_X \left(\sum_{n=1}^{\infty} f_n \right) \; d\mu = \sum_{n=1}^{\infty} \int_X f_n \; d\mu.$$

Proof. We have

$$\begin{aligned} \int_X \left(\sum_{n=1}^{\infty} f_n \right) \; d\mu &= \int_X \lim_{N\to\infty} \left(\sum_{n=1}^{N} f_n \right) \; d\mu \\ &= \lim_{N\to\infty} \int_X \left(\sum_{n=1}^{N} f_n \right) \; d\mu && \text{(MCT)} \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \int_X f_n \; d\mu && \text{(additivity of the integral)} \\ &= \sum_{n=1}^{\infty} \int_X f_n \; d\mu. \end{aligned}$$

□

Theorem 3.13: Fatou's Lemma

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions, $f_n : X \to [0, \infty]$. Then

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

PROOF. Let $f = \liminf_{n\to\infty} f_n$. Define $F_N = \inf_{n\geq N} f_n$. Then $0 \leq F_1 \leq F_2 \leq \dots$ and $F_N \to f$. Therefore,

$$\begin{aligned}\int_X f \, d\mu &= \lim_{N\to\infty} \int_X F_N \, d\mu && \text{(MCT)} \\ &\leq \lim_{N\to\infty} \inf_{n\geq N} \int_X f_n \, d\mu && \text{(monotonicity of the integral)} \\ &= \liminf_{N\to\infty} \int_X f_n \, d\mu.\end{aligned}$$

□

3. Integration of Real and Complex-Valued Functions

The method for integrating real and complex-valued functions involves decomposing these functions as linear combinations of nonnegative functions. An important observation is that such a decomposition can be done in a measurable way.

DEFINITION 3.14

Let X be a set and $f : X \to [-\infty, \infty]$. The *positive part* f^+ and *negative part* f^- of f are defined by

$$f^+ = \max\{f, 0\} \qquad \text{and} \qquad f^- = \max\{-f, 0\}.$$

Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Moreover, if $(X, \mathcal{B})$ is a measurable space and $f : X \to [-\infty, \infty]$ is measurable, then f^+ and f^- are measurable by Proposition 2.11.

DEFINITION 3.15

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

- An extended real-valued measurable function $f : X \to [-\infty, \infty]$ is *integrable* if
$$\int_X |f| \, d\mu < \infty.$$
In this case, the *integral of f* is defined by
$$\int_X f \, d\mu = \int_X f^+ \, d\mu - \int_X f^- \, d\mu.$$
- A complex-valued measurable function $f : X \to \mathbb{C}$ is *integrable* if
$$\int_X |f| \, d\mu < \infty,$$
and the *integral of f* is defined by
$$\int_X f \, d\mu = \int_X \mathrm{Re}\,(f) \, d\mu + i \int_X \mathrm{Im}\,(f) \, d\mu.$$
- Given a measurable set $E \in \mathcal{B}$, a measurable function f taking extended real or complex values is *integrable over E* if $f \cdot \mathbb{1}_E$ is integrable, and the *integral of f over E* is
$$\int_E f \, d\mu = \int_X f \cdot \mathbb{1}_E \, d\mu.$$

REMARK. By monotonicity of the integral (Proposition 3.9), if a function is integrable, then it is also integrable over every measurable subset of X.

4. Integral Identities and Inequalities

Proposition 3.16: Linearity of the Integral

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{C}$ be integrable functions, and let $c \in \mathbb{C}$. Then

(1) $f + g$ is integrable, and $\int_X (f+g)\ d\mu = \int_X f\ d\mu + \int_X g\ d\mu$.
(2) cf is integrable, and $\int_X cf\ d\mu = c \int_X f\ d\mu$.

Proof. **(1)** First, by the triangle inequality, we have $|f+g| \le |f| + |g|$. Therefore,

$$\int_X |f+g|\ d\mu \overset{(*)}{\le} \int_X (|f| + |g|)\ d\mu \overset{(**)}{=} \int_X |f|\ d\mu + \int_X |g|\ d\mu < \infty.$$

In step $(*)$, we have used monotonicity of the integral (Proposition 3.9), and in $(**)$, we have used additivity (Proposition 3.11).

Decomposing f and g into their real and imaginary parts, it suffices to prove the identity $\int_X (f+g)\ d\mu = \int_X f\ d\mu + \int_X g\ d\mu$ for real-valued functions f and g. Let $h = f + g$. Then $h = h^+ - h^- = f^+ - f^- + g^+ - g^-$. This can be rearranged to the identity $h^+ + f^- + g^- = h^- + f^+ + g^+$. Then using additivity of the integral for nonnegative functions (Proposition 3.11), we have

$$\begin{aligned}\int_X h^+\ d\mu + \int_X f^-\ d\mu + \int_X g^-\ d\mu &= \int_X (h^+ + f^- + g^-)\ d\mu \\ &= \int_X (h^- + f^+ + g^+)\ d\mu = \int_X h^-\ d\mu + \int_X f^+\ d\mu + \int_X g^+\ d\mu. \quad (3.1)\end{aligned}$$

Rearranging again,

$$\begin{aligned}\int_X (f+g)\ d\mu &= \int_X h^+\ d\mu - \int_X h^-\ d\mu && \text{(Definition 3.15)} \\ &= \int_X f^+\ d\mu - \int_X f^-\ d\mu + \int_X g^+\ d\mu - \int_X g^-\ d\mu && \text{(by (3.1))} \\ &= \int_X f\ d\mu + \int_X g\ d\mu && \text{(Definition 3.15)}\end{aligned}$$

(2) Note that $|cf| = |c||f|$, so

$$\int_X |cf|\ d\mu = \int_X |c||f|\ d\mu \overset{(*)}{=} |c| \int_X |f|\ d\mu < \infty,$$

where $(*)$ follows from Proposition 3.11. Hence, cf is integrable.

For computing the integral of cf, we consider several different cases.

Case 1. $c \ge 0$

When f is nonnegative, we have

$$\int_X cf\ d\mu = c \int_X f\ d\mu$$

by Proposition 3.11. The identity follows for a general complex-valued function f by decomposing $f = \left(\mathrm{Re}\,(f)^+ - \mathrm{Re}\,(f)^-\right) + i\left(\mathrm{Im}\,(f)^+ - \mathrm{Im}\,(f)^-\right)$.

Case 2. $c = -1$

For real-valued $f : X \to \mathbb{R}$, we use the identities $(-f)^+ = f^-$ and $(-f)^- = f^+$ to obtain
$$\int_X (-f)\ d\mu = \int_X f^-\ d\mu - \int_X f^+\ d\mu = -\int_X f\ d\mu.$$
Complex-valued functions can be handled by decomposing into real and imaginary parts.

Case 3. $c \in \mathbb{R}$

Combine Case 1 and Case 2.

Case 4. $c = i$

Noting that $\operatorname{Re}(if) = -\operatorname{Im}(f)$ and $\operatorname{Im}(if) = \operatorname{Re}(f)$, we have
$$\begin{aligned}
\int_X if\ d\mu &= \int_X (-\operatorname{Im}(f))\ d\mu + i\int_X \operatorname{Re}(f)\ d\mu && \text{(Definition 3.15)}\\
&= -\int_X \operatorname{Im}(f)\ d\mu + i\int_X \operatorname{Re}(f)\ d\mu && \text{(Case 2)}\\
&= i\left(\int_X \operatorname{Re}(f)\ d\mu + i\int_X \operatorname{Im}(f)\ d\mu\right)\\
&= i\int_X f\ d\mu && \text{(Definition 3.15)}
\end{aligned}$$

Case 5. $c \in \mathbb{C}$

Write $c = a + ib$ with $a, b \in \mathbb{R}$. Then
$$\begin{aligned}
\int_X cf\ d\mu &= \int_X (af + ibf)\ d\mu\\
&= \int_X af\ d\mu + \int_X ibf\ d\mu && \text{(by (1))}\\
&= \int_X af\ d\mu + i\int_X bf\ d\mu && \text{(Case 3)}\\
&= a\int_X f\ d\mu + ib\int_X f\ d\mu && \text{(Case 4)}\\
&= c\int_X f\ d\mu.
\end{aligned}$$

□

Let $(X, \mathcal{B}, \mu)$ be a measure space, and denote by $L^1(\mu)$ the set of integrable functions. Proposition 3.16 shows that $L^1(\mu)$ is a (complex) vector space. Moreover, in the course of the proof, we showed

$$\int_X |cf|\ d\mu = |c|\int_X |f|\ d\mu \qquad \text{and} \qquad \int_X |f+g|\ d\mu \le \int_X |f|\ d\mu + \int_X |g|\ d\mu.$$

In other words, if we let

$$\|f\|_1 = \int_X |f| \, d\mu,$$

then $\|\cdot\|_1$ defines a *seminorm* on the vector space of integrable functions on $(X, \mathcal{B}, \mu)$.

Definition 3.17

Let V be a real or complex vector space. A function $\|\cdot\| : V \to [0, \infty)$ is a *seminorm* if it satisfies:

- TRIANGLE INEQUALITY: $\|u + v\| \le \|u\| + \|v\|$ for all $u, v \in V$, and
- ABSOLUTE HOMOGENEITY: $\|cv\| = |c| \, \|v\|$ for all $v \in V$ and all scalars c.

A seminorm is a *norm* if it satisfies the additional property

- POSITIVE DEFINITE: if $v \in V$ and $\|v\| = 0$, then $v = 0$.

The seminorm $\|\cdot\|_1$ on the space of integrable functions may not be a norm in general, but a small modification will turn it into a norm. This will be discussed in greater detail later in the course, in the context of so-called L^p spaces. One of the important ingredients is a deeper understanding of *null sets*, which we will discuss in Section 5 below.

As another basic property of integration, we establish a version of the triangle inequality:

Proposition 3.18: Triangle Inequality for the Integral

Suppose $(X, \mathcal{B}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is an integrable function. Then

$$\left| \int_X f \, d\mu \right| \le \int_X |f| \, d\mu.$$

PROOF. First, suppose f is real-valued. Then by the triangle inequality and linearity,

$$\left| \int_X f \, d\mu \right| = \left| \int_X f^+ \, d\mu - \int_X f^- \, d\mu \right| \le \int_X f^+ \, d\mu + \int_X f^- \, d\mu = \int_X |f| \, d\mu.$$

Now suppose f is complex-valued. Let $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ such that $\left| \int_X f \, d\mu \right| = \lambda \int_X f \, d\mu$. Then by Proposition 3.16,

$$\left| \int_X f \, d\mu \right| = \int_X \lambda f \, d\mu.$$

Since this is a real number, we then have $\int_X \lambda f \, d\mu = \int_X \mathrm{Re}\,(\lambda f) \, d\mu$, so by monotonicity of the integral,

$$\left| \int_X f \, d\mu \right| = \int_X \mathrm{Re}\,(\lambda f) \, d\mu \le \int_X |\lambda f| \, d\mu = \int_X |f| \, d\mu.$$

□

5. Sets of Measure Zero

Definition 3.19

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

- A measurable set $N \in \mathcal{B}$ is a *null set* if $\mu(N) = 0$.
- We say that a property holds *almost everywhere* if there exists a null set $N \in \mathcal{B}$ such that the property holds for every point $x \in X \setminus N$.

Remark. An easy consequence of countable additivity and monotonicity of measures is that the family $\mathcal{N}$ of null sets forms a σ-*ideal* of $\mathcal{B}$:

- $\emptyset \in \mathcal{N}$;
- if $A \in \mathcal{N}$ and $B \in \mathcal{B}$ with $B \subseteq A$, then $B \in \mathcal{N}$; and
- if $(N_n)_{n\in\mathbb{N}}$ is a countable family of null sets, then $\bigcup_{n\in\mathbb{N}} N_n \in \mathcal{N}$.

Notation. The phrases "almost everywhere" or "almost every" are often abbreviated by a.e. or μ-a.e. if the measure needs to be specified. In a statement of the form "Property P holds a.e.," we interpret a.e. as "almost everywhere." For a statement of the form "Property P holds for a.e. $x \in X$," we read a.e. as "almost every," and the meaning is the same as in the previous example statement.

Null sets naturally arise and play an important role in integration theory. Some examples are provided by the next three propositions.

Proposition 3.20

Let $(X, \mathcal{B}, \mu)$ be a measure space. Suppose $f : X \to [-\infty, \infty]$ is an integrable function. Then $f(x) \in \mathbb{R}$ for μ-a.e. $x \in X$.

Proof. Let $N = \{x \in \mathbb{R} : |f(x)| = \infty\}$. We want to show that N is a null set. By monotonicity of the integral (Proposition 3.9),

$$\int_X |f| \, d\mu \geq \int_N |f| \, d\mu = \infty \cdot \mu(N).$$

On the other hand, by integrability of f,

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

Thus, $\infty \cdot \mu(N) < \infty$, so $\mu(N) = 0$. □

Corollary 3.21: Borel–Cantelli Lemma

Let $(X, \mathcal{B}, \mu)$ be a measure space. Suppose $(E_n)_{n\in\mathbb{N}}$ is a sequence of measurable sets and $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. Then

$$\mu\left(\{x \in X : x \in E_n \text{ for infinitely many } n \in \mathbb{N}\}\right) = 0.$$

Proof. One possible proof uses continuity from above of the measure μ. We will now give a different proof using integration.

Let $f = \sum_{n=1}^{\infty} \mathbb{1}_{E_n}$. Note that $f(x) = \infty$ if and only if $x \in E_n$ for infinitely many $n \in \mathbb{N}$. By Theorem 3.12,

$$\int_X f \, d\mu = \sum_{n=1}^{\infty} \underbrace{\int_X \mathbb{1}_{E_n} \, d\mu}_{\mu(E_n)} < \infty.$$

So by Proposition 3.20, $f < \infty$ a.e. That is,

$$\mu\left(\{x \in X : x \in E_n \text{ for infinitely many } n \in \mathbb{N}\}\right) = \mu(\{f = \infty\}) = 0.$$

□

Proposition 3.22

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g : X \to \mathbb{C}$ be measurable functions. Suppose $f = g$ a.e. Then f is integrable if and only if g is integrable. Moreover, if f and g are integrable, then

$$\int_X f \, d\mu = \int_X g \, d\mu.$$

Proof. Let $N = \{x \in X : f(x) \neq g(x)\}$. By assumption, N is a null set.

Step 1. Integrability

Suppose f is integrable. Then

$$\begin{aligned} \int_X |g| \, d\mu &= \int_{X \setminus N} |f| \, d\mu + \int_N |g| \, d\mu && \text{(linearity of the integral)} \\ &\leq \int_X |f| \, d\mu + \underbrace{\infty \cdot \mu(N)}_{0} && \text{(monotonicity of the integral)} \\ &= \int_X |f| \, d\mu < \infty, \end{aligned}$$

so g is integrable. Reversing the roles of f and g proves the converse.

Step 2. Integral

Assume f and g are integrable. Then

$$\begin{aligned} \left| \int_X g \, d\mu - \int_X f \, d\mu \right| &= \left| \int_X (g - f) \, d\mu \right| && \text{(linearity of the integral)} \\ &\leq \int_X |g - f| \, d\mu && \text{(triangle inequality for the integral)} \\ &= \int_{X \setminus N} 0 \, d\mu + \int_N |g - f| \, d\mu && \text{(linearity of the integral)} \\ &\leq 0 \cdot \mu(X \setminus N) + \infty \cdot \mu(N) = 0. \end{aligned}$$

□

Proposition 3.23

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f : X \to [0, \infty]$ be a measurable function. Then $\int_X f \, d\mu = 0$ if and only if $f = 0$ a.e.

Proof. If $f = 0$ a.e., then by Proposition 3.22, f is integrable and

$$\int_X f \, d\mu = \int_X 0 \, d\mu = 0 \cdot \mu(X) = 0.$$

Conversely, suppose $\int_X f \; d\mu = 0$. Then by Markov's inequality (Problem 3 on Exercise Sheet #3),

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

for every $c > 0$. Therefore, by continuity of μ from below,

$$\mu(\{f \neq 0\}) = \mu\left(\bigcup_{n \in \mathbb{N}} \left\{f > \frac{1}{n}\right\}\right) = \lim_{n\to\infty} \mu\left(\left\{f > \frac{1}{n}\right\}\right) = 0.$$

That is, $f = 0$ a.e. □

The examples above (especially Proposition 3.22) show that null sets are negligible from the point of view of integration, and we can very often ignore modifications that happen on null sets. There is one subtle issue that requires care, however: in general, a subset of a null set may not be measurable and non-measurable modifications on null sets may create issues. For this reason, it is often convenient to work with *complete* measure spaces, as defined below.

Definition 3.24

A measure space $(X, \mathcal{B}, \mu)$ is *complete* if every subset of every null set is measurable. That is, if $E \subseteq X$ and there exists $N \in \mathcal{B}$ with $E \subseteq N$ and $\mu(N) = 0$, then $E \in \mathcal{B}$.

The following proposition is a useful tool for passing to complete measure spaces.

Proposition 3.25

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $\mathcal{N} = \{N \in \mathcal{B} : \mu(N) = 0\}$ be the σ-ideal of μ-null sets. Then the family $\overline{\mathcal{B}} = \{E \cup F : E \in \mathcal{B}, F \subseteq N \in \mathcal{N}\}$ is a σ-algebra, and there is a unique extension $\overline{\mu}$ of μ to $\overline{\mathcal{B}}$.

Proof. Exercise. □

Definition 3.26

The *completion* of a measure space $(X, \mathcal{B}, \mu)$ is the space $(X, \overline{\mathcal{B}}, \overline{\mu})$, where $\overline{\mathcal{B}}$ and $\overline{\mu}$ are as defined in Proposition 3.25.

6. The Dominated Convergence Theorem

We have already seen two fundamental convergence theorems for integration against a measure: the monotone convergence theorem and Fatou's lemma. We are nearly ready to state another fundamental result about integration: the dominated convergence theorem. First, we need to introduce the two notions of convergence that will be related by the dominated convergence theorem.

Definition 3.27

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

- We say that a sequence $(f_n)_{n\in\mathbb{N}}$ of functions on X *converges almost everywhere* to a function f if $\lim_{n\to\infty} f_n(x) = f(x)$ for almost every $x \in X$.

- A sequence $(f_n)_{n\in\mathbb{N}}$ of integrable functions *converges in L^1* to $f \in L^1(\mu)$ if
$$\|f_n - f\|_1 = \int_X |f_n - f| \, d\mu \to 0$$
in $\mathbb{R}$ as $n \to \infty$.

The dominated convergence theorem says that any sequence that converges almost everywhere and is "L^1-dominated" will converge in L^1. The precise mathematical formulation is as follows:

Theorem 3.28: Dominated Convergence Theorem

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of integrable functions, $f_n : X \to \mathbb{C}$, and let $f : X \to \mathbb{C}$ be measurable. Suppose

- $f_n \to f$ a.e., and
- there is an integrable function $g : X \to [0, \infty)$ such that $\sup_{n\in\mathbb{N}} |f_n| \le g$ a.e.

Then f is integrable and $f_n \to f$ in $L^1(\mu)$. In particular,
$$\int_X f \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu.$$

Proof. First, $|f| \le |g|$ a.e., so f is integrable.

Observe:
$$\begin{aligned}
\int_X 2g \, d\mu - \limsup_{n\to\infty} \int_X |f - f_n| \, d\mu &= \liminf_{n\to\infty} \int_X (2g - |f - f_n|) \, d\mu \\
&\ge \int_X \liminf_{n\to\infty} (2g - |f - f_n|) \, d\mu && \text{(Fatou's lemma)} \\
&= \int_X 2g \, d\mu && (f_n \to f)
\end{aligned}$$
Rearranging, we conclude
$$\limsup_{n\to\infty} \int_X |f - f_n| \, d\mu \le 0.$$
Using the triangle inequality for the integral,
$$\left| \int_X f \, d\mu - \int_X f_n \, d\mu \right| \le \int_X |f - f_n| \, d\mu \to 0,$$
so
$$\int_X f \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu.$$
□

The assumption that the sequence $(f_n)_{n\in\mathbb{N}}$ is "dominated" by an integrable function g is a necessary assumption to avoid "escape of mass to infinity," as the following example demonstrates.

Example 3.29

Let $X = \mathbb{Z}$, $\mathcal{B} = \mathscr{P}(\mathbb{Z})$, and let μ be the counting measure. Let $f_n = \mathbb{1}_{\{n\}}$. Then $f_n(x) \to 0$ for every $x \in X$. However,
$$\int_X f_n \, d\mu = 1$$

for every $n \in \mathbb{N}$, while
$$\int_X \lim_{n\to\infty} f_n \, d\mu = \int_X 0 \, d\mu = 0 \neq 1.$$

Chapter Notes

In this chapter, we defined the integral of a real or complex-valued function only in the case that the function is integrable functions (see Definition 3.15). The reason for insisting on integrability is to avoid the problematic expression $\infty - \infty$. Nevertheless, as we saw in the definition of the integral for nonnegative measurable functions, it makes sense in certain instances to consider integrals taking infinite values. The general case in which we can make sense of the value of the integral of an extended real-valued function is to define
$$\int_X f \, d\mu = \int_X f^+ \, d\mu - \int_X f^- \, d\mu$$
whenever at least one of the integrals on the right hand side is finite. For this definition of the integral, we still have the properties $\int_X cf \, d\mu = c\int_X f \, d\mu$ for $c \in \mathbb{R}$ and $|\int_X f \, d\mu| \leq \int_X |f| \, d\mu$. However, one needs to be careful when working with sums $f + g$. For example, if $\int_X f^+ \, d\mu = \infty$ and $\int_X g^- \, d\mu = \infty$, then the integral of $f + g$ may be undefined.

For other presentations of integration on abstract measures spaces, see [**Fol99**, Section 2.1–2.3], [**Rud87**, Chapter 1], [**SS05**, Sections 2.1 and 6.2], and/or [**Tao11**, Section 1.3 and Subsection 1.4.4]. The development of integration in the books of Folland [**Fol99**] and Rudin [**Rud87**] is very similar to the presentation in these notes. By contrast, Stein and Shakarchi [**SS05**] and Tao [**Tao11**] first develop integration in the special case of the Lebesgue measure before moving to abstract spaces. The book of Stein and Shakarchi [**SS05**] also proves the fundamental convergence theorems in a different order, starting with a special case of the dominated convergence theorem known as the *bounded convergence theorem*, and then deducing Fatou's lemma, the monotone convergence theorem, and the general case of the dominated convergence theorem.

There is a very nice book of Oxtoby [**Oxt80**] that develops useful analogies between measure spaces and topological spaces and includes a discussion of null sets in relation to a σ-ideal of "topologically negligible" sets called *meager* sets or sets of *first category*.

Part 2

Constructions of Measures

CHAPTER 4

Lebesgue–Stieltjes Measures

Learning Objectives

At the end of this chapter, you will be able to:

- Reinterpret and resolve the problem of measurement in a measure-theoretic framework
- Construct nontrivial measures on the real line

Let us rephrase (an instance of) the problem of measurement using the language of abstract measure theory developed in Part 1.

Problem 4.1: Problem of Measurement in One Dimension

Construct a measure $\lambda : \mathrm{Borel}(\mathbb{R}) \to [0, \infty]$ such that $\lambda(I) = \mathrm{length}(I)$ for every interval $I \subseteq \mathbb{R}$. Is there a unique such measure? Can the measure be defined on all subsets of $\mathbb{R}$?

We will address this problem in a more general framework, where we allow for different assignments of measure to intervals.

Definition 4.2

A Borel measure $\mu : \mathrm{Borel}(\mathbb{R}) \to [0, \infty]$ is *locally finite* if $\mu(K) < \infty$ for every compact set $K \subseteq \mathbb{R}$. The *distribution function* of a locally finite Borel measure μ is the function $F_\mu : \mathbb{R} \to \mathbb{R}$ defined by

$$F_\mu(x) = \begin{cases} \mu((0, x]), & \text{if } x > 0; \\ 0, & \text{if } x = 0; \\ -\mu((x, 0]), & \text{if } x < 0. \end{cases}$$

By monotonicity of the measure μ, its distribution function F_μ is necessarily increasing. Moreover, by continuity from above and below, F_μ is a right-continuous function. The goal of this chapter is to prove the following theorem:

Theorem 4.3: Existence and Uniqueness of Lebesgue–Stieltjes Measures

Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing, right-continuous function with $F(0) = 0$. There exists a σ-algebra $\mathcal{M}_F$ containing the Borel subsets of $\mathbb{R}$ and a complete measure $\mu_F : \mathcal{M}_F \to [0, \infty]$ such that $F = F_{\mu_F}$. Moreover, if $\nu : \mathrm{Borel}(\mathbb{R}) \to [0, \infty]$ is a Borel measure satisfying $F_\nu = F$, then $\nu = \mu_F|_{\mathrm{Borel}(\mathbb{R})}$, and μ_F is the completion of ν.

Definition 4.4

Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing, right-continuous function. The unique complete measure μ_F given by Theorem 4.3 is called the *Lebesgue–Stieltjes measure* associated to F.

Proposition 4.5

Let F be an increasing, right-continuous function, and let μ_F be the Lebesgue–Stieltjes measure associated to F. Then

$$\lim_{x\to\infty} F(x) = \sup_{x\in\mathbb{R}} F(x) = \mu_F((0,\infty)) \qquad \text{and} \qquad \lim_{x\to-\infty} F(x) = \inf_{x\in\mathbb{R}} F(x) = -\mu_F((-\infty,0]).$$

Proof. This is an application of continuity from below of the measure μ_F. □

Notation. Given an increasing, right-continuous function, we will write $F(\infty)$ for the value $\lim_{x\to\infty} F(x)$ and $F(-\infty) = \lim_{x\to-\infty} F(x)$. In general, $F(\pm\infty)$ is an extended real number.

1. The π-λ Theorem and Uniqueness of Lebesgue–Stieltjes Measures

Before constructing Lebesgue–Stieltjes measures, let us prove that every locally finite Borel measure is uniquely determined by its distribution function. The key tool will be the π-λ theorem, for which we need a new definition.

Definition 4.6

Let X be a set.

- A family $\mathcal{P} \subseteq \mathscr{P}(X)$ of subsets of X is a *π-system* if $\mathcal{P}$ is closed under finite intersections.
- A family $\mathcal{L} \subseteq \mathscr{P}(X)$ is a *λ-system* if $\emptyset \in \mathcal{L}$ and $\mathcal{L}$ is closed under complements and countable disjoint unions.

Example 4.7

The following are examples of π systems:

- the collection $\mathcal{P} = \{(a,b] : a,b \in \mathbb{R}\}$ of half-open intervals in $\mathbb{R}$;
- the family of open sets of any topological space;
- given a measure space $(X,\mathcal{B},\mu)$, the family $\mathcal{P} = \{E \in \mathcal{B} : \mu(X \setminus E) = 0\}$ of co-null sets;
- given two measurable spaces $(X,\mathcal{B})$ and $(Y,\mathcal{C})$, the family $\mathcal{P} = \{B \times C : B \in \mathcal{B}, C \in \mathcal{C}\}$ of "rectangles" in $X \times Y$.

Examples of λ-systems include:

- for two probability measures μ,ν on a measurable space $(X,\mathcal{B})$, the family $\mathcal{L} = \{E \in \mathcal{B} : \mu(E) = \nu(E)\}$.

Another characterization of λ-systems is given by the following proposition:

Proposition 4.8

Let X be a set. A family $\mathcal{L} \subseteq \mathscr{P}(X)$ is a λ-system if and only if it satisfies the following three properties:

(1) $X \in \mathcal{L}$;
(2) if $A, B \in \mathcal{L}$ and $A \subseteq B$, then $B \setminus A \in \mathcal{L}$;
(3) if $A_1 \subseteq A_2 \subseteq \dots$ is an increasing sequence in $\mathcal{L}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{L}$.

Proof. Suppose $\mathcal{L}$ is a λ-system. We check that $\mathcal{L}$ satisfies properties (1)–(3).

(1) Since $\emptyset \in \mathcal{L}$ and $\mathcal{L}$ is closed under complements, we have $X \in \mathcal{L}$.

(2) Let $A, B \in \mathcal{L}$ with $A \subseteq B$. Then $B \setminus A = B \cap A^c = (B^c \cup A)^c$. The assumption $A \subseteq B$ means $B^c \cap A = \emptyset$, so we have represented $B \setminus A$ in terms of A and B using complementation and disjoint union. Hence, $B \setminus A \in \mathcal{L}$.

(3) Let $A_1 \subseteq A_2 \subseteq \dots$ be an increasing sequence in $\mathcal{L}$. Let $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. By property (ii), $B_n \in \mathcal{L}$ for every $n \in \mathbb{N}$. Therefore, $\bigcup_{n\in\mathbb{N}} A_n = \bigsqcup_{n\in\mathbb{N}} B_n \in \mathcal{L}$.

Conversely, suppose $\mathcal{L} \subseteq \mathscr{P}(X)$ is a family of sets satisfying (1), (2), and (3).

Applying property (2) with $A = B = X$, we have $\emptyset = X \setminus X \in \mathcal{L}$.

Let $A \in \mathcal{L}$. Combining (1) and (2), $A^c = X \setminus A \in \mathcal{L}$.

Finally, let $(A_n)_{n\in\mathbb{N}}$ be a pairwise disjoint sequence of sets in $\mathcal{L}$. Then $B_n = A_1 \sqcup \dots \sqcup A_n$ forms an increasing sequence, so by property (3), it suffices to prove that $B_n \in \mathcal{L}$. By induction, this reduces to showing that the disjoint union of two sets in $\mathcal{L}$ is an element of $\mathcal{L}$. Let $C, D \in \mathcal{L}$ with $C \cap D = \emptyset$. Then $C \sqcup D = (C^c \cap D^c)^c = (C^c \setminus D)^c$. We have already checked that $\mathcal{L}$ is closed under complementation. The disjointness of C and D implies $D \subseteq C^c$, so $C^c \setminus D \in \mathcal{L}$ by property (2). Thus, $C \sqcup D \in \mathcal{L}$. □

Theorem 4.9: π-λ Theorem (Sierpiński–Dynkin)

Let X be a set, and suppose $\mathcal{P} \subseteq \mathscr{P}(X)$ is a π-system. If $\mathcal{L} \subseteq \mathscr{P}(X)$ is a λ-system and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

We will prove the π-λ theorem with the help of several lemmas.

Lemma 4.10

Let X be a set. A family $\mathcal{B} \subseteq \mathscr{P}(X)$ of subsets of X is a σ-algebra if and only if $\mathcal{B}$ is both a π-system and a λ-system.

Proof. The definition of a λ-system is the same as the definition of a σ-algebra, except that one is only allowed to take unions of *disjoint* sets in the definition of a λ-system. It therefore suffices to check that being a π-system as well allows for taking countable unions of not necessarily disjoint sets.

Suppose $E_1, E_2, \dots \in \mathcal{B}$. Define $E_1' = E_1$, $E_2' = E_2 \setminus E_1$, $\dots$, $E_n' = E_n \setminus \bigcup_{i=1}^{n-1} E_i$. Then $E_1', E_2', \dots$ are pairwise disjoint and satisfy $\bigsqcup_{n\in\mathbb{N}} E_n' = \bigcup_{n\in\mathbb{N}} E_n$, so it suffices to check that $E_n' \in \mathcal{B}$ for each $n \in \mathbb{N}$. But this is clear upon rewriting $E_n' = E_n \cap \bigcap_{i=1}^{n-1} E_i^c$, since $E_i^c = X \setminus E_i \in \mathcal{B}$ (by the axioms of a λ-system) and a finite intersection of sets from $\mathcal{B}$ belongs to $\mathcal{B}$ (by the axioms of a π-system). □

Lemma 4.11

Let X be a set, and suppose $(\mathcal{L}_i)_{i\in I}$ is a collection of λ-systems $\mathcal{L}_i \subseteq \mathscr{P}(X)$. Then $\bigcap_{i\in I} \mathcal{L}_i$ is a λ-system.

Proof. The proof is the same as the proof of Proposition 2.4, except we only allow disjoint unions. □

Definition 4.12

Let X be a set and $\mathcal{S} \subseteq \mathscr{P}(X)$ a family of subsets of X. The *λ-system generated by $\mathcal{S}$* is the smallest λ-system containing $\mathcal{S}$:

$$\lambda(\mathcal{S}) = \bigcap\{\mathcal{L} \subseteq \mathscr{P}(X) : \mathcal{L} \text{ is a } \lambda\text{-system}, \mathcal{S} \subseteq \mathcal{L}\}.$$

Lemma 4.13

Let X be a set, and let $\mathcal{P} \subseteq \mathscr{P}(X)$ be a π-system. The λ-system $\lambda(\mathcal{P})$ generated by $\mathcal{P}$ is a σ-algebra.

Proof. By Lemma 4.10, it suffices to show that $\lambda(\mathcal{P})$ is a π-system.

Claim 1. For any set $A \in \lambda(\mathcal{P})$, the family $\mathcal{L}_A := \{B \subseteq X : A \cap B \in \lambda(\mathcal{P})\}$ is a λ-system.

Since $A \in \lambda(\mathcal{P})$, we see that $X \in \mathcal{L}_A$.

Suppose $B_1, B_2 \in \mathcal{L}_A$ and $B_1 \subseteq B_2$. Then

$$A \cap (B_2 \setminus B_1) = \underbrace{(A \cap B_2)}_{\in\lambda(\mathcal{P})} \setminus \underbrace{(A \cap B_1)}_{\in\lambda(\mathcal{P})} \in \lambda(\mathcal{P}),$$

so $B_2 \setminus B_1 \in \mathcal{L}_A$.

Finally, suppose $B_1 \subseteq B_2 \subseteq \cdots \in \mathcal{L}_A$. Then $A \cap \bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}}(A \cap B_n) \in \lambda(\mathcal{P})$, so $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{L}_A$.

This proves the claim.

Claim 2. For any $A \in \lambda(\mathcal{P})$ and any $B \in \mathcal{P}$, we have $A \cap B \in \lambda(\mathcal{P})$.

This follows from Claim 1: the family $\mathcal{L}_B$ is a λ-system, and $\mathcal{P} \subseteq \mathcal{L}_B$ by the definition of a π-system, so $\mathcal{L}_B \supseteq \lambda(\mathcal{P}) \ni A$.

Let $A, B \in \lambda(\mathcal{P})$. The family $\mathcal{L}_A$ is a λ-system (by Claim 1) containing $\mathcal{P}$ (by Claim 2), so $\mathcal{L}_A \supseteq \lambda(\mathcal{P}) \ni B$. Hence, $A \cap B \in \lambda(\mathcal{P})$. □

Now we can complete the proof of the π-λ theorem.

Proof of π-λ Theorem (Theorem 4.9). Let $\mathcal{P}$ be a π-system, $\mathcal{L}$ a λ-system, and suppose $\mathcal{P} \subseteq \mathcal{L}$. On the one hand, by Lemma 4.13, the λ-system $\lambda(\mathcal{P})$ generated by $\mathcal{P}$ is a σ-algebra, so $\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P})$. On the other hand, $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, so $\lambda(\mathcal{P}) \subseteq \mathcal{L}$. Combining these two observations completes the proof. □

Corollary 4.14: Uniqueness of Lebesgue–Stieltjes Measures

Suppose μ and ν are locally finite Borel measures on $\mathbb{R}$ with the same distribution function $F_\mu = F_\nu = F$. Then $\mu = \nu$.

Proof. Let $\mathcal{P}$ be the π-system $\mathcal{P} = \{(a,b] : a,b \in \mathbb{R}\}$ of half-open intervals. Define

$$\mathcal{L} = \{E \in \text{Borel}(\mathbb{R}) : \mu(E \cap (-N,N]) = \nu(E \cap (-N,N]) \text{ for every } N \in \mathbb{N}\}.$$

Claim 1. $\mathcal{L}$ is a λ-system

For every $N \in \mathbb{N}$,

$$\mu((-N,N]) = (\mu((-N,0]) + \mu((0,N])) = F(N) - F(-N).$$

The same holds for ν, so $\mathbb{R} \in \mathcal{L}$.

Suppose $E \in \mathcal{L}$, and let $N \in \mathbb{N}$. By additivity of μ and ν, we have

$$\begin{aligned}\mu(E^c \cap (-N,N]) &= \mu((-N,N]) - \mu(E \cap (-N,N]) \\ &= \nu((-N,N]) - \nu(E \cap (-N,N]) \\ &= \nu(E^c \cap (-N,N]),\end{aligned}$$

so $E^c \in \mathcal{L}$.

Finally, $\mathcal{L}$ is closed under countable disjoint unions as a consequence of countable additivity of the measures μ and ν.

Claim 2. $\mathcal{P} \subseteq \mathcal{L}$

The sets $(-N,N]$ belong to $\mathcal{P}$, which is a π-system, so it suffices to prove $\mu(P) = \nu(P)$ for every $P \in \mathcal{P}$. Let $P = (a,b] \in \mathcal{P}$. If $b \leq a$, then $P = \emptyset$, so $\mu(P) = \nu(P) = 0$. Suppose $a < b$. If $a \leq b$, then $\mu(P) = F(b) - F(a) = \nu(P)$.

By the π-λ theorem, $\sigma(\mathcal{P}) \subseteq \mathcal{L}$. But $\mathcal{P}$ generates the Borel σ-algebra (we essentially showed this in the proof of Proposition 2.11), so $\mathcal{L} = \text{Borel}(\mathbb{R})$. Hence, applying continuity from below, we have

$$\mu(E) = \lim_{N\to\infty} \mu(E \cap (-N,N]) = \lim_{N\to\infty} \nu(E \cap (-N,N]) = \nu(E)$$

for every $E \in \text{Borel}(\mathbb{R})$. □

Example 4.15

The locally finite condition cannot be dropped from Corollary 4.14. As an example, define a measure $\mu : \text{Borel}(\mathbb{R}) \to [0,\infty]$ by

$$\mu(B) = \#(B \cap \mathbb{Q}),$$

and let $\nu = \infty \cdot \lambda$, where λ is the Lebesgue measure. (We will construct λ later in this chapter, but for now, take it as a given that the Lebesgue measure exists.) Every non-empty interval in $\mathbb{R}$ contains infinitely many rational points, so

$$F_\mu(x) = \begin{cases} \infty, & \text{if } x > 0; \\ 0, & \text{if } x = 0; \\ -\infty, & \text{if } x < 0. \end{cases}$$

Similarly, every non-empty interval in $\mathbb{R}$ has positive Lebesgue measure, so multiplying by ∞, the measure ν has the same distribution function $F_\nu = F_\mu$. However, μ and ν are not the same measure, since, for instance, $\mu(\mathbb{R} \setminus \mathbb{Q}) = 0$, while $\nu(\mathbb{R} \setminus \mathbb{Q}) = \infty$, and $\mu(\{0\}) = 1$, while $\nu(\{0\}) = 0$.

2. Half-Open Intervals

Now we begin the construction of Lebesgue–Stieltjes measures. Let us define some basic objects that we will work with for the construction.

Definition 4.16

A *left-open, right-closed interval* is a set of the form

- $\mathbb{R}$,
- $\emptyset$,
- $(a, b]$ with $a, b \in \mathbb{R}$, $a < b$,
- $(-\infty, b]$ with $b \in \mathbb{R}$, or
- (a, ∞) with $a \in \mathbb{R}$.

The intersection of two left-open, right-closed intervals is a left-open, right-closed intervals, and the complement of a left-open, right-closed interval is either a left-open, right-closed interval or a disjoint union of two left-open, right-closed intervals (see Figure 4.1).

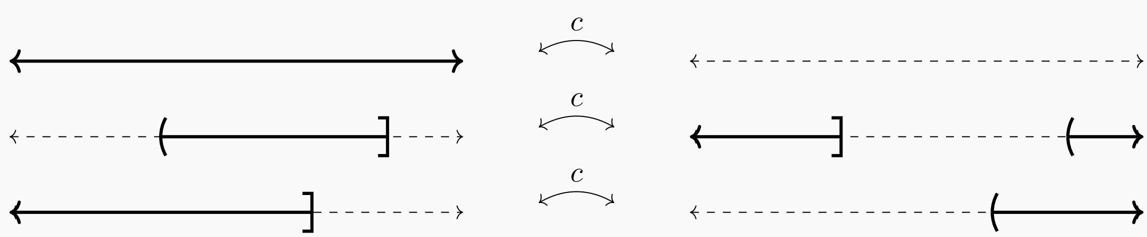


Figure 4.1. The complement of a left-open, right-closed intervals is either another left-open, right-closed interval or the union of two.

Therefore, the family of left-open, right-closed intervals forms a *semi-algebra* on $\mathbb{R}$. We recall the definition below.

Definition 4.17

Let X be a set. A family $\mathcal{S} \subseteq \mathscr{P}(X)$ of subsets of X is a *semi-algebra* if

- $\emptyset, X \in \mathcal{S}$;
- if $S, T \in \mathcal{S}$, then $S \cap T \in \mathcal{S}$;
- if $S \in \mathcal{S}$, then $X \setminus S = \bigsqcup_{i=1}^{n} C_i$ for some $C_1, \dots, C_n \in \mathcal{S}$.

Proposition 4.18

Let $\mathcal{S}$ be a semi-algebra on a set X. Then

$$\mathcal{A} = \left\{ \bigsqcup_{i=1}^{n} S_i : n \in \mathbb{N}, S_1, \dots, S_n \in \mathcal{S} \right\}$$

is an algebra.

Proof. The family $\mathcal{A}$ contains X, since $X \in \mathcal{S}$.

Let us check that $\mathcal{A}$ is closed under intersections. Let $A, B \in \mathcal{A}$, say $A = \bigsqcup_{i=1}^{n} S_i$ and $B = \bigsqcup_{j=1}^{m} T_j$ with $S_1, \dots, S_n \in \mathcal{S}$ disjoint and $T_1, \dots, T_m \in \mathcal{S}$ another disjoint collection. Then the sets $\{S_i \cap T_j : 1 \leq i \leq n, 1 \leq j \leq m\}$ are disjoint, so

$$A \cap B = \bigsqcup_{i,j} \underbrace{(S_i \cap T_j)}_{\in \mathcal{S}} \in \mathcal{A}.$$

Finally, we check that $\mathcal{A}$ is closed under complements. Let $A = \bigsqcup_{i=1}^n S_i \in \mathcal{A}$. For each $i \in \{1, \ldots, n\}$, we can write $X \setminus S_i = \bigsqcup_{j=1}^{m_i} C_{i,j}$ for some $C_{i,1}, \ldots, C_{i,m_i} \in \mathcal{S}$. Thus,

$$X \setminus A = \bigcap_{i=1}^n (X \setminus S_i) = \bigcap_{i=1}^n \left(\bigsqcup_{j=1}^{m_i} C_{i,j} \right)$$

is an intersection of elements of $\mathcal{A}$ so belongs to $\mathcal{A}$. □

Notation. We will denote the algebra generated by the semi-algebra of left-open, right-closed intervals by

$$\mathcal{A}_{int} = \left\{ \bigsqcup_{i=1}^n I_i : n \in \mathbb{N}, I_i \text{ is a left-open, right-closed interval} \right\}.$$

Note that the σ-algebra generated by $\mathcal{A}_{int}$ is the Borel σ-algebra.

3. Premeasures and Outer Measures

We will begin the construction of the Lebesgue–Stieltjes measure associated to a distribution function F by assigning a measure to each element of $\mathcal{A}_{int}$.

Definition 4.19

Let X be a set and $\mathcal{A} \subseteq \mathscr{P}(X)$ an algebra. A *premeasure* is a function $\mu_0 : \mathcal{A} \to [0, \infty]$ such that

- $\mu_0(\emptyset) = 0$;
- if $(A_n)_{n \in \mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{A}$ and $A = \bigsqcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$, then $\mu_0(A) = \sum_{n=1}^{\infty} \mu_0(A_n)$.

Note that if $\mathcal{A}$ is a σ-algebra, then a premeasure on $\mathcal{A}$ is the same thing as a measure. More generally, if $\mu : \mathcal{B} \to [0, \infty]$ is a measure on a measurable space $(X, \mathcal{B})$ and $\mathcal{A} \subseteq \mathcal{B}$ is an algebra, then $\mu_0 = \mu|_{\mathcal{A}}$ defines a premeasure on $\mathcal{A}$.

The next proposition shows that we can associate to an increasing right-continuous function F a premeasure on the algebra $\mathcal{A}_{int}$ generated by left-open, right-closed intervals. We will see afterwards how to extend this premeasure to a measure on the Borel subsets of $\mathbb{R}$.

Proposition 4.20

Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right-continuous. Define a function $\mu_{F,0} : \mathcal{A}_{int} \to [0, \infty]$ by

$$\mu_{F,0} \left(\bigsqcup_{i=1}^n (a_i, b_i] \right) = \sum_{i=1}^n (F(b_i) - F(a_i)).$$

Then $\mu_{F,0}$ is a premeasure on $\mathcal{A}_{int}$.

Proof. We will first show that $\mu_{F,0}$ is a well-defined function on $\mathcal{A}_{int}$ and then prove that it is a premeasure.

Step 1. $\mu_{F,0}$ is well-defined

Every element of $\mathcal{A}_{int}$ can always be written uniquely in the form

$$\bigsqcup_{i=1}^{n}(a_i, b_i]$$

with $-\infty \leq a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n \leq \infty$. Indeed, after writing the intervals in increasing order, if $a_{i+1} = b_i$ for some i, then the intervals $(a_i, b_i]$ and $(a_{i+1}, b_{i+1}]$ can be merged into the single interval $(a_i, b_{i+1}]$ (see Figure 4.2). This process of merging leaves the expression for $\mu_{F,0}$ unchanged, since if $b_i = a_{i+1}$, we have a telescoping phenomenon

$$(F(b_i) - F(a_i)) + (F(b_{i+1}) - F(a_{i+1})) = F(b_{i+1}) - F(a_i).$$

Thus, the formula for $\mu_{F,0}$ gives the same value for every possible expression of $A \in \mathcal{A}_{int}$ as a disjoint union of left-open, right-closed intervals.

Figure 4.2. Merging the red and blue intervals into a single left-open, right-closed interval (purple) leaves the value $\sum_i (F(b_i) - F(a_i))$ unchanged.

Step 2. If $(a, b] = \bigsqcup_{i=1}^{\infty}(a_i, b_i]$, then $\mu_{F,0}((a, b]) \leq \sum_{i=1}^{\infty} \mu_{F,0}((a_i, b_i])$.

Let $\delta > 0$ and let $\varepsilon_i > 0$ for $i \in \mathbb{N}$. Then $[a + \delta, b]$ is a closed interval covered by the union of open intervals $\bigcup_{i=1}^{\infty}(a_i, b_i + \varepsilon_i)$. By the Heine–Borel theorem (compactness of closed intervals in $\mathbb{R}$), there is a finite subcover $i_1, \ldots, i_n$ such that $[a + \delta, b] \subseteq \bigcup_{j=1}^{n}(a_{i_j}, b_{i_j} + \varepsilon_{i_j})$. Therefore, $(a + \delta, b] \subseteq \bigcup_{j=1}^{n}(a_{i_j}, b_{i_j} + \varepsilon_{i_j}]$, so by Step 1,

$$\mu_{F,0}\left((a + \delta, b]\right) \leq \sum_{j=1}^{n} \mu_{F,0}\left((a_{i_j}, b_{i_j} + \varepsilon_{i_j}]\right) \leq \sum_{i=1}^{\infty} \mu_{F,0}((a_i, b_i + \varepsilon_i]).$$

Letting $\delta \to 0$,

$$\lim_{\delta \to 0^+} \mu_{F,0}\left((a + \delta, b]\right) = F(b) - \lim_{\delta \to 0^+} F(a + \delta) = F(b) - F(a) = \mu_{F,0}((a, b]),$$

since F is right-continuous. Similarly, given $\varepsilon > 0$, we can take ε_i sufficiently small so that $\sum_{i=1}^{\infty} \mu_{F,0}((a_i, b_i + \varepsilon_i]) \leq \sum_{i=1}^{\infty} \mu_{F,0}((a_i, b_i]) + \varepsilon$. Then letting $\varepsilon \to 0$ proves the desired inequality.

Step 3. $\mu_{F,0}$ is countably additive.

Suppose $(A_n)_{n \in \mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{A}_{int}$, and $A = \bigsqcup_{n \in \mathbb{N}} A_n \in \mathcal{A}_{int}$. Each of the sets A_n, belonging to the algebra $\mathcal{A}_{int}$, can be written in the form $A_n = \bigsqcup_{m=1}^{M_n} S_{n,m}$, where $S_{n,m}$ is a left-open, right-closed interval. We know $\mu_{F,0}(A_n) = \sum_{m=1}^{M_n} \mu_{F,0}(S_{n,m})$ by definition. Replacing $(A_n)_{n \in \mathbb{N}}$ by $(S_{n,m})_{n \in \mathbb{N}, 1 \leq m \leq M_n}$, we may assume from the start that A_n is a left-open, right-closed interval for each $n \in \mathbb{N}$. Since $A \in \mathcal{A}$, we may also write $A = \bigsqcup_{m=1}^{M} S_m$ for some left-open, right-closed intervals S_m. Then

$$
\begin{aligned}
\mu_{F,0}(A) &= \sum_{m=1}^{M} \mu_{F,0}(S_m) && \text{(definition of } \mu_{F,0}\text{)} \\
&= \sum_{m=1}^{M} \mu_{F,0}\left(\bigsqcup_{n\in\mathbb{N}} (S_m \cap A_n)\right) && (A = \sqcup_{n\in\mathbb{N}} A_n) \\
&\leq \sum_{n,m} \mu_{F,0}(S_m \cap A_n) && \text{(Step 2)} \\
&= \sum_{n=1}^{\infty} \mu_{F,0}\left(\bigsqcup_{m=1}^{M} (S_m \cap A_n)\right) && \text{(definition of } \mu_{F,0}\text{)} \\
&= \sum_{n=1}^{\infty} \mu_{F,0}(A_n) && \text{(definition of } \mu_{F,0}\text{)}
\end{aligned}
$$

On the other hand, for $N \in \mathbb{N}$,

$$\sum_{n=1}^{N} \mu_{F,0}(A_n) = \mu_{F,0}\left(\bigsqcup_{n=1}^{N} A_n\right) \leq \mu_{F,0}(A),$$

so taking a limit as $N \to \infty$, $\sum_{n=1}^{\infty} \mu_{F,0}(A_n) \leq \mu_{F,0}(A)$.

□

The next stage in the construction is to extend the premeasure $\mu_{F,0}$ to an *outer measure* defined on all subsets of $\mathbb{R}$.

Definition 4.21

Let X be a set. A function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ is an *outer measure* if

- $\mu^*(\emptyset) = 0$;
- MONOTONE: if $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$; and
- COUNTABLY SUBADDITIVE: for any sequence of sets $(A_n)_{n\in\mathbb{N}}$, one has $\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

A premeasure can always be extended to an outer measure, as shown by the following proposition.

Proposition 4.22

Let $\mathcal{A}$ be an algebra on a set X, and suppose $\mu_0 : \mathcal{A} \to [0, \infty]$ is a premeasure. Then

$$\mu^*(E) = \inf\left\{\sum_{n=1}^{\infty} \mu_0(A_n) : E \subseteq \bigcup_{n\in\mathbb{N}} A_n, A_n \in \mathcal{A}\right\}$$

defines an outer measure on X with $\mu^*|_{\mathcal{A}} = \mu_0$.

PROOF. Let us check the properties one at a time.

First, $\emptyset \in \mathcal{A}$, so $\mu^*(\emptyset) \leq \mu_0(\emptyset) = 0$.

Next, suppose $A \subseteq B$. Then any set containing B also contains A, so the expression defining $\mu^*(A)$ involves an infimum over a larger collection than the expression defining $\mu^*(B)$. Hence, $\mu^*(A) \leq \mu^*(B)$.

Now let us prove countable subadditivity. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of subsets of X. If

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

there is nothing to show, so assume

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

Let $\varepsilon > 0$. For each n, let $(A_{n,k})_{k\in\mathbb{N}}$ be a sequence of elements of $\mathcal{A}$ such that $A_n \subseteq \bigcup_{k\in\mathbb{N}} A_{n,k}$ and

$$\sum_{k=1}^{\infty} \mu_0(A_{n,k}) < \mu^*(A_n) + \frac{\varepsilon}{2^n}.$$

Then $(A_{n,k})_{n,k\in\mathbb{N}}$ is a countable family of elements of the algebra $\mathcal{A}$, and $\bigcup_{n\in\mathbb{N}} A_n \subseteq \bigcup_{n,k\in\mathbb{N}} A_{n,k}$, so

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n,k} \mu_0(A_{n,k}) \leq \sum_{n=1}^{\infty}\left(\mu^*(A_n) + \frac{\varepsilon}{2^n}\right) = \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.$$

Letting $\varepsilon \to 0$ establishes countable subadditivity.

Finally, let us show $\mu^*|_{\mathcal{A}} = \mu_0$. Let $A \in \mathcal{A}$. Then by definition $\mu^*(A) \leq \mu_0(A)$. It remains to show $\mu^*(A) \geq \mu_0(A)$. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ such that $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$. Define a new sequence $(B_n)_{n\in\mathbb{N}}$ by $B_1 = A \cap A_1$ and $B_n = (A \cap A_n) \setminus (A_1 \cup \cdots \cup A_{n-1})$. Since $\mathcal{A}$ is an algebra, the sets B_n belong to $\mathcal{A}$. Moreover, $(B_n)_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint sets whose union is $A \in \mathcal{A}$, so

$$\mu_0(A) = \sum_{n=1}^{\infty} \mu_0(B_n) \leq \sum_{n=1}^{\infty} (\mu_0(B_n) + \mu_0(A_n \setminus B_n)) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

Taking an infimum over all such collections $(A_n)_{n\in\mathbb{N}}$ gives the desired inequality $\mu_0(A) \leq \mu^*(A)$. □

The outer measure μ_F^* obtained from the premeasure $\mu_{F,0}$ is not in general a measure on $\mathscr{P}(X)$. The problem is that, while μ_F^* is *subadditive*, it may fail to be additive. In order to obtain a measure, we restrict to the sets with better additive behavior. As a first step towards the correct notion of μ^*-measurability, any μ^*-measurable set $E \subseteq X$ should certainly satisfy $\mu^*(X) = \mu^*(E) + \mu^*(X \setminus E)$. Unfortunately, this is not sufficient to obtain a suitable notion of measurability, as the following example shows.

Example 4.23

Let $X = \{1, 2, 3\}$, and let $\mu^* : \mathscr{P}(X) \to \{0, 1, 2\}$ be the function defined by $\mu^*(\emptyset) = 0$, $\mu^*(X) = 2$, and $\mu^*(E) = 1$ for nonempty property subsets $\emptyset \neq E \subsetneq X$. It is easy to check that μ^* is an outer measure and moreover that $\mu^*(E) + \mu^*(X \setminus E) = 2 = \mu^*(X)$ for every subset $E \subseteq X$. But $\mu^*(\{1, 2\}) = 1$, while $\mu^*(\{1\}) + \mu^*(\{2\}) = 2$, so μ^* is not a measure.

In order to correct for the insufficiency of $\mu^*(X) = \mu^*(E) + \mu^*(X \setminus E)$ as a condition for measurability, we impose the stronger condition that E partitions *every* subset of X in an analogous manner.

DEFINITION 4.24

Let μ^* be an outer measure on a set X. A set $E \subseteq X$ is *μ^*-measurable* if for every $A \subseteq X$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E). \tag{4.1}$$

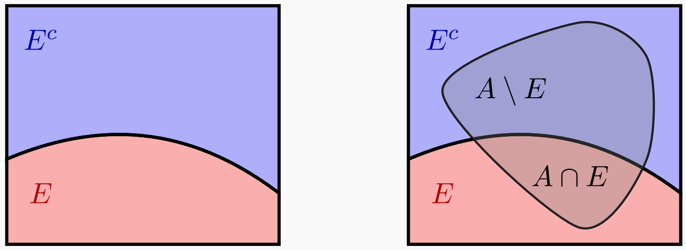


FIGURE 4.3. A set is μ^*-measurable if it partitions every set into two pieces that can be measured separately and then recombined.

REMARK. The notion of measurability defined here is due to Carathéodory, and (4.1) is often referred to as *Carathéodory's criterion for measurability*. Outer measures are subadditive, so (4.1) is equivalent to the *a priori* weaker inequality

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

THEOREM 4.25: CARATHÉODORY'S THEOREM

Let μ^* be an outer measure on a set X. Let $\mathcal{M} \subseteq \mathscr{P}(X)$ be the family of μ^*-measurable sets. Then $\mathcal{M}$ is a σ-algebra, and $\mu^*|_{\mathcal{M}}$ is a complete measure.

PROOF. We break the proof into several steps.

CLAIM 1. $X \in \mathcal{M}$.

Given $A \subseteq X$, we have $\mu^*(A \cap X) + \mu^*(A \setminus X) = \mu^*(A) + \mu^*(\emptyset) = \mu^*(A)$.

CLAIM 2. $\mathcal{M}$ is closed under complementation.

Rewriting $A \setminus E = A \cap E^c$, the measurability condition (4.1) is symmetric in E and E^c.

CLAIM 3. $\mathcal{M}$ is closed under finite unions.

Suppose $E, F \in \mathcal{M}$, and let $A \subseteq X$. The sets E and F together partition A into four parts that we will subsequently recombine (see Figure 4.4). Using subadditivity of μ^*, we have

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

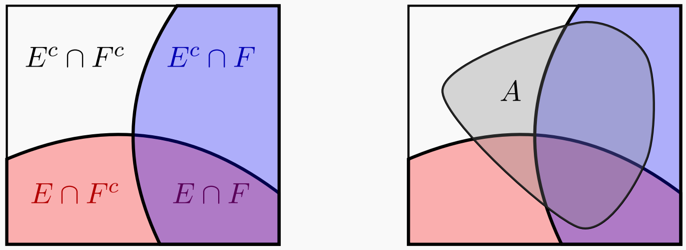


FIGURE 4.4. Partition of A into four pieces determined by E and F.

Now, by μ^*-measurability of F, the right hand side of the inequality can be rewritten as

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

which is in turn equal to $\mu^*(A)$ by μ^*-measurability of E. Thus,

$$\mu^*(A) \geq \mu^*(A \cap (E \cup F)) + \mu^*(A \setminus (E \cup F)),$$

so $E \cup F \in \mathcal{M}$.

Claims 1–3 show that $\mathcal{M}$ is an algebra. The next claim upgrades $\mathcal{M}$ to a σ-algebra and proves that $\mu^*|_{\mathcal{M}}$ is a measure.

CLAIM 4. $\mathcal{M}$ is closed under countable disjoint unions, and $\mu^*|_{\mathcal{M}}$ is countably additive.

Suppose $(E_n)_{n\in\mathbb{N}}$ be a sequence of pairwise disjoint sets in $\mathcal{M}$, and let $E = \bigsqcup_{n\in\mathbb{N}} E_n$. Let $A \subseteq X$. As in Claim 3, we want to show

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \setminus E). \tag{4.2}$$

If $\mu^*(A) = \infty$, there is nothing to check, so assume $\mu^*(A) < \infty$. Let $F_N = \bigsqcup_{n=1}^N E_n$. By induction, we have

$$\mu^*(A \cap F_N) = \sum_{n=1}^N \mu^*(A \cap E_n).$$

Hence, by countable subadditivity of μ^*,

$$\mu^*(A \cap E) \leq \sum_{n=1}^\infty \mu^*(A \cap E_n) = \lim_{N\to\infty} \mu^*(A \cap F_N).$$

For fixed $N \in \mathbb{N}$, $F_N \in \mathcal{M}$ by Claim 3, so

$$\mu^*(A) = \mu^*(A \cap F_N) + \mu^*(A \setminus F_N) \geq \mu^*(A \cap F_N) + \mu^*(A \setminus E).$$

Taking a limit as $N \to \infty$ gives (4.2).

Note that we actually proved the stronger inequality

$$\mu^*(A) \geq \sum_{n=1}^\infty \mu^*(A \cap E_n) + \mu^*(A \setminus E).$$

Taking $A = E$ establishes countable additivity of μ^*.

Finally, we check that $(X, \mathcal{M}, \mu^*|_{\mathcal{M}})$ is complete.

Claim 5. If $N \subseteq X$ and $\mu^*(N) = 0$, then $N \in \mathcal{M}$.

Let $A \subseteq X$. Then by monotonicity,

$$\mu^*(A \cap N) + \mu^*(A \setminus N) \leq \mu^*(N) + \mu^*(A) = \mu^*(A).$$

□

The last remaining piece to tie everything together is relating the σ-algebra $\mathcal{M}$ in Carathéodory's theorem to the algebra on which the premeasure μ_0 was defined.

Lemma 4.26

Let $\mathcal{A}$ be an algebra on a set X. Let $\mu_0 : \mathcal{A} \to [0, \infty]$ be a premeasure, and let μ^* be the outer measure extending μ_0 as in Proposition 4.22. Then every element of $\mathcal{A}$ is μ^*-measurable.

Proof. Let $A \in \mathcal{A}$, and let $B \subseteq X$ be an arbitrary set. We want to show

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \setminus A). \tag{4.3}$$

Let $(A_n)_{n\in\mathbb{N}}$ be family of elements of $\mathcal{A}$ such that $B \subseteq \bigcup_{n\in\mathbb{N}} A_n$. Then

$$\sum_{n=1}^{\infty} \mu_0(A_n) = \sum_{n=1}^{\infty} (\mu_0(A_n \cap A) + \mu_0(A_n \setminus A)) = \sum_{n=1}^{\infty} \mu_0(A_n \cap A) + \sum_{n=1}^{\infty} \mu_0(A_n \setminus A).$$

The union $\bigcup_{n\in\mathbb{N}}(A_n \cap A)$ contains $B \cap A$, and similarly, $\bigcup_{n\in\mathbb{N}}(A_n \setminus A)$ contains $B \setminus A$, so by the definition of the outer measure μ^*,

$$\sum_{n=1}^{\infty} \mu_0(A_n) \geq \mu^*(B \cap A) + \mu^*(B \setminus A).$$

Taking an infimum over all such collections $(A_n)_{n\in\mathbb{N}}$ gives (4.3). □

Putting everything together, we get the following theorem.

Theorem 4.27: Hahn–Kolmogorov Extension Theorem

Let $\mathcal{A} \subseteq \mathscr{P}(X)$ be an algebra on a set X, and let $\mu_0 : \mathcal{A} \to [0, \infty]$ be a premeasure. Then μ_0 extends to a complete measure $\mu : \mathcal{M} \to [0, \infty]$ defined on a σ-algebra $\mathcal{M} \supseteq \sigma(\mathcal{A})$. Moreover, if μ_0 is σ-finite, then the extension of μ_0 to $\sigma(\mathcal{A})$ is unique, and μ is the completion of this unique extension.

Proof. By Proposition 4.22, μ_0 extends to an outer measure $\mu^* : \mathscr{P}(X) \to [0, \infty]$. Let $\mathcal{M}$ be the σ-algebra of μ^*-measurable sets, and let $\mu = \mu^*|_{\mathcal{M}}$. Then μ is a complete measure by Carathéodory's theorem (Theorem 4.25), and by Lemma 4.26, $\mathcal{A} \subseteq \mathcal{M}$.

Uniqueness in the σ-finite case is a consequence of the π-λ theorem and follows on exactly the same lines as the proof of Corollary 4.14. □

Remark. When μ_0 is not σ-finite, it may have several different extensions to $\sigma(\mathcal{A})$. The outer measure construction is the maximal such extension in the sense that given any other extension $\nu : \sigma(\mathcal{A}) \to [0, \infty]$ of μ_0, one has $\nu(E) \leq \mu(E)$ for every $E \in \sigma(\mathcal{A})$. We will return to this subject in the context of product measures later in the course, where it will sometimes be useful to work with a different extension of a premeasure than the one obtained by Carathéodory's theorem.

Corollary 4.28: Existence of Lebesgue–Stieltjes Measures

Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right-continuous with $F(0) = 0$. There exists a σ-algebra $\mathcal{M}_F$ containing the Borel subsets of $\mathbb{R}$ and a complete measure $\mu_F : \mathcal{M}_F \to [0, \infty]$ with distribution function $F_{\mu_F} = F$.

Proof. Let $\mu_{F,0}$ be the premeasure on $\mathcal{A}_{int}$ given by Proposition 4.20. Then the extension μ_F given by the Hahn–Kolmogorov extension theorem (Theorem 4.27) is a complete measure with distribution function F. □

4. Lebesgue Measure

Definition 4.29

The *Lebesgue measure* on $\mathbb{R}$ is the Lebesgue–Stieltjes measure associated to the distribution function $F(x) = x$.

Proposition 4.30

Let λ be the Lebesgue measure on $\mathbb{R}$, and let $\mathcal{M}$ be the σ-algebra of Lebesgue measurable sets.

(1) translation-invariance: $\lambda(E + t) = \lambda(E)$ for every $E \in \mathcal{M}$ and $t \in \mathbb{R}$;
(2) reflection-invariance: $\lambda(-E) = \lambda(E)$ for every $E \in \mathcal{M}$;
(3) dilation property: $\lambda(tE) = |t|\lambda(E)$ for every $E \in \mathcal{M}$ and $t \in \mathbb{R}$;

Proof. Compute the distribution function of the transformed measure and apply uniqueness of Lebesgue–Stieltjes measures (Corollary 4.14). We leave the details as an exercise. □

Using the translation-invariance property of the Lebesgue measure, we can prove the existence of a non-measurable set.

Theorem 4.31

There exists a Lebesgue non-measurable subset of $\mathbb{R}$.

Proof. See Theorem 1.12. □

Remark. The axiom of choice plays a crucial role in the construction of Vitali sets. Using the set-theoretic notion of an *inaccessible cardinal*, Robert Solovay constructed a model of set theory under the ZF axioms without choice in which every subset of $\mathbb{R}$ is Lebesgue measurable [**Sol70**].

5. Cantor Measure

Another interesting example of a Borel measure on the real line is the "uniform" measure on the middle-thirds Cantor set, which we will construct now. Recall that the Cantor set $C \subseteq [0, 1]$ is obtained by starting with the full interval $[0, 1]$ and iteratively removing the middle third of each remaining interval at each step. We can therefore write $[0, 1] \setminus C = \bigcup_{n=0}^{\infty} \bigcup_{k=0}^{2^n-1} I_{n,k}$, where $I_{n,0}, \dots, I_{n,2^{n-1}}$ is an enumeration of the removed intervals of length $3^{-(n+1)}$ in increasing order. How should we define the distribution function for a uniform measure on C? Well, after step $n = 0$, where we remove the interval $(\frac{1}{3}, \frac{2}{3})$, we have half of of the Cantor set to the left of this interval and the other half to the right, so the distribution function should take the value $\frac{1}{2}$ on the entirety of this interval. Arguing similarly, the distribution function should take the value $\frac{2k+1}{2^{n+1}}$ on the interval $I_{n,k}$ for each $n \geq 0$ and $k \in \{0, 1, \ldots, 2^n - 1\}$. Since the union of intervals $\bigcup_{n=0}^{\infty} \bigcup_{k=0}^{2^n-1} I_{n,k}$ is dense in $[0, 1]$, there is a unique way of interpolating between the values on the intervals $I_{n,k}$ in order to obtain a continuous function. We call this continuous function the *Cantor function* and its associated Lebesgue–Stieltjes measure the *Cantor measure*.

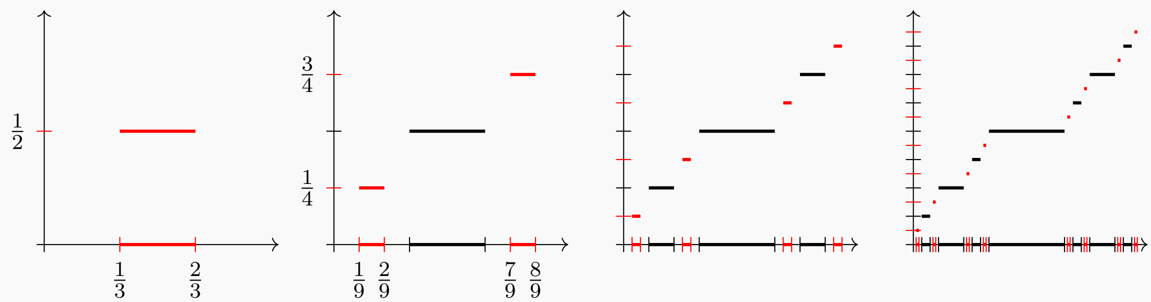


FIGURE 4.5. Iterative construction of the Cantor function

A more explicit description of the Cantor set is the collection of numbers in the interval $[0, 1]$ whose binary expansion consists entirely of the digits 0 and 2. From this description of the Cantor set, we can also obtain a formula for the Cantor function, namely

$$c(x) = \begin{cases} \sum_{j=1}^{\infty} \frac{a_j/2}{2^j}, & \text{if } x = \sum_{j=1}^{\infty} \frac{a_j}{3^j} \in C; \\ \sup_{y<x, y \in C} c(y), & \text{if } x \notin C. \end{cases}$$

The Cantor measure has a surprising combination of properties. Suppose you have a perfectly fair coin to flip, and you record the sequence of heads and tails as you flip the coin repeatedly. Recording heads as the digit 2 and tails as the digit 0, this sequence of coin flips produces a random element of the Cantor set in terms of its base 3 expansion. The distribution of this random element of the Cantor set is described by the Cantor measure.

The Cantor function is continuous, so the Cantor measure is a continuous probability measure. This is despite the fact that all of the mass of the Cantor measure is concentrated on the Cantor set, which is a set of Lebesgue measure zero! This makes the Cantor measure an example of what is called a *singular measure*, and as a result, even though we have given a reasonable probabilistic method for constructing Cantor-distributed random variables, the Cantor measure does not have a probability density function. That is, there is no function $f : [0, 1] \to [0, \infty)$ for which we can express $\mu_C((a, b]) = \int_a^b f(x)\, dx$. We will reencounter singular measures and deal with them systematically later in the course.

6. Regularity of Lebesgue–Stieltjes Measures

Built into the definition of Lebesgue–Stieltjes measures is the fact that they are locally finite, but it is not at all obvious that Lebesgue–Stieltjes measures should have additional regularity properties. However, from the outer measure construction, we can quickly deduce several useful and nontrivial properties of Lebesgue–Stieltjes measures.

Proposition 4.32

Let μ be a Lebesgue–Stieltjes measure on $\mathbb{R}$, and let $\mathcal{M}_\mu$ be the σ-algebra of μ-measurable sets.

(1) Let $E \subseteq \mathbb{R}$. The following are equivalent:

  (i) $E \in \mathcal{M}_\mu$;

  (ii) for any $\varepsilon > 0$, there exists a closed set F and an open set G such that $F \subseteq E \subseteq G$ and $\mu(G \setminus F) < \varepsilon$;

  (iii) there exists an F_σ set A and a G_δ set B such that $A \subseteq E \subseteq B$ and $\mu(B \setminus A) = 0$.

(2) OUTER REGULARITY: If $E \in \mathcal{M}_\mu$, then

$$\mu(E) = \inf \{\mu(U) : U \supseteq E \text{ open}\}.$$

(3) INNER REGULARITY: If $E \in \mathcal{M}_\mu$, then

$$\mu(E) = \sup \{\mu(K) : K \subseteq E \text{ compact}\}.$$

Proof. We will first prove outer regularity (2), then prove the measurability conditions (1), and end with inner regularity (3).

(2) Let $E \in \mathcal{M}_\mu$. By monotonicity of μ, it suffices to show $\mu(E) \geq \inf\{\mu(U) : U \supseteq E \text{ open}\}$. As usual, if $\mu(E) = \infty$, there is nothing to check, so assume $\mu(E) < \infty$, and let $\varepsilon > 0$. From the outer measure construction, we have

$$\mu(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu((a_n, b_n]) : E \subseteq \bigcup_{n\in\mathbb{N}} (a_n, b_n] \right\}.$$

Hence, there exists a family of left-open, right-closed intervals $((a_n, b_n])_{n\in\mathbb{N}}$ such that $E \subseteq \bigcup_{n\in\mathbb{N}}(a_n, b_n]$ and $\sum_{n=1}^{\infty} \mu((a_n, b_n]) < \mu(E) + \frac{\varepsilon}{2}$. For each $n \in \mathbb{N}$, let $\delta_n > 0$ such that $\mu((a_n, b_n + \delta_n)) < \mu((a_n, b_n]) + 2^{-(n+1)}\varepsilon$. Such δ_n exists by continuity of μ from above. Then for the open set $U = \bigcup_{n\in\mathbb{N}}(a_n, b_n + \delta_n)$, we have

$$\mu(U) \leq \sum_{n=1}^{\infty} \mu((a_n, b_n + \delta_n)) < \sum_{n=1}^{\infty}(\mu((a_n, b_n]) + 2^{-(n+1)}\varepsilon) < \mu(E) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \mu(E) + \varepsilon.$$

But $\varepsilon > 0$ was arbitrary, so we are done.

(1) We will prove the chain of implications (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (i).

(i) $\Longrightarrow$ (ii). Suppose $E \in \mathcal{M}_\mu$, and let $\varepsilon > 0$. Let $E_n = E \cap (n, n+1]$ for $n \in \mathbb{Z}$. By (2), there exists an open set $G_n \supseteq E_n$ such that $\mu(G_n) < \mu(E_n) + 2^{-|n|}\frac{\varepsilon}{6}$. Let $G = \bigcup_{n\in\mathbb{Z}} G_n$. Then G is open, $E \subseteq G$, and

$$\mu(G \setminus E) \leq \sum_{n=-\infty}^{\infty} \mu(G_n \setminus E_n) < \frac{\varepsilon}{2}.$$

Applying the same argument to E^c, we find an open set $U \subseteq \mathbb{R}$ such that $E^c \subseteq U$ and $\mu(U \setminus E^c) < \frac{\varepsilon}{2}$. Let $F = U^c$. Then F is closed, $F \subseteq E$, and $\mu(E \setminus F) = \mu(U \setminus E^c) < \frac{\varepsilon}{2}$. Therefore, $\mu(G \setminus F) \leq \mu(G \setminus E) + \mu(E \setminus F) < \varepsilon$.

(ii) $\Longrightarrow$ (iii). For each $n \in \mathbb{N}$, choose $F_n \subseteq E \subseteq G_n$ such that $\mu(G_n \setminus F_n) < \frac{1}{n}$ by (ii). Let $A = \bigcup_{n\in\mathbb{N}} F_n$ and $B = \bigcap_{n\in\mathbb{N}} G_n$. Then A is an F_σ set, B is a G_δ set, $A \subseteq E \subseteq B$, and $\mu(B \setminus A) = 0$.

(iii) $\Longrightarrow$ (i). Suppose (iii) holds. Then we may write $E = A \cup N$, where A is an F_σ set and $N = E \setminus A \subseteq B \setminus A$. Since μ is complete and $\text{Borel}(\mathbb{R}) \subseteq \mathcal{M}_\mu$, we have $E \in \mathcal{M}_\mu$.

(3) Let $E \in \mathcal{M}_\mu$, and let $\varepsilon > 0$. Then by (1), there exists a closed set $F \subseteq E$ such that $\mu(E \setminus F) < \varepsilon$. Let $F_n = F \cap [-n, n]$ for $n \in \mathbb{N}$. Then F_n is compact, $F_1 \subseteq F_2 \subseteq \ldots$, and $F = \bigcup_{n\in\mathbb{N}} F_n$. Therefore, by continuity of μ from below, $\mu(F_n) \to \mu(F)$ as $n \to \infty$. Hence, $\sup\{\mu(K) : K \subseteq E \text{ compact}\} \geq \mu(F) \geq \mu(E) - \varepsilon$. □

Chapter Notes

The construction of Lebesgue–Stieltjes measures presented in this chapter is along the same lines as in the book of Folland [**Fol99**, Sections 1.4 and 1.5]. A similar approach is taken in [**SS05**, Section 3.3] and [**Tao11**, Sections 1.7.1–1.7.3].

The interesting Example 4.23 motivating Carathéodory's criterion for measurability was brought to my attention by [**BBT08**, Example 2.29].

CHAPTER 5

Borel Measures on Locally Compact Hausdorff Spaces

Learning Objectives

At the end of this chapter, you will be able to:

- Construct measures on locally compact Hausdorff spaces
- Describe the relationship between Radon measures and positive linear functionals

In the previous chapter, we constructed locally finite Borel measures on the real line. This is already sufficient for many purposes in probability theory, where the structure of the underlying measure space is often insignificant and the main object of study is (real-valued) random variables. However, in other contexts, the underlying structure of the measure space may play a prominent role (for example, if one is interested in measures on manifolds), and for this, we need additional tools to construct measures on more general topological spaces. The goal of this section is to construct Borel measures on locally compact Hausdorff spaces that are "compatible with the topology" in a sense that will be made precise below.

1. Locally Compact Hausdorff Spaces

Definition 5.1

A topological space X is

- *Hausdorff* if every pair of points can be separated by open sets: if $x, y \in X$ and $x \neq y$, then there are open set $U \ni x$ and $V \ni y$ such that $U \cap V = \emptyset$;
- *locally compact* if for every point has a compact neighborhood: for $x \in X$, there is an open set U and a compact set K such that $x \in U \subseteq K$.

If X is both locally compact and Hausdorff, we say X is a *locally compact Hausdorff space* or an *LCH space* for short.

Example 5.2

Examples of locally compact Hausdorff spaces include:

- the unit interval $[0, 1]$
- the middle-thirds Cantor set
- Euclidean space $\mathbb{R}^d$ for $d \in \mathbb{N}$
- topological manifolds
- discrete spaces

Non-examples include:

- the rational numbers $\mathbb{Q}$ (not locally compact)
- infinite-dimensional real or complex vector spaces (not locally compact)
- an infinite set with the co-finite topology (not Hausdorff)

Definition 5.3

Let X be an LCH space and $f : X \to \mathbb{C}$ a continuous function. The *support of f* is the set

$$\operatorname{supp}(f) = \overline{\{f \neq 0\}}.$$

We say that f is *compactly supported* if $\operatorname{supp}(f)$ is a compact subset of X.

Notation. We denote the space of compactly supported continuous functions on a topological space X by $C_c(X)$.

2. Radon Measures and the Riesz Representation Theorem

Definition 5.4

Let X be an LCH space. A Borel measure μ on X is a *Radon measure* if μ is

- locally finite: $\mu(K) < \infty$ for every compact $K \subseteq X$;
- outer regular: for every Borel set $E \subseteq X$,

  $$\mu(E) = \inf\{\mu(U) : U \text{ is open and } E \subseteq U\};$$

  and
- inner regular on open sets: for every open set $G \subseteq X$,

  $$\mu(G) = \sup\{\mu(K) : K \text{ is compact and } K \subseteq G\}.$$

Let X be an LCH space, and suppose μ is a Radon measure on X. Given $f \in C_c(X)$, we have

$$\int_X |f| \, d\mu \leq \sup_{x \in X} |f(x)| \cdot \mu(\operatorname{supp}(f)).$$

The quantity $\sup_{x \in X} |f(x)|$ is actually a maximum and is finite by the extreme value theorem, while $\mu(\operatorname{supp}(f)) < \infty$ since μ is locally finite. Hence, $C_c(X) \subseteq L^1(\mu)$. Integration against the measure μ thus induces a *positive linear functional* on $C_c(X)$.

Definition 5.5

A *linear functional* on $C_c(X)$ is a linear map $\varphi : C_c(X) \to \mathbb{C}$. We say that a linear functional $\varphi : C_c(X) \to \mathbb{C}$ is *positive* if $\varphi(f) \geq 0$ for every $f \in C_c(X)$ with $f \geq 0$.

It turns out that all positive linear functionals on $C_c(X)$ arise via integration against a measure.

Theorem 5.6: Riesz Representation Theorem

Let X be a locally compact Hausdorff space. Given a positive linear functional $\varphi : C_c(X) \to \mathbb{C}$, there exists a unique Radon measure μ such that

$$\varphi(f) = \int_X f \, d\mu \tag{5.1}$$

for every $f \in C_c(X)$.

Example 5.7

Let $R : C_c(\mathbb{R}) \to \mathbb{C}$ be the functional induced by Riemann integration. That is, if $f : \mathbb{R} \to \mathbb{C}$ with compact support, say $\operatorname{supp}(f) \subseteq [a, b]$, then $R(f) = \int_a^b f(x)\, dx$. The measure representing the functional R is the Lebesgue measure on $\mathbb{R}$.

3. Topological Lemmas

Lemma 5.8

Let X be an LCH space. Suppose $K \subseteq X$ is compact, $U \subseteq X$ is open, and $K \subseteq U$. Then there exists an open set $V \subseteq X$ such that $\overline{V}$ is compact and $K \subseteq V \subseteq \overline{V} \subseteq U$.

Proof. We first handle the case $K = \{x\}$, illustrated in Figure 5.1. Since X is locally compact, there is an open neighborhood $W \subseteq X$ such that $x \in W$ and $\overline{W}$ is compact. If $\overline{W} \subseteq U$, then we are done. Suppose $\overline{W} \not\subseteq U$. Then $L = \overline{W} \setminus U$ is a compact set, and $x \notin L$. For each point $y \in L$, let V_y and O_y be open sets such that $x \in V_y$, $y \in O_y$, and $V_y \cap O_y = \emptyset$. (The sets V_y and O_y exist, since X is Hausdorff.) By compactness of L, there is a finite collection $y_1, \ldots, y_n \in L$ such that $L \subseteq \bigcup_{j=1}^n O_{y_j}$. Let $V = W \cap \bigcap_{j=1}^n V_{y_j}$. Then V is open, $x \in V$, and $\overline{V} \subseteq \overline{W}$ is compact. Moreover, if $z \in \overline{V}$, then $z \in \overline{V}_{y_j}$ for every $j \in \{1, \ldots, n\}$. Hence, $z \notin O_{y_j}$, so $z \notin L$. Therefore, $\overline{V} \subseteq \overline{W} \setminus L \subseteq U$.

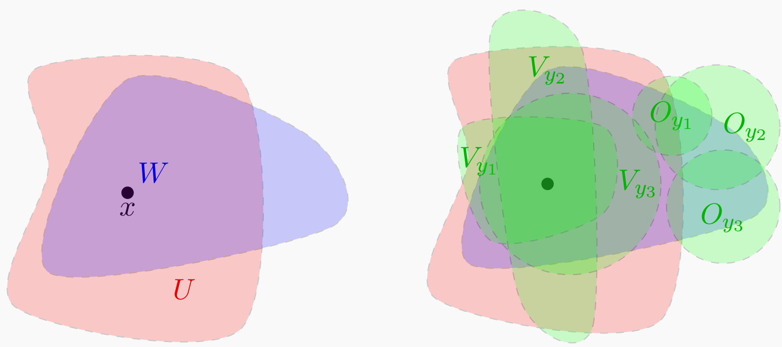


Figure 5.1. Constructing an open set V satisfying $x \in V \subseteq \overline{V} \subseteq U$ using properties of LCH spaces.

Suppose now that K is an arbitrary compact set. By the above, we may find open sets V_x, $x \in K$, such that $x \in V_x \subseteq \overline{V}_x \subseteq U$. By compactness, there is a finite subcover $x_1, \ldots, x_n \in K$ such that $K \subseteq \bigcup_{j=1}^n V_{x_j}$. We then take $V = \bigcup_{j=1}^n V_{x_j}$. □

Corollary 5.9

Let X be an LCH space. Suppose $K \subseteq X$ is compact, $U_1, \ldots, U_N \subseteq X$ are open, and $K \subseteq \bigcup_{n=1}^N U_n$. Then there exists open sets $V_n \subseteq X$ such that $\overline{V}_n \subseteq U_n$ is compact and $K \subseteq \bigcup_{n=1}^N V_n$.

Proof. We do a proof by induction on N. The base case ($N = 1$) is Lemma 5.8. Suppose the statement holds for some $N \in \mathbb{N}$, and let K be a compact set and $U_1, \ldots, U_{N+1}$ an open cover of K. Let $K_1 = K \setminus U_{N+1}$. Then K_1 is a compact set covered by $U_1, \dots, U_N$, so by the inductive hypothesis, there exist open sets $V_1, \dots, V_N$ such that $\overline{V}_n \subseteq U_n$ is compact for $n \leq N$ and $K_1 \subseteq \bigcup_{n=1}^{N} V_n$. Now let $K_2 = K \setminus (\bigcup_{n=1}^{N} V_n)$. Then K_2 is a compact subset of U_{N+1}, so by Lemma 5.8 there exists an open set V_{N+1} such that $\overline{V}_{N+1} \subseteq U_{N+1}$ is compact and $K_2 \subseteq V_{N+1}$. Then $K \subseteq \bigcup_{n=1}^{N+1} V_n$, so the corollary holds by induction.

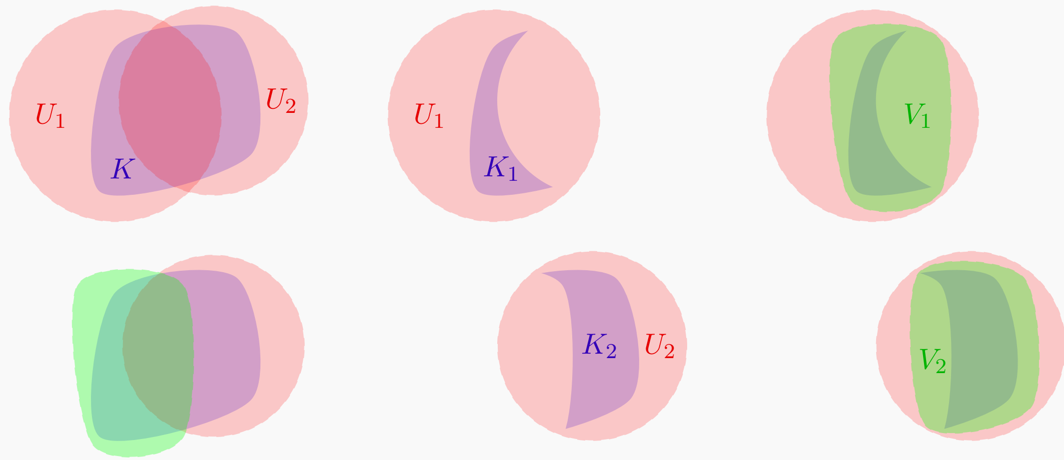


FIGURE 5.2. An illustration of the induction step in the case $N = 2$.

□

LEMMA 5.10: URYSOHN'S LEMMA FOR LCH SPACES

Let X be an LCH space. Given a compact set $K \subseteq X$ and an open set $U \subseteq X$ with $K \subseteq U$, there exists a compactly supported continuous function $f : X \to [0, 1]$ such that $f = 1$ on K and $\operatorname{supp}(f) \subseteq U$.

PROOF. Let K be a compact subset of X and $U \subseteq X$ an open set with $K \subseteq U$. By Lemma 5.8, let V be an open set such that $\overline{V}$ is compact and $K \subseteq V \subseteq \overline{V} \subseteq U$. We construct a function f supported on $\overline{V}$ in terms of its sub-level sets.

Let $K(1) = K$, $V(0) = V$. Put $V(1) = \emptyset$ and $K(0) = \overline{V}$.

CLAIM 1. There are families of open sets $V(r)$ and compact sets $K(r)$ indexed by dyadic rationals $r \in [0, 1]$ such that

- for every dyadic rational $r \in [0, 1]$, $V(r) \subseteq K(r)$, and
- for dyadic rationals $r, s \in [0, 1]$, if $r > s$, then $K(r) \subseteq V(s)$.

We will prove the claim by induction on the denominators of dyadic rationals. By Lemma 5.8, let $V(1/2)$ be an open set such that $K(1/2) = \overline{V(1/2)}$ is compact and $K = K(1) \subseteq V(1/2) \subseteq K(1/2) \subseteq V(0) = V$.

Suppose we have constructed sets $V(r)$ and $K(r)$ with the desired properties for dyadic rationals $r \in (0,1)$ with denominators 2^n for $n < N$. Let $r \in (0,1)$ by a dyadic rational with denominator 2^N, say $r = \frac{2j-1}{2^N}$, $j \in \{1, \ldots, 2^{N-1}\}$. By the induction hypothesis, we have a compact set $K\left(\frac{j}{2^{N-1}}\right)$ and an open set $V\left(\frac{j-1}{2^{N-1}}\right)$ such that $K\left(\frac{j}{2^{N-1}}\right) \subseteq V\left(\frac{j-1}{2^{N-1}}\right)$. Applying Lemma 5.8, we then obtain an open set $V(r)$ such that $K(r) = \overline{U(r)}$ is compact, and $K\left(\frac{j}{2^{N-1}}\right) \subseteq V(r) \subseteq K(r) \subseteq V\left(\frac{j-1}{2^{N-1}}\right)$. The claim thus holds by induction.

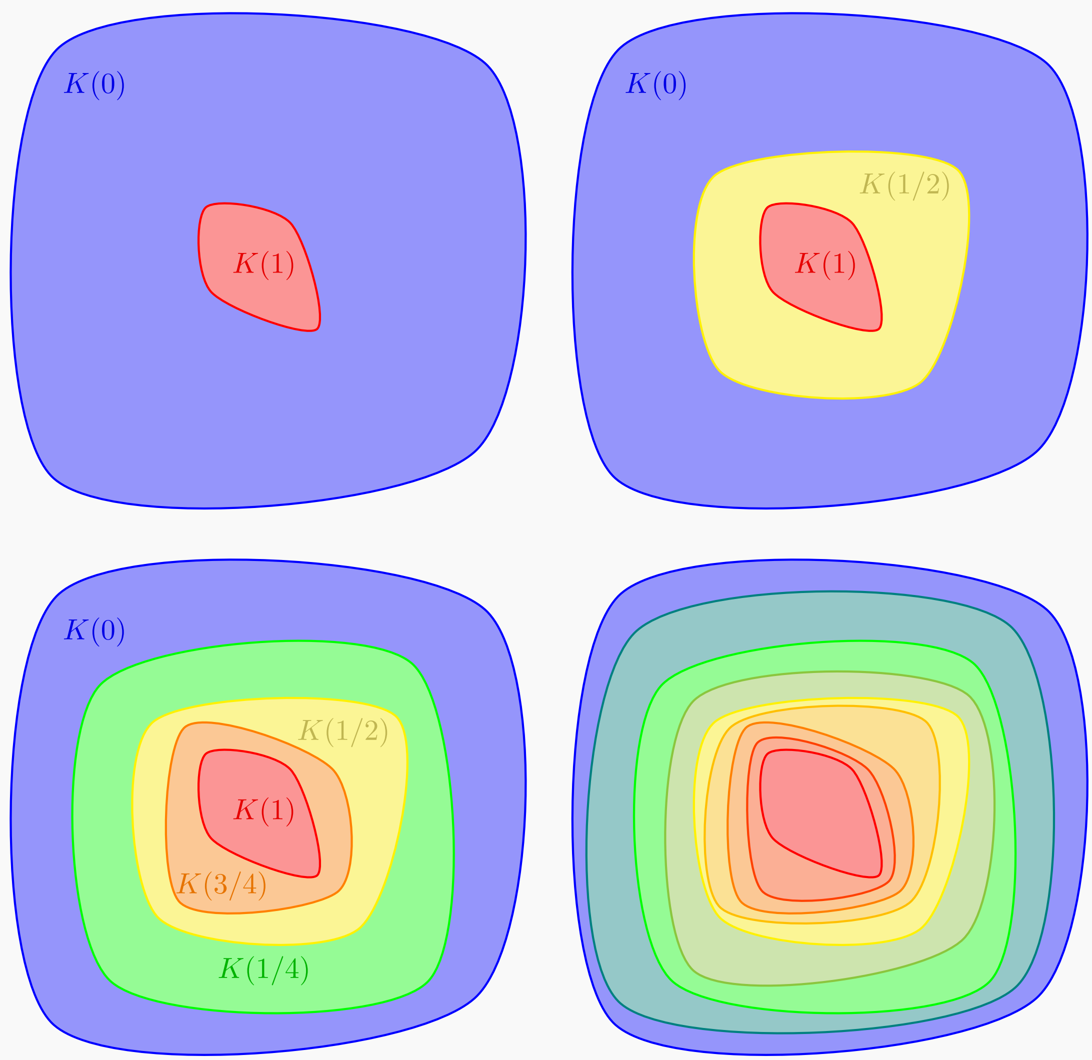


FIGURE 5.3. Inductive construction of sub-level sets.

Define $f(x) = 0$ if $x \notin V(0)$ and $f(x) = \sup\{r \geq 0 : x \in V(r)\}$ otherwise. By construction $K = K(1) \subseteq V(r)$ for every dyadic rational $r \in [0,1)$, so $f = 1$ on K. Moreover, $\operatorname{supp}(f) \subseteq K(0) \subseteq U$. It remains to show that f is continuous.

Claim 2. For $a \in \mathbb{R}$,

$$\{f > a\} = \begin{cases} \emptyset, & \text{if } a \geq 1; \\ \bigcup_{r>a} V(r), & \text{if } 0 \leq a < 1; \\ X, & \text{if } a < 0. \end{cases}$$

The function f takes values between 0 and 1, so the cases $a < 0$ and $a \geq 1$ are immediate. Let $0 \leq a < 1$. Suppose $x \in X$ and $f(x) > a$. Then by the definition of f, there exists $r > a$ such that $x \in V(r)$. Hence, $x \in \bigcup_{r>a} V(r)$. Conversely, if $x \in \bigcup_{r>a} V(r)$, then $f(x) \geq r > a$.

Claim 3. For $b \in \mathbb{R}$,

$$\{f < b\} = \begin{cases} X, & \text{if } b > 1; \\ \bigcup_{r<b}(X \setminus K(r)), & \text{if } 0 < b \leq 1; \\ \emptyset, & \text{if } b \leq 0. \end{cases}$$

As in the previous claim, since f takes values in the interval $[0, 1]$, the cases $b > 1$ and $b \leq 0$ are immediate. Let $0 < b \leq 1$. Suppose $f(x) < b$. Taking $s \in (f(x), b)$, we have $x \notin V(s)$. Let $r \in (s, b)$. Then since $K(r) \subseteq V(s)$, we conclude $x \notin K(r)$. Hence, $x \in \bigcup_{r<b}(X \setminus K(r))$. Conversely, if $x \notin K(r)$ for some $r < b$, then $x \notin V(s)$ for $s \geq r$, so $f(x) \leq r < b$.

Combining Claims 2 and 3, for any $a, b \in \mathbb{R}$, the set $\{a < f < b\}$ is an intersection of two open sets and therefore open. Thus, f is continuous. □

Notation. For a compact set $K \subseteq X$ and a function $f : X \to [0, 1]$, we write $K \prec f$ if $f = 1$ on K. Given an open set $U \subseteq X$ and a function $f : X \to [0, 1]$, we write $f \prec U$ if $\operatorname{supp}(f) \subseteq U$. With this notation, the conclusion of Urysohn's lemma reads $K \prec f \prec U$.

Corollary 5.11: Partition of Unity

Let X be an LCH space. Let $K \subseteq X$ be a compact set and $U_1, \dots, U_N \subseteq X$ an open cover of K. Then there exist compactly supported continuous functions $h_n : X \to [0, 1]$ such that $h_n \prec U_n$ for each $n \in \{1, \dots, N\}$ and $\sum_{n=1}^{N} h_n = 1$ on K.

Proof. First apply Corollary 5.9 to obtain open sets $V_1, \dots, V_N$ such that $\overline{V}_n \subseteq U_n$ is compact and $K \subseteq \bigcup_{n=1}^{N} V_n$. Then by Urysohn's lemma, let $f_n \in C_c(X)$ with $\overline{V}_n \prec f_n \prec U_n$. Define $h_1 = f_1, h_2 = (1 - f_1)f_2, \dots, h_N = (1 - f_1) \dots (1 - f_{N-1})f_N$. For each $n \in \{1, \dots, N\}$, $h_n \leq f_n$, so h_n has compact support, and $h_n \prec U_n$. Moreover, it can be checked by induction on N that

$$\sum_{n=1}^{N} h_n = 1 - \prod_{n=1}^{N}(1 - f_n).$$

For $x \in K$, at least one of the functions $f_n(x)$ is equal to 1, so $\sum_{n=1}^{N} h_n(x) = 1$. □

Lemma 5.12

Let X be a locally compact Hausdorff space, and suppose μ is a Radon measure on X. Then for any open set $U \subseteq X$,

$$\mu(U) = \sup\left\{\int_X f\ d\mu : f \in C_c(X), 0 \leq f \prec U\right\}.$$

Proof. Clearly $\mu(U) \geq \int_X f\ d\mu$ for any $f \in C_c(x)$ with $0 \leq f \leq \mathbb{1}_U$. Let us prove the reverse inequality. Let $c < \mu(U)$ be arbitrary. Then by inner regularity of μ on open sets, there exists a compact set $K \subseteq U$ such that $\mu(K) > c$. Then by Urysohn's lemma, there is a continuous function $f \in C_c(X)$ such that $K \prec f \prec U$. By monotonicity of the integral, we then have

$$\int_X f\ d\mu \geq \mu(K) > c.$$

□

Proof of Riesz Representation Theorem. We will carry out the proof in several steps.

Step 1. Uniqueness.

Suppose μ and ν are two Radon measures satisfying (5.1). By Lemma 5.12, μ and ν must agree on all open subsets of X. But then by outer regularity, μ and ν agree on all Borel sets.

Step 2. Defining an Outer Measure.

Motivated by Lemma 5.12, we define

$$m(U) = \sup\{\varphi(f) : f \in C_c(X), 0 \leq f \prec U\}$$

for open subsets $U \subseteq X$, and let

$$\mu^*(E) = \inf\{m(U) : U \text{ is open and } E \subseteq U\}$$

for $E \subseteq X$.

We must check that μ^* is an outer measure. That $\mu^*(\emptyset) = 0$ and μ^* is monotone are both easy consequences of the definition of μ^*. Suppose $(E_n)_{n\in\mathbb{N}}$ is a countable family of subsets of X. We want to show

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} E_n\right) \leq \sum_{n=1}^{\infty} \mu^*(E_n).$$

If the sum diverges, there is nothing to show, so assume the sum is finite. Let $\varepsilon > 0$ be arbitrary. Take $(U_n)_{n\in\mathbb{N}}$ open sets such that $E_n \subseteq U_n$ and $\mu^*(E_n) > m(U_n) - 2^{-n}\varepsilon$. Let $f \in C_c(X)$ with $0 \leq f \prec \bigcup_{n\in\mathbb{N}} U_n$. Since the sets $(U_n)_{n\in\mathbb{N}}$ are open and $\operatorname{supp}(f)$ is a compact set, there exists $N \in \mathbb{N}$ such that $\operatorname{supp}(f) \subseteq \bigcup_{n=1}^N U_n$. By partition of unity (Corollary 5.11), let $h_1, \ldots, h_N \in C_c(X)$ such that $h_n \prec U_n$ and $\sum_{n=1}^N h_n = 1$ on $\operatorname{supp}(f)$. Letting $f_n = f \cdot h_n$, we have $f = \sum_{n=1}^N f_n$. Therefore,

$$\varphi(f) = \sum_{n=1}^{N} \varphi(f_n) \leq \sum_{n=1}^{N} m(U_n) \leq \sum_{n=1}^{\infty} m(U_n) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon.$$

Taking a supremum over all such f, we conclude

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} E_n\right) \leq m\left(\bigcup_{n\in\mathbb{N}} U_n\right) = \sup\left\{\varphi(f) : f \in C_c(X), 0 \leq f \prec \bigcup_{n\in\mathbb{N}} U_n\right\} \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon.$$

Since ε was arbitrary, this proves that μ^* is countably subadditive and therefore an outer measure.

STEP 3. Borel Sets are μ^*-Measurable.

By Carathéodory's theorem (Theorem 4.25), the family of μ^*-measurable sets is a σ-algebra, so it suffices to check that every open set is μ^*-measurable. Let $U \subseteq X$ be an open set. We want to show

$$\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U) \tag{5.2}$$

for every $E \subseteq X$ with $\mu^*(E) < \infty$. Let $E \subseteq X$ be any set with $\mu^*(E) < \infty$. Let $\varepsilon > 0$ be arbitrary, and choose an open set $V \subseteq X$ with $E \subseteq V$ and $\mu^*(E) > m(V) - \varepsilon$. The set $V \cap U$ is open, so choose a function $f_1 \in C_c(X)$ with $0 \leq f_1 \prec V \cap U$ such that $\varphi(f_1) > m(V \cap U) - \varepsilon$. Let $K = \operatorname{supp}(f_1)$, and then choose a function $f_2 \in C_c(X)$ with $0 \leq f_2 \prec V \setminus K$ such that $\varphi(f_2) > m(V \setminus K) - \varepsilon$. Then

$$\mu^*(E) > m(V) - \varepsilon \geq \varphi(f_1 + f_2) - \varepsilon > m(V \cap U) + m(V \setminus K) - 3\varepsilon \geq \mu^*(E \cap U) + \mu^*(E \setminus U) - 3\varepsilon.$$

Taking $\varepsilon \to 0$ gives (5.2).

We can now define a Borel measure μ by $\mu = \mu^*|_{\mathrm{Borel}(X)}$.

STEP 4. Regularity of μ.

The measure μ is outer regular by construction. We will check that it is also inner regular on open sets. Let $U \subseteq X$ be open, and let $c < \mu(U)$. By the definition of μ, there exists a function $f \in C_c(X)$ with $0 \leq f \prec U$ such that $\varphi(f) > c$. Let $K = \operatorname{supp}(f)$. We claim $\mu(K) \geq \varphi(f) > c$. From the definition of μ, it suffices to show: if $V \supseteq K$ is open, then there exists a function $g \in C_c(X)$ with $0 \leq g \prec V$ with $\varphi(g) > c$. But this is immediate upon taking $g = f$.

STEP 5. Local Finiteness.

Let $K \subseteq X$ be compact. We want to show $\mu(K) < \infty$. It suffices to find an open set $U \supseteq K$ with $m(U) < \infty$. Since X is locally compact, there exists an open set $U \supseteq K$ such that $\overline{U}$ is compact (see Lemma 5.8). Let $f \in C_c(X)$ with $\overline{U} \prec f$. Suppose $g \in C_c(X)$ with $0 \leq g \prec U$. Then $f - g \geq 0$, so $\varphi(g) \leq \varphi(f)$. Thus,

$$m(U) = \sup\{\varphi(g) : g \in C_c(x), 0 \leq g \leq 1, \operatorname{supp}(g) \subseteq U\} \leq \varphi(f) < \infty.$$

STEP 6. Proving the Identity (5.1).

Let $f \in C_c(X)$, and let $K = \operatorname{supp}(f)$. By decomposing f into real and imaginary parts, then positive and negative parts and scaling, we may assume $0 \le f \le 1$. Given $N \in \mathbb{N}$, we decompose $K = \bigsqcup_{n=0}^{N} K_n$, where $K_0 = \{x \in K : f(x) = 0\}$ and $K_n = \left\{x \in K : f(x) \in \left(\frac{n-1}{N}, \frac{n}{N}\right]\right\}$ for $n \ge 1$. For $n \in \{1, \ldots, N\}$, let

$$f_n(x) = \begin{cases} 0, & \text{if } x \in K_m, m < n \\ f(x) - \frac{n-1}{N}, & \text{if } x \in K_n \\ \frac{1}{N}, & \text{if } x \in K_m, m > n. \end{cases}$$

Then $f_n \in C_c(X)$ and $f = \sum_{n=1}^{N} f_n$. Moreover, $\bigsqcup_{m>n} \overline{K}_m \prec N f_n \prec \bigsqcup_{m \ge n} K_m$. We can therefore estimate $\varphi(f)$ and $\int_X f \, d\mu$ as follows:

$$\frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m>n} K_m\right) \le \varphi(f) \le \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m \ge n} K_m\right).$$

and

$$\frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m>n} K_m\right) \le \int_X f \, d\mu \le \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m \ge n} K_m\right).$$

All that remains is to check that the two sides of the inequality become arbitrarily close as $N \to \infty$. Observe:

$$\frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m \ge n} K_m\right) - \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigsqcup_{m>n} K_m\right) = \frac{1}{N} \sum_{n=1}^{N} \mu(K_n) = \frac{1}{N}\mu(K),$$

and $\mu(K) < \infty$ by Step 5. Therefore, taking $N \to \infty$ and applying the squeeze theorem, we conclude

$$\int_X f \, d\mu = \varphi(f).$$

□

5. Convergence in the Space of Measures

Let X be an LCH space. The Riesz representation theorem establishes an identification of the space of Radon measures on X with the space of positive linear functionals on $C_c(X)$. Spaces of functionals can be equipped rather naturally with topological structure, and we can use this to define a topology on the space of Radon measures.

Notation. For the remainder of this section, we denote the space of Radon measures on X by $\mathcal{M}(X)$.

Definition 5.13

The *vague topology* on $\mathcal{M}(X)$ is the topology generated by the basis of open sets

$$B(\mu, \mathcal{F}, \varepsilon) = \left\{\nu \in \mathcal{M}(X) : \left|\int_X f \, d\nu - \int_X f \, d\mu\right| < \varepsilon \text{ for all } f \in \mathcal{F}\right\}$$

for $\mu \in \mathcal{M}(X)$, $\mathcal{F} \subseteq C_c(X)$ a finite set of functions, and $\varepsilon > 0$.

Remark. As alluded to above, the vague topology arises from viewing $\mathcal{M}(X)$ as a space of functionals. It is a particular instance of a more general topological construction from functional analysis called the *weak* topology*. For this reason, the vague topology is also often called the weak* topology.

For general LCH spaces X, it can be difficult to fully grasp the vague topology. However, we can give a more concrete description in the case that X is a compact metric space, and this is already sufficient for many applications in probability theory, harmonic analysis, and ergodic theory.

Proposition 5.14

Let X be a compact metric space. Then $(C(X), \|\cdot\|_{\sup})$ is a separable metric space.

Proof. Compact metric spaces are separable. Indeed, for each $k \in \mathbb{N}$, we may cover X by finitely many open balls of radius $\frac{1}{k}$. Taking S_k to be the finite set of centers of these open balls, the set $S = \bigcup_{k\in\mathbb{N}} S_k$ is countable and dense. Let $(x_n)_{n\in\mathbb{N}}$ be an enumeration of S. For $n \in \mathbb{N}$, let $f_n : X \to \mathbb{R}$ be the continuous function $f_n(x) = d(x, x_n)$. Let $\mathcal{A}$ be the algebra of functions generated by $(f_n)_{n\in\mathbb{N}}$. That is,

$$\mathcal{A} = \operatorname{span}_{\mathbb{C}} \left\{ \prod_{n\in F} f_n : F \subseteq \mathbb{N} \text{ finite} \right\},$$

where the empty product is taken to be the constant function $\mathbb{1}$. Then $\mathcal{A}$ is closed under conjugation and contains the constant functions. Moreover, if $x \neq y$, then by density of the sequence $(x_n)_{n\in\mathbb{N}}$, there exists $n \in \mathbb{N}$ such that $d(x, x_n) \neq d(y, x_n)$, so $f_n(x) \neq f_n(y)$. Hence, $\mathcal{A}$ separates points. By the Stone–Weierstrass theorem, it follows that $\mathcal{A}$ is dense in $C(X)$. To finish, we note that the generating set $(f_n)_{n\in\mathbb{N}}$ is countable, so the subset

$$\mathcal{A}_{\mathbb{Q}} = \operatorname{span}_{\mathbb{Q}(i)} \left\{ \prod_{n\in F} f_n : F \subseteq \mathbb{N} \text{ finite} \right\}$$

is countable and dense in $\mathcal{A}$. □

Proposition 5.15

Let X be a compact metric space. Then $\mathcal{M}(X)$ with the vague topology is metrizable.

Proof. We give an explicit formula for the metric. Let $g_1, g_2, \dots$ be a countable dense sequence in $C(X)$. For convenience, we will assume $g_1 = \mathbb{1}$. Define

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{\left| \int_X g_n \, d\mu - \int_X g_n \, d\nu \right|}{2^n \|g_n\|_{\sup}}.$$

Claim 1. d is a metric on $\mathcal{M}(X)$.

It is immediate from the definition that d is nonnegative and symmetric. Moreover,

$$d(\mu, \nu) \leq \mu(X) + \nu(X) < \infty$$

for every $\mu, \nu \in \mathcal{M}(X)$. The triangle inequality for d follows from the usual triangle inequality for complex numbers:

$$d(\mu,\rho) = \sum_{n=1}^{\infty} \frac{\left|\int_X g_n \, d\mu - \int_X g_n \, d\rho\right|}{2^n \|g_n\|_{\sup}}$$
$$\leq \sum_{n=1}^{\infty} \frac{\left|\int_X g_n \, d\mu - \int_X g_n \, d\nu\right| + \left|\int_X g_n \, d\nu - \int_X g_n \, d\rho\right|}{2^n \|g_n\|_{\sup}} = d(\mu,\nu) + d(\nu,\rho).$$

Finally, suppose $d(\mu,\nu) = 0$. Then for every $n \in \mathbb{N}$, $\int_X g_n \, d\mu = \int_X g_n \, d\nu$. Let $f \in C(X)$ be arbitrary. There exists a sequence $(n_k)_{k\in\mathbb{N}}$ such that $g_{n_k} \to f$ uniformly. Hence, by the dominated convergence theorem (the functions g_{n_k} are bounded pointwise by $\sup|f| + 1$ for all large enough k), we have

$$\int_X f \, d\mu = \lim_{k\to\infty} \int_X g_{n_k} \, d\mu = \lim_{k\to\infty} \int_X g_{n_k} \, d\nu = \int_X f \, d\nu.$$

By the uniqueness part of the Riesz representation theorem, we conclude that $\mu = \nu$.

Claim 2. The metric d induces the vague topology on $\mathcal{M}(X)$.

We need to show that every vaguely open set contains a d-open ball and *vice versa*.

Suppose U is an open neighborhood of a point $\mu \in \mathcal{M}(X)$ in the vague topology. Then $B(\mu, \mathcal{F}, \varepsilon) \subseteq U$ for some finite family of functions $\mathcal{F} \subseteq C(X)$ and $\varepsilon \in (0,1)$. For each $f \in \mathcal{F}$, we may find $k \in \mathbb{N}$ such that $\left\|g_{k_f} - f\right\|_{\sup} < \frac{\varepsilon}{3\mu(X)+1}$. Let $K = \max\{k_f : f \in \mathcal{F}\}$. Suppose $\nu \in \mathcal{M}(X)$ with $d(\mu,\nu) < \frac{\varepsilon}{3\cdot 2^K}$. Then

$$\left|\int_X g_k \, d\mu - \int_X g_k \, d\nu\right| < \frac{\varepsilon}{3}$$

for each $k \in \{1,\dots,K\}$. In particular (taking $k=1$), we have $|\mu(X) - \nu(X)| < \frac{\varepsilon}{3}$, and

$$\left|\int_X g_{k_f} \, d\mu - \int_X g_{k_f} \, d\nu\right| < \frac{\varepsilon}{3}$$

for every $f \in \mathcal{F}$. Thus, for each $f \in \mathcal{F}$,

$$\left|\int_X f \, d\mu - \int_X f \, d\nu\right|$$
$$\leq \left|\int_X f \, d\mu - \int_X g_{k_f} \, d\mu\right| + \left|\int_X g_{k_f} \, d\mu - \int_X g_{k_f} \, d\nu\right| + \left|\int_X g_{k_f} \, d\nu - \int_X f \, d\nu\right|$$
$$\leq \frac{\varepsilon}{3\mu(X)+1} \cdot \mu(X) + \frac{\varepsilon}{3} + \frac{\varepsilon}{3\mu(X)+1} \cdot \nu(X) < \varepsilon.$$

Therefore, the ball of radius $\frac{\varepsilon}{3\cdot 2^K}$ around μ is a subset of $B(\mu,\mathcal{F},\varepsilon) \subseteq U$.

Conversely, suppose $\mu \in \mathcal{M}(X)$ and $r > 0$. Let $N \in \mathbb{N}$ such that $2^{-N} = \sum_{n>N} 2^{-n} < \frac{r}{2}$. Suppose $\nu \in B\left(\mu, \{g_1,\dots,g_N\}, \frac{r}{2} \cdot \min_{1\leq n\leq N} \|g_n\|_{\sup}\right)$. Then

$$d(\mu,\nu) = \sum_{n=1}^{\infty} \frac{\left|\int_X g_n \, d\mu - \int_X g_n \, d\nu\right|}{2^n \|g_n\|_{\sup}} \leq \sum_{n=1}^{N} \frac{\frac{r}{2} \cdot \min_{1\leq k\leq N} \|g_k\|_{\sup}}{2^n \|g_n\|_{\sup}} + \sum_{n>N} 2^{-n} < \frac{r}{2} + \frac{r}{2} = r.$$

□

Remark. In a metric space, the topology is fully described by convergence of sequences. For compact metric X, we can describe the vague topology on $\mathcal{M}(X)$ by:

$$\mu_n \to \mu \text{ vaguely } \iff \forall f \in C(X), \int_X f\ d\mu = \lim_{n\to\infty} \int_X f\ d\mu_n.$$

In other words, the vague topology on $\mathcal{M}(X)$ is the topology of pointwise convergence when viewing measures as functionals on $C(X)$ (and functions $f \in C(X)$ as points).

Assuming X is compact, every Radon measure on X is finite, so we can write $\mu \in \mathcal{M}(X)$ as $\mu = c\mu'$ for some probability measure μ'. Moreover, in this context of compact metric X, every finite Borel measure is Radon:

Proposition 5.16

Let X be a compact metric space. A Borel measure on X is Radon if and only if it is finite.

Proof. The fact that Radon measures on X are finite is immediate from the definition of a Radon measure.

Conversely, suppose $\mu : \text{Borel}(X) \to [0, \infty)$ is a finite measure. The measure μ induces a positive linear function $I_\mu : f \mapsto \int_X f\ d\mu$ on $C(X)$. By the Riesz representation theorem, let μ' be the Radon measure representing I_μ. That is, $\int_X f\ d\mu' = \int_X f\ d\mu$ for every $f \in C(X)$.

Claim 1. If $U \subseteq X$ is open, then $\mu'(U) = \mu(U)$.

If $U = X$, then $\mu'(X) = \int_X 1\ d\mu' = \int_X 1\ d\mu = \mu(X)$. Suppose $U \neq X$. Let $D(x) = d(x, X \setminus U) = \min\{d(x,y) : y \in X \setminus U\}$. (Note that this minimum exists by the extreme value theorem, since $X \setminus U$ is a compact set.) Then $x \in U$ if and only if $D(x) > 0$. Define a sequence of functions $f_n : X \to [0,1]$ by $f_n(x) = \max\{nD(x), 1\}$. Then $(f_n)_{n\in\mathbb{N}}$ is an increasing sequence of continuous functions with $f_n \to \mathbb{1}_U$, so by the monotone convergence theorem,

$$\mu'(U) = \int_X \mathbb{1}_U\ d\mu' = \lim_{n\to\infty} \int_X f_n\ d\mu' = \lim_{n\to\infty} \int_X f_n\ d\mu = \int_X \mathbb{1}_U\ d\mu = \mu(U).$$

Claim 2. The family of subsets $\mathcal{L} = \{E \in \text{Borel}(X) : \mu'(E) = \mu(E)\}$ is a λ-system.

See Example 4.7.

The collection of open subsets of X is a π-system generating the Borel σ-algebra, so by the π-λ theorem (Theorem 4.9), $\mathcal{L} = \text{Borel}(X)$. That is, $\mu' = \mu$. Since μ' is Radon by construction, we conclude that original measure μ is Radon. □

In this way, we can understand every Radon measure on X in terms of probability measures.

Notation. We will denote the space of Borel probability measures on X by $\mathcal{M}_1(X)$.

A deep theorem from functional analysis (the Banach–Alaoglu theorem) has the following important corollary.

Theorem 5.17

Let X be a compact metric space. Then $\mathcal{M}_1(X)$ is compact in the vague topology.

Corollary 5.18

Let X be a compact metric space. Every sequence $(\mu_n)_{n\in\mathbb{N}}$ of Borel probability measures on X has a convergent subsequence.

CHAPTER 6

Products of Measure Spaces

Learning Objectives

At the end of this chapter, you will be able to:

- Define product measures.
- Construct product measures with at least two different methods.
- Identify when there is a unique product measure.
- Apply product measures and the Fubini–Tonelli theorem to solve problems in analysis.

Definition 6.1

Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces. A *measurable rectangle* in $X \times Y$ is a set of the form $B \times C$ such that $B \in \mathcal{B}$ and $C \in \mathcal{C}$. The *product σ-algebra* $\mathcal{B} \otimes \mathcal{C}$ on $X \times Y$ is the σ-algebra generated by measurable rectangles.

Remark. The product can also be defined in a category-theoretic way. Let $\pi_X : X \times Y$ be the projection onto the first coordinate, $\pi_X(x, y) = x$, and let $\pi_Y : X \times Y$ be the projection onto the second coordinate, $\pi_Y(x, y) = y$. The maps π_X and π_Y are easily checked to be measurable maps defined on $(X \times Y, \mathcal{B} \otimes \mathcal{C})$. The product measurable space $(X \times Y, \mathcal{B} \otimes \mathcal{C})$ satisfies the following universal property (see Figure 6.1): for any measurable space $(Z, \mathcal{D})$ and any measurable functions $f : Z \to X$ and $g : Z \to Y$, there is a unique measurable function $h : Z \to X \times Y$ such that $\pi_X \circ h = f$ and $\pi_Y \circ h = g$. This universal property characterizes the product space $(X \times Y, \mathcal{B} \otimes \mathcal{C})$ uniquely up to isomorphism.

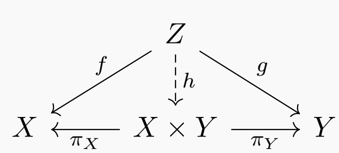


Figure 6.1. Universal property of product spaces.

Definition 6.2

Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be measure spaces. A measure $\rho : \mathcal{B} \otimes C \to [0, \infty]$ is a *product measure* of μ and ν if $\rho(B \times C) = \mu(B)\nu(C)$ for every $B \in \mathcal{B}$ and $C \in \mathcal{C}$.

Theorem 6.3

Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be measure spaces. There exists a product measure $\rho : \mathcal{B} \otimes \mathcal{C} \to [0, \infty]$. Moreover, if μ and ν are σ-finite, then there is a unique product measure.

The uniqueness part of Theorem 6.3 follows by the π-λ theorem, so we present its proof first.

Proof of uniqueness of product measure for σ-finite spaces. Suppose $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ are σ-finite measure spaces, and suppose $\rho_1, \rho_2 : \mathcal{B} \otimes \mathcal{C} \to [0, \infty]$ are product measures of μ and ν. We want to show that $\rho_1 = \rho_2$. We will use the π-λ theorem.

The family $\mathcal{R}$ of measurable rectangles is a π-system (see Example 4.7).

Since μ and ν are σ-finite measures, we may write $X = \bigcup_{n \in \mathbb{N}} X_n$ with $X_1 \subseteq X_2 \subseteq \dots$ such that $\mu(X_n) < \infty$ and $Y = \bigcup_{n \in \mathbb{N}} Y_n$ with $Y_1 \subseteq Y_2 \subseteq \dots$ such that $\nu(Y_n) < \infty$. Define a family

$$\mathcal{L} = \{E \in \mathcal{B} \otimes \mathcal{C} : \rho_1(E \cap (X_n \times Y_n)) = \rho_2(E \cap (X_n \times Y_n)) \text{ for every } n \in \mathbb{N}\}.$$

By the same argument as in the proof of Corollary 4.14, $\mathcal{L}$ is a λ-system on $X \times Y$.

Moreover, $\mathcal{R} \subseteq \mathcal{L}$. Indeed, if $E = B \times C$ is a measurable rectangle, then $E \cap (X_n \times Y_n) = (B \cap X_n) \times (C \cap Y_n)$ is also a measurable rectangle for every $n \in \mathbb{N}$, so

$$\rho_1 (E \cap (X_n \times Y_n)) = \mu(B \cap X_n)\nu(C \cap Y_n) = \rho_2 (E \cap (X_n \times Y_n))$$

by the definition of a product measure.

Thus, by the π-λ theorem, $\mathcal{L} \supseteq \sigma(\mathcal{R}) = \mathcal{B} \otimes \mathcal{C}$. Applying continuity from below of the measures ρ_1 and ρ_2, given an arbitrary measurable set $E \in \mathcal{B} \otimes \mathcal{C}$, we have

$$\rho_1(E) = \lim_{n \to \infty} \rho_1 (E \cap (X_n \times Y_n)) = \lim_{n \to \infty} \rho_2 (E \cap (X_n \times Y_n)) = \rho_2(E).$$

That is, $\rho_1 = \rho_2$. □

The preceding proof shows that it makes sense to talk about *the* product measure of a pair of σ-finite measures.

For the existence part of Theorem 6.3, several different constructions of product measures are possible, and in the case of non-σ-finite spaces, different constructions may produce different measures.

1. Cross-Sectional Product Measures and the Fubini–Tonelli Theorem

Definition 6.4

Let X and Y be sets, and let $(x, y) \in X \times Y$.

- for a set $E \subseteq X \times Y$, the *x-section* E_x and the *y-section* E^y of E are defined by

$$E_x = \{v \in Y : (x, v) \in E\} \qquad \text{and} \qquad E^y = \{u \in X : (u, y) \in E\}.$$

- for a function f defined on $X \times Y$, the *x-section* f_x and the *y-section* f^y of f are defined by

$$f_x(v) = f(x, v) \qquad \text{and} \qquad f^y(u) = f(u, y)$$

  for $v \in Y$ and $u \in X$.

Remark. If $E \subseteq X \times Y$, then we have the identities $(\mathbb{1}_E)_x = \mathbb{1}_{E_x}$ and $(\mathbb{1}_E)^y = \mathbb{1}_{E^y}$.

Proposition 6.5

Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces.

(1) If $E \in \mathcal{B} \otimes \mathcal{C}$, then $E_x \in \mathcal{C}$ for every $x \in X$ and $E^y \in \mathcal{B}$ for every $y \in Y$.

(2) If f is a $(\mathcal{B} \otimes \mathcal{C})$-measurable function on $X \times Y$, then f_x is $\mathcal{C}$-measurable for every $x \in X$ and f^y is $\mathcal{B}$-measurable for every $y \in Y$.

Proof. **(1)** Consider the family

$$\mathcal{F} = \{E \subseteq X \times Y : E_x \in \mathcal{C} \text{ for every } x \in X \text{ and } E^y \in \mathcal{B} \text{ for every } y \in Y\}.$$

Then $\mathcal{F}$ contains all measurable rectangles, since

$$(B \times C)_x = \begin{cases} C, & \text{if } x \in B; \\ \emptyset, & \text{if } x \notin B, \end{cases} \qquad \text{and} \qquad (B \times C)^y = \begin{cases} B, & \text{if } y \in C; \\ \emptyset, & \text{if } y \notin C. \end{cases} \tag{6.1}$$

Moreover, since taking cross-sections is compatible with (countable) unions and complements, $\mathcal{F}$ is a σ-algebra. Hence, $\mathcal{B} \otimes \mathcal{C} \subseteq \mathcal{F}$, which proves (1).

(2) This follows from (1) by noting that pre-images are compatible with cross-sections in the sense that $(f_x)^{-1}(E) = (f^{-1}(E))_x$ and $(f^y)^{-1}(E) = (f^{-1}(E))^y$. □

Informally, item (1) in Proposition 6.5 says that every horizontal and vertical slice of a product-measurable set is measurable. The following example shows that the converse is not true: there exist nonmeasurable sets such that every cross-section is nevertheless measurable.

Example 6.6

Take $X = Y = \mathbb{R}$ with the Borel σ-algebra $\mathcal{B} = \mathcal{C} = \text{Borel}(\mathbb{R})$. Let E be an arbitrary non-Borel subset of $\mathbb{R}$ and consider $\widetilde{E} = \{(t,t) : t \in E\} \subseteq \mathbb{R}^2$. Every cross-section of $\widetilde{E}$ is either empty or a singleton and hence Borel. However, $\widetilde{E}$ is not an element of the product σ-algebra $\mathcal{B} \otimes \mathcal{C} = \text{Borel}(\mathbb{R}^2)$. To see this, note that $E = f^{-1}(\widetilde{E})$ for the continuous function $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (t,t)$. Continuous functions are Borel-measurable, so the preimage of every Borel set under f is Borel. But E is non-Borel by assumption, so we conclude that $\widetilde{E}$ is also non-Borel.

Theorem 6.7

Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be measure spaces.

(1) If ν is s-finite,[a] then the map $x \mapsto \nu(E_x)$ is measurable, and $\rho : \mathcal{B} \otimes \mathcal{C} \to [0, \infty]$ defined by

$$\rho(E) = \int_X \nu(E_x)\, d\mu(x)$$

is a product measure of μ and ν.

(2) If μ and ν are both s-finite, then

$$\int_X \nu(E_x)\, d\mu(x) = \int_Y \mu(E^y)\, d\nu(y)$$

for every $E \in \mathcal{B} \otimes \mathcal{C}$.

[a] A measure is s-finite if it can be expressed as a countable sum of finite measures; see Appendix B.

Proof. **(1)** Suppose ν is s-finite.

Claim 1. The map $x \mapsto \nu(E_x)$ is measurable.

We can write $\nu = \sum_{n=1}^{\infty} \nu_n$ for some finite measures $\nu_n : \mathcal{C} \to [0, \infty)$. Since a countable sum of measurable functions is measurable, it suffices to prove measurability under the stronger hypothesis that ν is finite.

Consider the family of sets

$$\mathcal{L} = \{E \subseteq X \times Y : x \mapsto \nu(E_x) \text{ is measurable}\}.$$

Using (6.1), we see that $\mathcal{L}$ contains the π-system of measurable rectangles. By the π-λ theorem, it therefore suffices to prove that $\mathcal{L}$ is a λ-system.

The set $X \times Y$ is a measurable rectangle so belongs to $\mathcal{L}$.

Suppose $E \in \mathcal{L}$. Noting that $((X \times Y) \setminus E)_x = Y \setminus E_x$, we have that

$$\nu(((X \times Y) \setminus E)_x) = \nu(Y) - \nu(E_x)$$

is measurable, so $(X \times Y) \setminus E \in \mathcal{L}$.

Finally, if $(E_n)_{n \in \mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{L}$ and $E = \bigsqcup_{n \in \mathbb{N}} E_n$, then

$$\nu(E_x) = \sum_{n=1}^{\infty} \nu((E_n)_x)$$

is measurable, so $E \in \mathcal{L}$.

Measurability of $x \mapsto \nu(E_x)$ means that ρ is a well-defined function.

Claim 2. ρ is a measure on $(X \times Y, \mathcal{B} \otimes \mathcal{C})$.

Since $\emptyset_x = \emptyset$ for every $x \in X$, we have

$$\rho(\emptyset) = \int_X \nu(\emptyset)\ d\mu = \int_X 0\ d\mu = 0 \cdot \mu(X) = 0.$$

Suppose $(E_n)_{n \in \mathbb{N}}$ is a sequence of pairwise disjoint measurable subsets of $X \times Y$. Then

$$\begin{aligned}
\rho\left(\bigsqcup_{n \in \mathbb{N}} E_n\right) &= \int_X \nu\left(\bigsqcup_{n \in \mathbb{N}} (E_n)_x\right)\ d\mu(x) \\
&= \int_X \sum_{n=1}^{\infty} \nu((E_n)_x)\ d\mu(x) \\
&\overset{(*)}{=} \sum_{n=1}^{\infty} \int_X \nu((E_n)_x)\ d\mu(x) \\
&= \sum_{n=1}^{\infty} \rho(E_n).
\end{aligned}$$

In step $(*)$, we used Theorem 3.12.

Claim 3. ρ is a product measure of μ and ν.

Let $B \in \mathcal{B}$ and $C \in \mathcal{C}$. As noted previously (see (6.1)), $(B \otimes C)_x = C$ if $x \in B$ and $(B \otimes C)_x = \emptyset$ if $x \notin B$. Hence, the function $x \mapsto \nu((B \otimes C)_x)$ is a simple function, and integrating with respect to μ gives

$$\rho(B \times C) = \int_X \nu((B \otimes C)_x)\, d\mu(x) = \nu(C) \cdot \mu(B) + \nu(\emptyset) \cdot \mu(X \setminus B) = \mu(B)\nu(C).$$

(2) Now suppose μ and ν are s-finite. Let

$$\rho_1(E) = \int_X \nu(E_x)\, d\mu(x) \qquad \text{and} \qquad \rho_2(E) = \int_Y \mu(E^y)\, d\mu(y).$$

Claim 3. $\rho_1 = \rho_2$

Write $\mu = \sum_{n=1}^\infty \mu_n$ and $\nu = \sum_{n=1}^\infty \nu_n$ for some finite measures μ_n, ν_n. Then by Theorem 3.12,

$$\rho_1(E) = \sum_{m,n} \int_X \nu_n(E_x)\, d\mu_m(x) \qquad \text{and} \qquad \rho_2(E) = \sum_{m,n} \int_Y \mu_m(E^y)\, d\nu_n(y). \tag{6.2}$$

For each $m, n \in \mathbb{N}$, the measures

$$\rho_{1,m,n}(E) = \int_X \nu_n(E_x)\, d\mu_m(x) \qquad \text{and} \qquad \rho_{2,m,n}(E) = \int_Y \mu_m(E^y)\, d\nu_n(y)$$

are product measures of μ_m and ν_n (by Claims 1 and 2). But the product of (σ-)finite measures is unique, so $\rho_{1,m,n} = \rho_{2,m,n}$. Hence, by (6.2), $\rho_1 = \rho_2$.

□

Definition 6.8

Given s-finite measure spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$, we call the product measure obtained by Theorem 6.7 the *cross-sectional product measure* and denote it by $\mu \overset{cs}{\otimes} \nu$.

Theorem 6.7 extends to a result about integration of measurable functions on products of s-finite measure spaces.

Theorem 6.9: Fubini–Tonelli Theorem

Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be s-finite measure spaces.

(1) (Tonelli) Let $f : X \times Y \to [0, \infty]$ be a $(\mathcal{B} \otimes \mathcal{C})$-measurable function. Then $x \mapsto \int_Y f_x\, d\nu$ and $y \mapsto \int_X f^y\, d\mu$ are measurable functions, and

$$\int_{X \times Y} f\, d(\mu \overset{cs}{\otimes} \nu) = \int_X \left(\int_Y f(x,y)\, d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\nu(y). \tag{6.3}$$

(2) (Fubini) Suppose $f \in L^1(\mu \overset{cs}{\otimes} \nu)$. Then $f_x \in L^1(\nu)$ for μ-a.e. $x \in X$, $f_y \in L^1(\mu)$ for ν-a.e. $y \in Y$, the almost-everywhere defined functions $x \mapsto \int_Y f_x\, d\nu$ and $y \mapsto \int_X f^y\, d\mu$ belong to $L^1(\mu)$ and $L^1(\nu)$ respectively, and (6.3) holds.

Proof. **(1)** If $f = \mathbb{1}_E$ for some $E \in \mathcal{B} \otimes \mathcal{C}$, then (1) holds by Theorem 6.7. Hence, (1) holds for simple functions. For general f, let $(f_n)_{n \in \mathbb{N}}$ be a sequence of simple functions such that $0 \le f_1 \le f_2 \le \dots$ and $f_n \to f$ pointwise as in Proposition 3.7. Then $(f_n)_x$ increases to f_x

and $(f_n)^y$ increases to f^y, so (6.3) holds by repeated application of the monotone convergence theorem.

(2) Let $f \in L^1(\mu \overset{cs}{\otimes} \nu)$. By (1),

$$\int_{X\times Y} |f| \, d(\mu \overset{cs}{\otimes} \nu) = \int_X \left(\int_Y |f_x| \, d\nu \right) \, d\mu(x) = \int_Y \left(\int_X |f^y| \, d\mu \right) \, d\nu(y).$$

This integral is finite, so $\int_Y |f_x| \, d\nu < \infty$ for μ-a.e. $x \in X$ and $\int_X |f^y| \, d\mu < \infty$ for ν-a.e. $y \in Y$ by Proposition 3.20. That is, $f_x \in L^1(\nu)$ for μ-a.e. $x \in X$, and $f^y \in L^1(\mu)$ for ν-a.e. $y \in Y$. Moreover, by the triangle inequality for integrals,

$$\int_X \left| \int_Y f_x \, d\nu \right| \, d\mu(x) \leq \int_X \left(\int_Y |f_x| \, d\nu \right) \, d\mu(x) < \infty$$

and similarly for the iterated integral in the other order.

The identity (6.3) holds for the positive and negative parts of the real and imaginary parts of f by (1), and these can be recombined to conclude (6.3) for the function f itself. □

Example 6.10

Let λ be the Lebesgue measure on $\mathbb{R}$. For an integrable function f, the *Fourier transform* of f is the function $\widehat{f} : \mathbb{R} \to \mathbb{R}$ defined by $\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} \, dx$. Given integrable functions f and g, we define the *convolution* $(f * g)(x) = \int_{\mathbb{R}} f(x-y) g(y) \, dy$. Then $f * g$ is well-defined almost everywhere, integrable, and $\widehat{(f * g)}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$.

To see this, consider the function $\Phi(x, y) = f(x - y)g(y)$. Since f and g are measurable, Φ is also measurable. Moreover, by Tonelli's theorem,

$$\int_{\mathbb{R}^2} |\Phi(x,y)| \, d(x,y) = \int_{\mathbb{R}} |g(y)| \left(\int_{\mathbb{R}} |f(x-y)| \, dx \right) \, dy = \left(\int_{\mathbb{R}} |f| \, d\lambda \right) \left(\int_{\mathbb{R}} |g| \, d\lambda \right) < \infty.$$

Therefore, by Fubini's thoerem, $f * g$ is almost everywhere well-defined, integrable, and satisifies

$$\begin{aligned}
\widehat{(f * g)}(\xi) &= \int_{\mathbb{R}} (f * g)(x) e^{-2\pi i \xi x} \, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y) g(y) e^{-2\pi i \xi x} \, dy \, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y) g(y) e^{-2\pi i \xi x} \, dx \, dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(t) g(y) e^{-2\pi i \xi (t+y)} \, dt \, dy \\
&= \int_{\mathbb{R}} f(t) e^{-2\pi i \xi t} \, dt \int_{\mathbb{R}} g(y) e^{-2\pi i \xi y} \, dy \\
&= \widehat{f}(\xi)\widehat{g}(\xi).
\end{aligned}$$

As in the example above, the utility of Fubini's theorem is often interchanging the order of an iterated integral, and the product measure acts simply as an auxiliary object to justify this swap. In practice, this means that we do not need to be particularly concerned by the fact that there may be more than one product measure. The validity of the Fubini–Tonelli theorem for s-finite (and not necessarily σ-finite) measures has found applications in the theory of Markov processes [**Get90**].

2. The Maximal Product Measure

Let $(X,\mathcal{B})$ and $(Y,\mathcal{C})$ be measurable spaces. The intersection of measurable rectangles $B_1 \times C_1$ and $B_2 \times C_2$ is again a measurable rectangle: $(B_1 \times C_1) \cap (B_2 \times C_2) = (B_1 \cap B_2) \times (C_1 \cap C_2)$. The complement of a measurable rectangle is a disjoint union of three measurable rectangles: $(B \times C)^c = (B^c \times C) \sqcup (B \times C^c) \sqcup (B^c \times C^c)$. Thus, the family of measurable rectangles is a semi-algebra on $X \times Y$. We can therefore build a product measure using an outer measure construction similar to what appeared in Section 3.

We may define an algebra

$$\mathcal{A} = \left\{ \bigsqcup_{i=1}^{n} (B_i \times C_i) : n \in \mathbb{N}, B_1 \times C_1, \dots, B_n \times C_n \text{ pairwise disjoint measurable rectangles} \right\}.$$

Given measures $\mu : \mathcal{B} \to [0,\infty]$ and $\nu : \mathcal{C} \to [0,\infty]$ on X and Y respectively, we define a premeasure ρ_0 on $\mathcal{A}$ by

$$\rho_0\left(\bigsqcup_{i=1}^{n} (B_i \times C_i) \right) = \sum_{i=1}^{n} \mu(B_i)\nu(C_i).$$

We can then extend ρ_0 to an outer measure

$$\begin{aligned}\rho^*(E) &= \inf\left\{ \sum_{n=1}^{\infty} \rho_0(A_n) : E \subseteq \bigcup_{n\in\mathbb{N}} A_n, A_n \in \mathcal{A} \right\} \\ &= \inf\left\{ \sum_{n=1}^{\infty} \mu(B_n)\nu(C_n) : E \subseteq \bigcup_{n\in\mathbb{N}} (B_n \times C_n), B_n \in \mathcal{B}, C_n \in \mathcal{C} \right\}\end{aligned}$$

by Proposition 4.22. By Lemma 4.26, elements of $\mathcal{A}$ are ρ^*-measurable, and so $\rho = \rho^*|_{\mathcal{B}\otimes\mathcal{C}}$ defines a product measure by Theorem 4.25. When it is ambiguous (i.e., when dealing with non-σ-finite spaces), we will denote this product measure by $\mu \overset{max}{\otimes} \nu$ and refer to it as the *maximal product measure* of μ and ν. The reason for this terminology is the following theorem.

Theorem 6.11

Let $(X,\mathcal{B},\mu)$ and $(Y,\mathcal{C},\nu)$ be measure spaces. Then for every $E \in \mathcal{B}\otimes\mathcal{C}$,

$$(\mu \overset{max}{\otimes} \nu)(E) = \sup\{\rho(E) : \rho \text{ is a product measure of } \mu \text{ and } \nu\}.$$

Moreover, if $(\mu \overset{max}{\otimes} \nu)(E) < \infty$, then $\rho(E) = (\mu \overset{max}{\otimes} \nu)(E)$ for every product measure ρ.

Example 6.12

Let $X = [0,1]$, $\mathcal{B} = \text{Borel}([0,1])$, and let μ be the Lebesgue measure on $[0,1]$. Let $Y = [0,1]$ also but with the σ-algebra $\mathcal{C} = \mathscr{P}([0,1])$ and counting measure ν. Note that, since Y is uncountable, ν is not s-finite. However, μ is (s-)finite, so can define a cross-sectional product measure $\rho : \mathcal{B}\otimes\mathcal{C} \to [0,\infty]$ by

$$\rho(E) = \int_Y \mu(E^y)\, d\nu(y) = \sum_{y\in Y} \mu(E^y).$$

To see that this is different than the product measure $\mu \overset{max}{\otimes} \nu$, consider the diagonal $\Delta = \{(t,t) : t \in [0,1]\}$.

Using the cross-sectional product measure, we have

$$\rho(\Delta) = \sum_{y \in Y} \mu(\{y\}) = 0.$$

Now let us compute the measure $(\mu \overset{max}{\otimes} \nu)(\Delta)$. Let $(B_n \times C_n)_{n\in\mathbb{N}}$ be a family of measurable rectangles such that $\Delta \subseteq \bigcup_{n\in\mathbb{N}}(B_n \times C_n)$. Let $S = \{n \in \mathbb{N} : \mu(B_n) = 0\}$, and let $E = \bigcup_{n\in S} B_n$. Then $F = [0,1] \setminus E$ has $\mu(F) = 1$, and $\Delta_F = \{(t,t) : t \in F\}$ satisfies $\Delta_F \subseteq \bigcup_{n\notin S}(B_n \times C_n)$. Since $\mu(F) = 1$, F is uncountable. But $F \subseteq \bigcup_{n\notin S} C_n$, so C_{n_0} is uncountable (in particular, infinite) for some $n_0 \notin S$. Therefore,

$$\sum_{n=1}^{\infty} \mu(B_n)\nu(C_n) \geq \underbrace{\mu(B_{n_0})}_{>0}\underbrace{\nu(C_{n_0})}_{\infty} = \infty.$$

This proves $(\mu \overset{max}{\otimes} \nu)(\Delta) = \infty$.

Proof of Theorem 6.11. Let ρ be a product measure of μ and ν and let $E \in \mathcal{B}\otimes\mathcal{C}$. Suppose $(B_n \times C_n)_{n\in\mathbb{N}}$ is a family of measurable rectangles such that $E \subseteq \bigcup_{n\in\mathbb{N}}(B_n \times C_n)$. Then by countable subadditivity of ρ and the fact that ρ is a product measure, we have

$$\rho(E) \leq \sum_{n=1}^{\infty} \rho(B_n \times C_n) = \sum_{n=1}^{\infty} \mu(B_n)\nu(C_n).$$

Taking an infimum over such families, we conclude $\rho(E) \leq (\mu \overset{max}{\otimes} \nu)(E)$. This proves the first part of the theorem.

Suppose $E \in \mathcal{B}\otimes\mathcal{C}$ and $(\mu \overset{max}{\otimes} \nu)(E) < \infty$. Let $\varepsilon > 0$. There exists a family $(B_n \times C_n)_{n\in\mathbb{N}}$ of measurable rectangles such that $E \subseteq \bigcup_{n\in\mathbb{N}}(B_n \times C_n)$ and

$$\sum_{n=1}^{\infty} \mu(B_n)\nu(C_n) \leq (\mu \overset{max}{\otimes} \nu)(E) + \varepsilon.$$

Let $A = \bigcup_{n\in\mathbb{N}}(B_n \times C_n)$. By countable subadditivity and monotonicity,

$$(\mu \overset{max}{\otimes} \nu)(E) \leq (\mu \overset{max}{\otimes} \nu)(A) \leq \sum_{n=1}^{\infty} \mu(B_n)\nu(C_n) \leq (\mu \overset{max}{\otimes} \nu)(E) + \varepsilon.$$

In particular, $(\mu \overset{max}{\otimes} \nu)(A\setminus E) < \varepsilon$. By the first part of the theorem, it follows that $\rho(A\setminus E) < \varepsilon$.

Define a sequence $(A_n)_{n\in\mathbb{N}}$ in the algebra $\mathcal{A}$ generated by measurable rectangles by $A_1 = B_1 \times C_1$ and $A_n = (B_n \times C_n) \setminus (A_1 \cup \dots A_{n-1})$ for $n \geq 2$. Then the sets $(A_n)_{n\in\mathbb{N}}$ are pairwise disjoint and $\bigcup_{n\in\mathbb{N}} A_n = A$. Moreover, since the rectangles satisfy $\mu(B_n)\nu(C_n) < \infty$ for every $n \in \mathbb{N}$, additivity of the arbitrary product measure ρ implies that the value $\rho(A_n)$ is the same for every product measure. Therefore,

$$\rho(A) = \sum_{n=1}^{\infty} \rho(A_n) = \sum_{n=1}^{\infty} (\mu \overset{max}{\otimes} \nu)(A_n) = (\mu \overset{max}{\otimes} \nu)(A).$$

Thus,

$$\rho(E) = \rho(A) - \rho(A \setminus E) > (\mu \overset{max}{\otimes} \nu)(A) - \varepsilon \geq (\mu \overset{max}{\otimes} \nu)(E) - \varepsilon.$$

Combining with the first part of the theorem, we conclude $\rho(E) = (\mu \overset{max}{\otimes} \nu)(E)$. □

Part 3

Additional Topics in Measure Theory

CHAPTER 7

L^p Spaces

Learning Objectives

At the end of this chapter, you will be able to:

- Define topological vector spaces, normed vector spaces, and Banach spaces and explain the relationship between these objects.
- Define convexity of sets and functions and describe the implications of convexity for L^p spaces.
- Prove that L^p spaces are complete.

1. Topological Vector Spaces

Definition 7.1

A *topological vector space* is a pair (V, τ) such that V is a (real or complex) vector space and τ is a topology on V such that

- the addition map $V \times V \to V$, $(u, v) \mapsto u + v$, is continuous, and
- the map $\mathbb{K} \times V \to V$, $(c, v) \mapsto cv$ is continuous, where $\mathbb{K}$ is the field of scalars.

Example 7.2

The Euclidean space $V = \mathbb{R}^d$ with the standard topology is a topological vector space for $d \in \mathbb{N}$.

A typical means of defining a topology on a vector space is with a norm. Recall (Definition 3.17) that a *norm* on a vector space V is a function $\|\cdot\| : V \to [0, \infty)$ such that

- TRIANGLE INEQUALITY: $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$;
- ABSOLUTE HOMOGENEITY: $\|cv\| = |c| \, \|v\|$ for all $v \in V$ and all scalars c; and
- POSITIVE DEFINITENESS: if $v \in V$ and $\|v\| = 0$, then $v = 0$.

The triangle inequality and positive definiteness imply that the function $d : V \times V \to [0, \infty)$ defined by $d(u, v) = \|u - v\|$ is a metric on V, so we can endow V with this metric space topology.

Definition 7.3

A normed vector space $(V, \|\cdot\|)$ is a *Banach space* if it is complete (as a metric space).

Example 7.4

The standard Euclidean norm on $\mathbb{R}^d$, i.e. $\|(x_1, \ldots, x_d)\| = \sqrt{x_1^2 + \cdots + x_d^2}$, makes $\mathbb{R}^d$ into a Banach space.

One way of studying normed spaces (or more general topological vector spaces) is in terms of linear functionals. We have already seen the utility of understanding linear functionals on a normed space in the Riesz representation theorem, which described the positive linear functionals on the normed space $C_c(X)$ as Radon measures on the underlying LCH space X.

Definition 7.5

Let $\mathbb{K}$ be either $\mathbb{R}$ or $\mathbb{C}$. Let $(V, \|\cdot\|_V)$ be a normed vector space over $\mathbb{K}$. A *linear functional* is a linear map $\varphi : V \to \mathbb{K}$. The *dual space* V^* is the vector space of continuous linear functionals on V with norm

$$\|\varphi\|_{V^*} = \sup_{\|v\|_V \leq 1} |\varphi(v)|.$$

Proposition 7.6

Let V be a normed vector space. Then the dual V^* is a Banach space.

Proof. We need to prove two things: (1) $\|\cdot\|_{V^*}$ defines a norm on V^* and (2) $(V^*, \|\cdot\|_{V^*})$ is complete.

Claim 1. $\|\cdot\|_{V^*}$ is a norm on V^*.

Let $\varphi, \psi \in V^*$. Then for $v \in V$ with $\|v\|_V \leq 1$, we have

$$|(\varphi + \psi)(v)| \leq |\varphi(v)| + |\psi(v)| \leq \|\varphi\|_{V^*} + \|\psi\|_{V^*},$$

so $\|\cdot\|_{V^*}$ satisfies the triangle inequality.

Suppose $\varphi \in V^*$ and $c \in \mathbb{K}$. Then for $v \in V$, we have $|(c\varphi)(v)| = |c\varphi(v)| = |c||\varphi(v)|$. Taking a supremum over $v \in V$ with $\|v\|_V \leq 1$ gives absolute homogeneity of $\|\cdot\|_{V^*}$.

Finally, if $\|\varphi\|_{V^*} = 0$, then $\varphi(v) = 0$ for every $v \in V$ with $\|v\|_V \leq 1$. Hence, for arbitrary $v \in V \setminus \{0\}$, we have $\varphi(v) = \|v\|_V \, \varphi\left(\frac{v}{\|v\|_V}\right) = 0$, so $\varphi = 0$. Thus, $\|\cdot\|_{V^*}$ is positive definite.

Claim 2. $(V^*, \|\cdot\|_{V^*})$ is complete.

Suppose $(\varphi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in V^*. Then for $v \in V$, the sequence $(\varphi_n(v))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{K}$, since $|\varphi_n(v) - \varphi_m(v)| \leq \|\varphi_n - \varphi_m\|_{V^*} \|v\|_V$. Since the field $\mathbb{K}$ is complete, we may define $\varphi(v) = \lim_{n\to\infty} \varphi_n(v)$ for every $v \in V$. We must check that $\varphi \in V^*$ and $\varphi_n \to \varphi$ in V^*.

First, $\|\varphi\|_{V^*} \leq \lim_{n\to\infty} \|\varphi_n\|_{V^*} < \infty$, so $\varphi \in V^*$ (see Exercise **??**). Next, given $v \in V$ with $\|v\|_V \leq 1$, we have

$$|\varphi(v) - \varphi_n(v)| = |\lim_{m\to\infty} \varphi_m(v) - \varphi_n(v)| \leq \sup_{m\geq n} |\varphi_m(v) - \varphi_n(v)| \leq \sup_{m\geq n} \|\varphi_m - \varphi_n\|_{V^*}.$$

Thus,

$$\limsup_{n\to\infty} \|\varphi - \varphi_n\|_{V^*} \leq \lim_{n\to\infty} \sup_{m\geq n} \|\varphi_m - \varphi_n\|_{V^*} = 0,$$

so $\varphi_n \to \varphi$.

□

2. L^p Norms

For the remainder of this chapter, we will focus on vector spaces of functions associated with a measure space.

Definition 7.7

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

- For $1 \le p < \infty$, the *L^p norm* of a measurable function $f : X \to \mathbb{C}$ is the quantity
$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p}.$$
- The *(μ-)essential supremum* of a nonnegative measurable function $f : X \to [0, \infty]$ is
$$\text{ess sup}(f) = \inf \{c \ge 0 : f \le c \ \mu\text{-a.e.}\} = \inf \{c \ge 0 : \mu(\{f > c\}) = 0\}.$$
- The *L^∞ norm* of a measurable function $f : X \to \mathbb{C}$ is $\|f\|_\infty = \text{ess sup}(|f|)$.
- For $1 \le p \le \infty$, the *L^p space* $L^p(\mu)$ is the space
$$L^p(\mu) = \left\{ [f]_\mu : \|f\|_p < \infty \right\},$$
where $[f]_\mu$ is the equivalence class $[f]_\mu = \{g : X \to \mathbb{C} \text{ measurable} : g = f \ \mu\text{-a.e.}\}$.

Remark. If $f, g : X \to \mathbb{C}$ are measurable functions and $f = g$ μ-a.e., then $\int_X |f|^p \, d\mu = \int_X |g|^p \, d\mu$, so the value of the L^p norm only depends on the equivalence class and not the choice of representative.

While for technical reasons, we view $L^p(\mu)$ as a space of equivalence classes of functions, it is often more convenient to work with actual functions (i.e., representatives of the equivalence classes). As long as we only perform "countable operations" on functions, we can safely pass between the two different points of view, since the family of null sets is a σ-ideal. For this reason, it is standard practice in mathematics to write expressions like "$f \in L^p(\mu)$," and we will also engage in this abuse of notation.

Lemma 7.8

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f : X \to \mathbb{C}$ be a measurable function. Then $|f| \le \|f\|_\infty$ μ-a.e.

Proof. Let $M = \|f\|_\infty$. If $M = \infty$, there is nothing to show, so assume $M < \infty$. By the definition of the essential supremum, we have $\mu(\{|f| > M + \frac{1}{n}\}) = 0$ for each $n \in \mathbb{N}$. Taking a union over $n \in \mathbb{N}$ and applying continuity from below of the measure μ, we conclude $\mu(\{|f| > M\}) = 0$. That is, $|f| \le M$ a.e. □

One of the main goals of this chapter is to prove that $\|\cdot\|_p$ defines a norm for every $p \in [1, \infty]$ and the vector space $L^p(\mu)$ is a Banach space.

3. Convexity and the Inequalities of Jensen and Minkowski

Definition 7.9

Let V be a real vector space.

- A set $C \subseteq V$ is *convex* if it contains the entire line segment between every pair of points in C. That is, for every $x, y \in C$ and every $t \in [0,1]$, one has $tx + (1-t)y \in C$.
- Let C be a convex set. A function $\varphi : C \to \mathbb{R}$ is *convex* if $\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ for all $x, y \in C$ and $t \in [0,1]$.

By induction, one can show that convex sets are closed under convex combinations: if $x_1, \ldots, x_n \in C$ and $\lambda_1, \ldots, \lambda_n \geq 0$ are coefficients with $\sum_{j=1}^n \lambda_j = 1$, then $\sum_{j=1}^n \lambda_j x_j \in C$. Similarly, if f is a convex function, then its behavior on convex combinations is governed by the inequality $\varphi\left(\sum_{j=1}^n \lambda_j x_j\right) \leq \sum_{j=1}^n \lambda_j \varphi(x_j)$. This is the discrete version of a fundamental inequality for convex functions known as Jensen's inequality.

Theorem 7.10: Jensen's Inequality

Let $(X, \mathcal{B}, \mu)$ be a probability space. Let $f : X \to I$ be an integrable function taking values in an interval $I \subseteq \mathbb{R}$, and suppose $\varphi : I \to \mathbb{R}$ is convex. Then

$$\varphi\left(\int_X f \, d\mu\right) \leq \int_X \varphi \circ f \, d\mu.$$

Proof. Let $t = \int_X f \, d\mu \in I$. We want to show $\int_X \varphi \circ f \, d\mu \geq \varphi(t)$.

If $t \in \partial I$, then $f = t$ a.e., so $\varphi \circ f = \varphi(t)$ a.e. and $\int_X \varphi \circ f \, d\mu = \varphi(t)$.

Suppose $t \in \operatorname{int}(I)$. By Exercise **??**, let $m \in \mathbb{R}$ such that

$$\varphi(s) \geq m(s-t) + \varphi(t) \quad (\forall s \in I).$$

Then

$$\int_X \varphi \circ f \, d\mu \geq \int_X \left(m(f-t) + \varphi(t)\right) \, d\mu = m \underbrace{\int_X f \, d\mu}_{t} - mt + \varphi(t) = \varphi(t).$$

□

The central inequality for the L^p norm is Minkowski's inequality, which shows that $\|\cdot\|_p$ is indeed a norm.

Theorem 7.11: Minkowski's Inequality

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $p \in [1, \infty]$. Let $f, g : X \to \mathbb{C}$ be measurable functions. Then

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

Before proving Minkowski's inequality, let us confirm that it makes $(L^p(\mu), \|\cdot\|_p)$ into a normed vector space.

Corollary 7.12

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $p \in [1, \infty]$. Then $L^p(\mu)$ is a vector space, and $\|\cdot\|_p : L^p(\mu) \to [0, \infty)$ is a norm.

Proof. Let $f, g \in L^p(\mu)$. (Technically, we should say $[f]_\mu, [g]_\mu \in L^p(\mu)$, but this can become notationally cumbersome and is not standard practice in mathematics; see the remark following Definition 7.7.) By Minkowski's inequality, $\|f+g\|_p \le \|f\|_p + \|g\|_p < \infty$, so $f+g \in L^p(\mu)$. Moreover, if $p < \infty$, then given $f \in L^p(\mu)$ and $c \in \mathbb{C}$, we have

$$\|cf\|_p = \left(\int_X |cf|^p \, d\mu \right)^{1/p} = \left(|c|^p \int_X |f|^p \, d\mu \right)^{1/p} = |c| \, \|f\|_p < \infty, \tag{7.1}$$

so $cf \in L^p(\mu)$. If $p = \infty$, then we can similarly check that $\|cf\|_\infty = |c| \, \|f\|_\infty$. This proves that $L^p(\mu)$ is a vector space. Minkowski's inequality is the triangle inequality for the L^p norm, and (7.1) proves absolute homogeneity. Finally, $\|\cdot\|_p$ is positive definite by Proposition 3.23 in the case $p < \infty$ and by Lemma 7.8 when $p = \infty$. □

Remark. The final line of the proof of Corollary 7.12 reveals why we want to view L^p as a space of equivalence classes rather than a space of functions: on the space of functions, $\|\cdot\|_p$ is only a seminorm, but working with equivalence classes of functions turns $\|\cdot\|_p$ into a proper norm.

Proof of Minkowski's Inequality. We split the proof into two cases: $p < \infty$ or $p = \infty$.

Case 1. $p < \infty$

Let $B = \{w : X \to \mathbb{C} \text{ measurable} : \|w\|_p \le 1\}$. We will show that B is a convex set. Let $u, v \in B$ and $t \in [0,1]$. Then

$$\begin{aligned}
\int_X |tu + (1-t)v|^p \, d\mu &\le \int_X (t|u| + (1-t)|v|)^p \, d\mu && \text{(triangle inequality)} \\
&\le \int_X (t|u|^p + (1-t)|v|^p) \, d\mu && (x \mapsto x^p \text{ is convex}) \\
&= t \, \|u\|_p^p + (1-t) \, \|v\|_p^p \le 1.
\end{aligned}$$

Now let $f, g : X \to \mathbb{C}$ be measurable functions. If $\|f\|_p = \infty$ or $\|g\|_p = \infty$, there is nothing to prove. Moreover, if $\|f\|_p = 0$, then $f = 0$ a.e., so $f + g = g$ a.e., and there is nothing to prove. Similarly, there is nothing to show in the case that $\|g\|_p = 0$. Thus, we may assume $0 < \|f\|_p, \|g\|_p < \infty$. Then $u = \frac{f}{\|f\|_p}, v = \frac{g}{\|g\|_p} \in B$. Hence, letting $t = \frac{\|f\|_p}{\|f\|_p + \|g\|_p}$, we have

$$\frac{\|f+g\|_p}{\|f\|_p + \|g\|_p} = \|tu + (1-t)v\|_p \le 1.$$

Case 2. $p = \infty$

By Lemma 7.8, the sets $N_1 = \{|f| > \|f\|_\infty\}$ and $N_2 = \{|g| > \|g\|_\infty\}$ are μ-null sets. Therefore, $N = N_1 \cup N_2$ is a null set. Suppose $x \in X \setminus N$. Then

$$|f(x) + g(x)| \le |f(x)| + |g(x)| \le \|f\|_\infty + \|g\|_\infty .$$

Therefore, $\{|f+g| > \|f\|_\infty + \|g\|_\infty\} \subseteq N$ is a null set, so $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$.

□

4. Riesz–Fischer Theorem

Our next goal is prove that $(L^p(\mu), \|\cdot\|_p)$ is a Banach space.

Theorem 7.13

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $1 \le p < \infty$. Suppose $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence $L^p(\mu)$. Then there is a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ that converges a.e. to a function $f \in L^p(\mu)$ and $\|f_n - f\|_p \to 0$.

Proof. The strategy of the proof is to choose a subsequence along which f_{n_k} is "quickly Cauchy" and then to show that "quickly Cauchy" sequences converge a.e. and in L^p.

Claim 1. There exists a subsequence $n_1 < n_2 < \dots$ such that

$$\left\|f_{n_{k+1}} - f_{n_k}\right\|_p < 2^{-k} \tag{7.2}$$

for every $k \in \mathbb{N}$.

We construct $(n_k)_{k\in\mathbb{N}}$ by induction. Choose $n_1 \in \mathbb{N}$ such that

$$\sup_{m\ge n_1} \|f_{n_1} - f_m\|_p < \frac{1}{2}.$$

Given $n_1, \dots, n_k$, choose $n_{k+1} > n_k$ such that

$$\sup_{m\ge n_{k+1}} \left\|f_{n_{k+1}} - f_m\right\|_p < 2^{-(k+1)}.$$

Then by induction, we have constructed a sequence $n_1 < n_2 < \dots$ satisfying (7.2).

Let $g_k = f_{n_{k+1}} - f_{n_k}$ for $k \in \mathbb{N}$, and let $G_k = \sum_{j=1}^k |g_j|$. By Minkowski's inequality, $\|G_k\|_p \le \sum_{j=1}^k \|g_j\|_p < 1$. Letting $G = \sum_{j=1}^\infty |g_j|$, we have by the monotone convergence theorem that

$$\int_X G^p \, d\mu = \lim_{k\to\infty} \underbrace{\int_X G_k^p \, d\mu}_{\|G_k\|_p^p} \le 1.$$

In particular, $G < \infty$ a.e., so the series $g(x) = \sum_{j=1}^\infty g_j(x)$ converges absolutely for a.e. $x \in X$. Let $f = g + f_{n_1}$.

Claim 2. $f_{n_k} \to f$ a.e as $k \to \infty$

Let $S_k = \sum_{j=1}^{k-1} g_j$. Then $S_k \to g$ a.e. (by the definition of g), so $f_{n_1} + S_k \to f$ a.e. as $k \to \infty$. But expanding g_j, the sum S_k is telescoping and we have

$$f_{n_1} + S_k = f_{n_1} + (f_{n_2} - f_{n_1}) + \dots + (f_{n_k} - f_{n_{k-1}}) = f_{n_k}.$$

Claim 3. $\|f_n - f\|_p \to 0$.

Note that $|f_{n_k} - f| = |\sum_{j=k+1}^{\infty} g_j| \leq \sum_{j=k+1}^{\infty} |g_j|$, so by Minkowski's inequality, for $n \geq n_k$,

$$\|f_n - f\|_p \leq \|f_n - f_{n_k}\|_p + \|f_{n_k} - f\|_p \leq 2^{-k} + \sum_{j=k+1}^{\infty} 2^{-j} = 2 \cdot 2^{-k}$$

Therefore, $\|f_n - f\|_p \to 0$ as $n \to \infty$.

Claim 4. $f \in L^p(\mu)$

By Minkowski's inequality, $\|f\|_p \leq \|f_n\|_p + \|f_n - f\|_p$. The term $\|f_n\|_p$ is finite, and $\|f_n - f\|_p < 1$ for all large enough n.

□

Theorem 7.14

Let $(X, \mathcal{B}, \mu)$ be a measure space. Suppose $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence $L^\infty(\mu)$. Then there exists a measurable set $X_0 \subseteq X$ such that $\mu(X \setminus X_0) = 0$ and $(f_n)_{n\in\mathbb{N}}$ is uniformly bounded and uniformly Cauchy on X_0. In particular, $(f_n)_{n\in\mathbb{N}}$ converges a.e. and in L^∞ to a function $f \in L^\infty(\mu)$.

Proof. Put

$$X_0 = \underbrace{\bigcap_{n\in\mathbb{N}} \{|f_n| \leq \|f_n\|_\infty\}}_{\text{uniformly bounded}} \cap \underbrace{\bigcap_{n,m\in\mathbb{N}} \{|f_n - f_m| \leq \|f_n - f_m\|_\infty\}}_{\text{uniformly Cauchy}}.$$

Since $(f_n)_{n\in\mathbb{N}}$ is uniformly Cauchy on X_0, we may define $f(x) = \lim_{n\to\infty} f_n(x)$ as the (uniform) limit of f_n on X_0. The values of f outside of X_0 will not play any role, so let us set $f(x) = 0$ for $x \in X \setminus X_0$. Then f is measurable and uniformly bounded, so $f \in L^\infty(\mu)$. Moreover, $f_n \to f$ uniformly outside of the null set $X \setminus X_0$, so $\|f_n - f\|_\infty \leq \sup_{x\in X_0} |f_n(x) - f(x)| \to 0$. □

Combining Theorem 7.13 (for $p < \infty$) and Theorem 7.14 (for $p = \infty$), we obtain the Riesz–Fischer theorem.

Theorem 7.15: Riesz–Fischer Theorem

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $1 \leq p \leq \infty$. Then $(L^p(\mu), \|\cdot\|_p)$ is a Banach space.

5. Hölder's Inequality

Lemma 7.16: Weighted Arithmetic Mean-Geometric Mean Inequality

Let $n \in \mathbb{N}$, $x_1, \ldots, x_n \geq 0$, and $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\sum_{j=1}^{n} \lambda_j = 1$. Then

$$\prod_{j=1}^{n} x_j^{\lambda_j} \leq \sum_{j=1}^{n} \lambda_j x_j.$$

Proof. The function $x \mapsto \log x$ is strictly increasing, so if suffices to prove the inequality after taking the logarithm of both sides. But $x \mapsto -\log x$ is also a convex function, so by Jensen's inequality,

$$-\log\left(\sum_{j=1}^n \lambda_j x_j\right) \leq -\sum_{j=1}^n \lambda_j \log x_j = -\log\left(\prod_{j=1}^n x_j^{\lambda_j}\right).$$

□

Theorem 7.17: Hölder's Inequality

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $f, g : X \to \mathbb{C}$ be measurable functions. Let $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

More generally, if $n \in \mathbb{N}$, $f_1, \ldots, f_n : X \to \mathbb{C}$ are measurable functions, and $p_1, \ldots, p_n \in [1, \infty]$ such that $\sum_{j=1}^n \frac{1}{p_j} = 1$, then

$$\left\|\prod_{j=1}^n f_j\right\|_1 \leq \prod_{j=1}^n \|f_j\|_{p_j}.$$

Proof. Let $n \in \mathbb{N}$, let $f_1, \ldots, f_n : X \to \mathbb{C}$ be measurable functions, and suppose $p_1, \ldots, p_n \in [1, \infty]$ such that $\sum_{j=1}^n \frac{1}{p_j} = 1$. If $\|f_j\|_{p_j} = 0$ for some j, then $f_j = 0$ a.e., so $\prod_{j=1}^n f_j = 0$ a.e., and there is nothing to prove. We may therefore assume $\|f_j\|_{p_j} > 0$ for all j. Now if $\|f_j\|_{p_j} = \infty$ for some j, then $\prod_{j=1}^n \|f_j\|_{p_j} = \infty$, and again there is nothing to prove, so we may assume $0 < \|f_j\|_{p_j} < \infty$.

For each j, let $u_j = \frac{f_j}{\|f_j\|_{p_j}}$ so that $\|f_j\|_{p_j} = 1$. Let $J = \{1 \leq j \leq n : p_j = \infty\}$, and note that $|u_j| \leq 1$ a.e. for $j \in J$. Moreover, $\sum_{j \notin J} \frac{1}{p_j} = 1$, so by Lemma 7.16,

$$\frac{\left\|\prod_{j=1}^n f_j\right\|_1}{\prod_{j=1}^n \|f_j\|_{p_j}} \leq \int_X \prod_{j=1}^n |u_j| \, d\mu = \int_X \prod_{j \notin J} \left(|u_j|^{p_j}\right)^{1/p_j} \, d\mu \leq \sum_{j \notin J} \frac{1}{p_j} \underbrace{\int_X |u_j|^{p_j} \, d\mu}_{\|u_j\|_{p_j}^{p_j} = 1} = 1.$$

□

Definition 7.18

The exponents $p \in [1, \infty]$ and $q \in [1, \infty]$ are called *conjugate* if $\frac{1}{p} + \frac{1}{q} = 1$.

The only self-conjugate exponent is $p = 2$, and this provides the space $L^2(\mu)$ with the additional structure of an inner product space. Namely, defining $\langle \cdot, \cdot \rangle : L^2(\mu) \times L^2(\mu) \to \mathbb{C}$ by $\langle f, g \rangle = \int_X f\overline{g} \, d\mu$, we have the following properties:

- CONJUGATE SYMMETRY: $\langle f, g \rangle = \overline{\langle g, f \rangle}$;
- LINEARITY IN THE FIRST ARGUMENT: $\langle cf + g, h \rangle = c \langle f, h \rangle + \langle g, h \rangle$;
- POSITIVE DEFINITENESS: $f \neq 0 \implies \langle f, f \rangle > 0$.

The norm induced by this inner product is the L^2 norm; that is, $\|f\|_2 = \langle f, f \rangle$. The special case of Hölder's inequality for L^2 functions is the *Cauchy–Schwarz inequality*: $|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$. Complete inner product spaces (such as $L^2(\mu)$) are called *Hilbert spaces* and have many advantages over general normed spaces. For example, one can discuss orthogonality of elements of $L^2(\mu)$ and work with *orthonormal bases* for the space. Also, Hilbert spaces are self-dual in the sense that every linear functional on a Hilbert space can be represented as an inner product with a fixed element of the Hilbert space. We will not discuss Hilbert spaces in more detail in this course, but they play an important role in functional analysis and have many applications in physics.

6. Convolutions and Young's Inequality

Definition 7.19

Let $f, g : \mathbb{R} \to \mathbb{C}$ be Lebesgue-measurable functions. The *convolution* of f and g is defined by

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y)\, dy$$

when $y \mapsto f(x - y)g(y)$ is integrable.

We saw in the section on Fubini's theorem that the convolution of integrable functions is defined a.e. and integrable (see Example 6.10). Young's inequality provides a substantial generalization for convolutions of functions from other L^p spaces.

Theorem 7.20: Young's Inequality

Suppose $1 \le p, q, r \le \infty$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Let $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$. Then $f * g$ is defined a.e. and

$$\|f * g\|_r \le \|f\|_p \|g\|_q .$$

Proof. When $r = \infty$ (i.e., p and q are conjugate), then $f * g$ is defined everywhere and is bounded by $\|f\|_p \|g\|_q$ by Hölder's inequality.

For $r < \infty$, defining new constants s, t by $\frac{1}{s} = 1 - \frac{1}{q}$ and $\frac{1}{t} = 1 - \frac{1}{p}$, we have the identity

$$ab = (a^p b^q)^{1/r} (a^p)^{1/s} (b^q)^{1/t} \tag{7.3}$$

for $a, b \ge 0$. Moreover, $\frac{1}{r} + \frac{1}{s} + \frac{1}{t} = 1$, so the generalized Hölder inequality applies. We leave the details of the computations as an exercise.

□

CHAPTER 8

Regularity and Littlewood's Principles

Learning Objectives

At the end of this chapter, you will be able to:

- Give several instantiations of Littlewood's principles in different contexts.
- Use Littlewood's principles for problem solving.

The British mathematician J. E. Littlewood laid out three principles for real analysis [**Lit44**, Section 4.1]:

> *There are three principles, roughly expressible in the following terms: Every (measurable) set is nearly a finite sum of intervals; every function (of class L^p) is nearly continuous; every convergent sequence of functions is nearly uniformly convergent.*

Littlewood's principles were formulated in the context of the Lebesgue measure on $\mathbb{R}$, but they are in fact a useful guide to measure theory on very general spaces. To make sense of "intervals" or of functions being "nearly continuous" requires a topology, so we will restrict to LCH spaces with Radon measures to address the first and second principle. However, the third principle can be substantiated in fully general measure spaces.

1. The First Principle: Approximation of Measurable Sets

Littlewood's first principle can be recast as a statement about regularity of measures. Let us first interpret it for the Lebesgue measure on $\mathbb{R}$ and then try to generalize. By Proposition 4.32, if $E \subseteq \mathbb{R}$ is a Lebesgue-measurable set, then for every $\varepsilon > 0$, there exists an open set $U \subseteq \mathbb{R}$ such that $E \subseteq U$ and $\lambda(U \setminus E) < \varepsilon$. Approximating open sets by finite unions of intervals, we deduce the following statement of Littlewood's first principle (in the slightly more general context of Lebesgue–Stieltjes measures).

Proposition 8.1

Let μ be a Lebesgue–Stieltjes measure on $\mathbb{R}$, and suppose $E \subseteq \mathbb{R}$ is a μ-measurable set with $\mu(E) < \infty$. Then for every $\varepsilon > 0$, there is a finite family of disjoint open intervals $(a_1, b_1), \ldots, (a_n, b_n)$ such that

$$\mu\left(E \triangle \bigsqcup_{j=1}^{n} (a_j, b_j)\right) < \varepsilon.$$

Proof. Suppose E is μ-measurable with $\mu(E) < \infty$. Let $\varepsilon > 0$. By outer regularity of μ (see Proposition 4.32), let $U \subseteq \mathbb{R}$ be an open set such that $E \subseteq \mathbb{R}$ and $\mu(U \setminus E) < \frac{\varepsilon}{2}$. Now we may write U as a countable union of disjoint open intervals, say $U = \bigsqcup_{j=1}^{\infty} (a_j, b_j)$. Then applying continuity of μ from below, there exists $n \in \mathbb{N}$ such that $\mu\left(\bigsqcup_{j=1}^n (a_j, b_j)\right) > \mu(U) - \frac{\varepsilon}{2}$. Thus,

$$\mu\left(E \triangle \bigsqcup_{j=1}^n (a_j, b_j)\right) \leq \mu(U \setminus E) + \mu\left(U \setminus \bigsqcup_{j=1}^n (a_j, b_j)\right) < \varepsilon.$$

□

One application of Proposition 8.1 that we have already seen in the exercises is Steinhuas's theorem.

Theorem 8.2: Steinhaus's Theorem

Let λ be the Lebesgue measure on $\mathbb{R}$. Suppose $E \subseteq \mathbb{R}$ is a Lebesgue-measurable set and $\lambda(E) > 0$. Then $E - E = \{x - y : x, y \in E\}$ contains an open interval around 0.

Another application is the Riemann–Lebesgue lemma from Fourier analysis.

Theorem 8.3: Riemann–Lebesgue Lemma

Let $f \in L^1(\mathbb{R})$, and define $\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x}\, dx$ for $\xi \in \mathbb{R}$. Then $\lim_{|\xi| \to \infty} \widehat{f}(\xi) = 0$.

Proof. First suppose $f = \mathbb{1}_I$, where $I = (a, b)$ is an open interval. Then for $\xi \neq 0$,

$$\widehat{f}(\xi) = \int_a^b e^{-2\pi i \xi x}\, dx = \frac{e^{-2\pi i \xi b} - e^{-2\pi i \xi a}}{-2\pi i \xi}.$$

In particular, $|\widehat{f}(\xi)| \leq \frac{1}{\pi|\xi|}$, so $\widehat{f}(\xi) \to 0$ as $|\xi| \to \infty$. By linearity, the conclusion also holds for finite (disjoint) unions of open intervals.

Now suppose $f = \mathbb{1}_E$ for an arbitrary measurable set $E \subseteq \mathbb{R}$ with $\lambda(E) < \infty$. Let $\varepsilon > 0$. By Proposition 8.1, there is a set J that is a finite union of open intervals such that $\lambda(E \triangle J) < \varepsilon$. Hence,

$$|\widehat{f}(\xi)| = \left|\int_{\mathbb{R}} \mathbb{1}_E(x) e^{-2\pi i \xi x}\, dx\right| \leq \underbrace{\left|\int_{\mathbb{R}} \mathbb{1}_J(x) e^{-2\pi i \xi x}\, dx\right|}_{\to 0 \text{ as } |\xi| \to 0} + \underbrace{\int_{\mathbb{R}} |\mathbb{1}_E(x) - \mathbb{1}_J(x)|\, dx}_{\lambda(E \triangle J) < \varepsilon}.$$

Therefore, by linearity, the conclusion holds for integrable simple functions.

Finally, suppose $f \in L^1(\mathbb{R})$ is an arbitrary integrable function. Let $\varepsilon > 0$. Then there is a simple function $s \in L^1(\mathbb{R})$ such that $\|f - s\|_1 < \varepsilon$, so

$$|\widehat{f}(\xi)| \leq \underbrace{|\widehat{s}(\xi)|}_{\to 0 \text{ as } |\xi| \to \infty} + \underbrace{|\widehat{(f - s)}(\xi)|}_{\leq \|f - s\|_1 < \varepsilon}.$$

□

Littlewood's first principle has a natural generalization to second countable LCH spaces, which can be used to prove versions of Steinhaus's theorem and the Riemann–Lebesgue lemma in more general topological groups. (We will work with second countable spaces in order to have sets playing the role of intervals.) As preparation for stating a generalized form of Proposition 8.1, let us revisit regularity properties of Radon measures in the context of second countable spaces.

Proposition 8.4

Let X be a locally compact Hausdorff space, and let μ be a Radon measure on X. Then μ is inner regular on all σ-finite sets. That is, if $E \subseteq X$ is a Borel set and $E = \bigcup_{n\in\mathbb{N}} E_n$, where each set E_n is a Borel set with $\mu(E_n) < \infty$, then

$$\mu(E) = \sup\{\mu(K) : K \text{ is compact and } K \subseteq E\}. \tag{8.1}$$

Proof. We will first handle the case that $\mu(E) < \infty$. Let $\varepsilon > 0$. By outer regularity of μ, there is an open set $U \supseteq E$ such that $\mu(U) < \mu(E) + \frac{\varepsilon}{2}$. Applying outer regularity again, we may find an open set $V \supseteq U \setminus E$ such that $\mu(V) < \frac{\varepsilon}{2}$. Since μ is inner regular on open sets by assumption, let $K \subseteq U$ be a compact set with $\mu(K) > \mu(U) - \frac{\varepsilon}{2}$. Then $K \setminus V$ is a compact subset of E, and

$$\mu(K \setminus V) \geq \mu(K) - \mu(V) > \mu(U) - \varepsilon \geq \mu(E) - \varepsilon.$$

But ε was arbitrary, so $\mu(E) \leq \sup\{\mu(K) : K \subseteq E \text{ compact}\}$.

Now we handle the general σ-finite case. Write $E = \bigcup_{n\in\mathbb{N}} E_n$ with $E_n \subseteq X$ Borel and $\mu(E_n) < \infty$ for each $n \in \mathbb{N}$. By the finite case above, there exist compact subsets $K_n \subseteq E_n$ with $\mu(E_n \setminus K_n) < 2^{-n}$. For each $N \in \mathbb{N}$, we then have that $\bigcup_{n=1}^N K_n$ is a compact subset of E, and

$$\mu\left(\bigcup_{n=1}^N K_n\right) \geq \mu\left(\bigcup_{n=1}^N E_n\right) - \sum_{n=1}^N \mu(E_n \setminus K_n) > \mu\left(\bigcup_{n=1}^N E_n\right) - 1.$$

Moreover, $\sup_{N\in\mathbb{N}} \mu\left(\bigcup_{n=1}^N E_n\right) = \mu(E) = \infty$, so $\sup_{N\in\mathbb{N}} \mu\left(\bigcup_{n=1}^N K_n\right) = \infty$ as desired. □

In second countable LCH spaces, locally finite measures are σ-finite, so Proposition 8.4 says that Radon measures are inner regular on all Borel subsets of second countable LCH spaces. (Recall that the definition of a Radon measure only requires inner regularity on the open subsets of X.) When a measure is both inner and outer regular on all Borel sets, we say that it is a *regular* Borel measure. In Proposition 4.32, we showed that every locally finite measure on $\mathbb{R}$ is regular. This extends to second countable LCH spaces.

Theorem 8.5

Let X be a second countable LCH space or, more generally, an LCH space in which every open set is σ-compact. Then every locally finite Borel measure on X is regular.

Proof. Let X be an LCH space in which every open set is σ-compact, and let μ be a locally finite Borel measure on X. Since X is σ-compact by assumption, the measure μ is σ-finite, so it suffices by Proposition 8.4 to show that μ is a Radon measure.

Because μ is locally finite, compactly supported continuous functions are μ-integrable, so we may define a positive linear function $I_\mu : C_c(X) \to \mathbb{C}$ by $I_\mu(f) = \int_X f \, d\mu$. By the Riesz representation theorem, there is a unique Radon measure ν such that $I_\mu(f) = \int_X f \, d\nu$ for every $f \in C_c(X)$. Our goal is thus to show $\mu = \nu$ so that μ is Radon.

Claim 1. If U is open, then $\mu(U) = \nu(U)$.

Let $U \subseteq X$ be open. We may write $U = \bigcup_{n\in\mathbb{N}} K_n$ for some compact sets K_n. Let $f_1 \in C_c(X)$ with $K_1 \prec f_1 \prec U$. Then construct inductively $f_n \in C_c(X)$ such that $\bigcup_{j=1}^{n-1} \operatorname{supp}(f_j) \cup K_n \prec f_n \prec U$. Then $0 \leq f_1 \leq f_2 \leq \dots$, and $f_n \to \mathbb{1}_U$ pointwise. Thus, by the monotone convergence theorem,

$$\mu(U) = \lim_{n\to\infty} \int_X f_n \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\nu = \nu(U).$$

Write $X = \bigcup_{n\in\mathbb{N}} K_n$ with K_n compact. Then there are open sets V_n such that $\overline{V}_n$ is compact and $K_n \subseteq V_n$ by Lemma 5.8. Let $X_n = \bigcup_{j=1}^{n} V_j$ so that $X_1 \subseteq X_2 \subseteq \dots$ is an increasing family of open sets with compact closure such that $\bigcup_{n\in\mathbb{N}} X_n = X$.

Claim 2. The family $\mathcal{L} = \{E \in \operatorname{Borel}(X) : \mu(E \cap X_n) = \nu(E \cap X_n) \text{ for every } n \in \mathbb{N}\}$ is a λ-system.

The proof is the same as Claim 1 in Corollary 4.14.

The family of open subsets of a topological space is a π-system, so combining Claims 1 and 2 with the π-λ theorem, we conclude that $\mu(E \cap X_n) = \nu(E \cap X_n)$ for every Borel set $E \subseteq X$ and every $n \in \mathbb{N}$. Applying continuity from below, it follows that $\mu = \nu$. □

Corollary 8.6

Let X be a second countable LCH space, and let $\mu : \operatorname{Borel}(X) \to [0, \infty]$ be a locally finite measure on X. Let $\mathcal{U}$ be a countable base for the topology on X. Suppose $E \in \operatorname{Borel}(X)$ and $\mu(E) < \infty$. Then for any $\varepsilon > 0$, there is a finite collection of basic open sets $U_1, \dots, U_n \in \mathcal{U}$ such that

$$\mu\left(E \triangle \bigcup_{j=1}^{n} U_j\right) < \varepsilon.$$

Proof. By Theorem 8.5, the measure μ is Radon. In particular, μ is outer regular, so we may follow the argument in the proof of Proposition 8.1. □

2. The Second Principle: L^p Functions are Nearly Continuous

There are (at least) two different ways in which L^p functions are "nearly" continuous. One is in terms of the L^p norm.

Proposition 8.7

Let X be an LCH space, and let μ be a Radon measure on X. Then for every $p \in [1, \infty)$, $C_c(X)$ is a dense subspace of $L^p(\mu)$.

Proof. Let $p \in [1, \infty)$.

Step 1. $C_c(X) \subseteq L^p(\mu)$.

Let $f \in C_c(X)$, and let $K = \operatorname{supp}(f) \subseteq X$. Then

$$\int_X |f|^p \, d\mu \leq \max_{x \in K} |f(x)|^p \cdot \mu(K) < \infty$$

by the extreme value theorem and local finiteness of μ, so $f \in L^p(\mu)$.

We will show $C_c(X)$ is dense in $L^p(\mu)$ be approximating successively by more convenient families of functions. Let $\mathcal{S} = \{s : X \to \mathbb{C} : s \text{ is simple and } \mu(\{s \neq 0\}) < \infty\}$.

STEP 2. $\mathcal{S}$ is dense in $L^p(\mu)$.

Note that a simple function $s = \sum_{j=1}^n c_j \mathbb{1}_{E_j}$ written in standard form has L^p norm satisfying $\|s\|_p^p = \sum_{j=1}^n |c_j|^p \mu(E_j)$. Therefore, $s \in L^p(\mu)$ if and only if $\mu(E_j) < \infty$ for every j such that $c_j \neq 0$. In other words, $\mathcal{S}$ is exactly the collection of simple functions that belong to $L^p(\mu)$.

Let $f \in L^p(\mu)$ be arbitrary. Let $(s_n)_{n \in \mathbb{N}}$ be a sequence of simple functions such that $0 \leq |s_1| \leq |s_2| \leq \cdots \leq |f|$ and $s_n \to f$ a.e. (Such a sequence exists by applying Proposition 3.7 to the positive and negative parts of the real and imaginary parts of f.) Then $|s_n - f|^p \to 0$ a.e. and $|s_n - f|^p \leq 2^p |f|^p$, so by the dominated convergence theorem, $s_n \to f$ in $L^p(\mu)$.

STEP 3. $C_c(X)$ is dense in $\mathcal{S}$ (with respect to $\|\cdot\|_p$).

Given a simple function $s = \sum_{j=1}^n c_j \mathbb{1}_{E_j} \in \mathcal{S}$, it suffices by Minkowski's inequality to approximate each of the functions $\mathbb{1}_{E_j}$. Let $E \in \operatorname{Borel}(X)$ with $\mu(E) < \infty$, and let $\varepsilon > 0$. By Proposition 8.4, the measure μ is inner regular on E, and μ is outer regular on all Borel sets, so we may find a compact set K and an open set U such that $K \subseteq E \subseteq U$ and $\mu(U \setminus K) < \varepsilon$. By Urysohn's lemma, let $f \in C_c(X)$ with $K \prec f \prec U$. Then since $f = \mathbb{1}_E = 1$ on K and $f = \mathbb{1}_E = 0$ on $X \setminus U$, we have

$$\int_X |f - \mathbb{1}_E|^p \, d\mu \leq \mu(U \setminus K) < \varepsilon.$$

□

Another sense in which measurable functions are nearly continuous is provided by Lusin's theorem.

THEOREM 8.8: LUSIN'S THEOREM, VERSION I

Let X be an LCH space, and let Y be a second countable topological space. Let μ be a regular Borel measure on X. Suppose $f : X \to Y$ is a Borel measurable function. Then for any $\varepsilon > 0$, there is a closed set $E \subseteq X$ such that $\mu(X \setminus E) < \varepsilon$ and $f|_E$ is continuous.

PROOF. Let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a countable base for the topology on Y. For $n \in \mathbb{N}$, let $B_n = f^{-1}(U_n) \in \operatorname{Borel}(X)$. Since μ is regular, we may find $F_n \subseteq B_n \subseteq G_n$ such that F_n is closed, G_n is open, and $\mu(G_n \setminus F_n) < 2^{-n}\varepsilon$. Let $E = X \setminus \bigcup_{n \in \mathbb{N}} (G_n \setminus F_n)$. Then $\mu(X \setminus E) \leq \sum_{n=1}^{\infty} \mu(G_n \setminus F_n) < \varepsilon$.

CLAIM 1. E is closed.

For each $n \in \mathbb{N}$, the set $G_n \setminus F_n$ is a open, so $\bigcup_{n\in\mathbb{N}}(G_n \setminus F_n)$ is open. Therefore, E is the complement of an open set, so E is closed.

CLAIM 2. $f|_E$ is continuous.

Let $g = f|_E : E \to Y$. Then for each $n \in \mathbb{N}$,

$$g^{-1}(U_n) = B_n \cap E = G_n \cap E$$

is open in E, so g is continuous.

□

It is important to note that Lusin's theorem does NOT say that the set of points of continuity of f has large measure. For example, $\mathbb{1}_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ is a measurable function that is nowhere continuous. However, we can give another interpretation of Lusin's theorem , saying that measurable functions (under some condition) agree with a continuous function outside of a set of small measure.

THEOREM 8.9: LUSIN'S THEOREM, VERSION II

Let X be an LCH space, and let μ be a Radon measure on X. Suppose $f : X \to \mathbb{C}$ is measurable and $\mu(\{f \neq 0\}) < \infty$. Then for any $\varepsilon > 0$, there exists $g \in C_c(X)$ such that $\mu(\{f \neq g\}) < \varepsilon$.

One of the ingredients for the proof is the Tietze extension theorem (which we will not prove).

THEOREM 8.10: TIETZE EXTENSION THEOREM

Let X be an LCH space and let $K \subseteq X$ be compact and $U \subseteq X$ open such that $K \subseteq U$. If $f : K \to \mathbb{C}$ is a continuous function, then there exists $g \in C_c(X)$ such that $g = f$ on K and $\operatorname{supp}(g) \subseteq U$.

Note that if $f = 1$ on K, then we recover Urysohn's lemma from the Tietze extension theorem.

PROOF OF LUSIN'S THEOREM, VERSION II. Let $E = \{f \neq 0\}$. Since E has finite measure, μ is regular on E by Proposition 8.4. Therefore, there exists a compact set $K \subseteq E$ and an open set $U \supseteq E$ such that $\mu(U \setminus K) < \frac{\varepsilon}{2}$. Applying version I of Lusin's theorem on the space K, there exists a closed set $C \subseteq K$ with $\mu(K \setminus C) < \frac{\varepsilon}{2}$ such that $f|_C$ is continuous. Then by Tietze's extension theorem, there exists $g \in C_c(X)$ such that $g = f$ on C and $\operatorname{supp}(g) \subseteq U$. Hence, $\{f \neq g\} \subseteq U \setminus C$, so $\mu(\{f \neq g\}) < \varepsilon$. □

3. The Third Principle: Convergent Sequences of Functions are Nearly Uniformly Convergent

Now we turn to the third principle, expressed by Egorov's theorem.

THEOREM 8.11: EGOROV'S THEOREM

Let $(X, \mathcal{B}, \mu)$ be a finite measure space, and let Y be a separable metric space. Suppose $f_n : X \to Y$ is a sequence of measurable functions that converges a.e. to a measurable function $f : X \to Y$. Then for any $\varepsilon > 0$, there is a set $E \in \mathcal{B}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on $X \setminus E$.

Proof. For $k, n \in \mathbb{N}$, let $E_{k,n} = \bigcup_{m \geq n} \left\{ d(f_m, f) \geq \frac{1}{k} \right\}$.

Claim 1. For each $k, n \in \mathbb{N}$, we have $E_{k,n} \in \mathcal{B}$.

It suffices to check that $d(f_m, f)$ is a measurable function for each $m \in \mathbb{N}$. Let $S \subseteq Y$ be a countable dense subset. Then for $y, z \in Y$, we have $d(y, z) = \inf_{s \in S}(d(y, s) + d(z, s))$. For each $s \in S$, let $D_s : Y \to [0, \infty)$ be the function $D_s(y) = d(y, s)$. The function D_s is continuous, hence Borel measurable. Thus,

$$d(f_m, f) = \inf_{s \in S} (D_s \circ f_m + D_s \circ f)$$

is measurable by Proposition 2.11.

Claim 2. For any $k \in \mathbb{N}$, $\mu\left(\bigcap_{n \in \mathbb{N}} E_{k,n}\right) = 0$.

The set $\bigcap_{n \in \mathbb{N}} E_{k,n}$ is the set of points $x \in X$ such that $d(f_m, f) \geq \frac{1}{k}$ for infinitely many $m \in \mathbb{N}$. Hence, $\bigcap_{n \in \mathbb{N}} E_{k,n} \subseteq \{x \in X : f_n(x) \not\to f(x)\}$ is a null set.

Since μ is a finite measure, we may apply continuity from above and let $n_k \in \mathbb{N}$ such that $\mu(E_{k,n_k}) < 2^{-k}\varepsilon$. Let $E = \bigcup_{k \in \mathbb{N}} E_{k,n_k}$. Then $\mu(E) \leq \sum_{k=1}^{\infty} \mu(E_{k,n_k}) < \varepsilon$.

Claim 3. $f_n \to f$ uniformly on $X \setminus E$.

By definition, if $x \notin E_{k,n_k}$, then $d(f_m(x), f(x)) < \frac{1}{k}$ for all $m \geq n_k$. Therefore,

$$\sup_{x \in X \setminus E_{k,n_k}} d(f_m(x), f(x)) < \frac{1}{k}$$

for all $m \geq n_k$ and all $k \in \mathbb{N}$, so $f_n \to f$ uniformly on $X \setminus E = \bigcap_{k \in \mathbb{N}} (X \setminus E_{k,n_k})$.

□

Corollary 8.12: Bounded Convergence Theorem

Let $(X, \mathcal{B}, \mu)$ be a finite measure space. Suppose $f_n : X \to \mathbb{C}$ is a sequence of measurable functions that converges almost everywhere to a measurable function $f : X \to \mathbb{C}$. If there exists $M < \infty$ such that $|f_n| \leq M$ a.e., then $f_n \to f$ in $L^1(\mu)$. In particular,

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

Proof. This is a special case of the dominated convergence theorem: since $\mu(X) < \infty$, the function $g(x) = M$ is integrable. We will give a proof using Egorov's theorem (which importantly does not rely on the dominated convergence theorem).

Let $\varepsilon > 0$. By Egorov's theorem, let $E \in \mathcal{B}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on $X \setminus E$. Then

$$\int_X |f_n - f| \, d\mu = \int_E |f_n - f| \, d\mu + \int_{X \setminus E} |f_n - f| \, d\mu.$$

The first term $\int_E |f_n - f| \, d\mu$ can be bounded using the triangle inequality by $2M\mu(E) < 2M\varepsilon$. The second term is bounded by

$$\sup_{x \in X \setminus E} |f_n(x) - f(x)| \cdot \underbrace{\mu(X \setminus E)}_{\leq \mu(X) < \infty} \xrightarrow[n \to \infty]{} 0.$$

Hence, $\limsup_{n\to\infty} \|f_n - f\|_1 \leq 2M\varepsilon$. But $\varepsilon > 0$ was arbitrary, so $f_n \to f$ in $L^1(\mu)$. □

Chapter Notes

Some authors prove Lusin's theorem as a consequence of Egorov's theorem. For example, a special case of version I of Lusin's theorem appears as Exercise 44 in [**Fol99**, Section 2.4], with a hint to use Egorov's theorem, and version II is proved using Egorov's theorem in [**Fol99**, Theorem 7.10]. The proofs provided in these lecture notes are shorter and more direct than the deduction from Egorov's theorem.

Bibliography

[BBT08] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson. *Real Analysis*, 2nd edition (2008). Available online at: https://classicalrealanalysis.info/documents/BBT-AlllChapters-Landscape.pdf.

[Fol99] Gerald B. Folland. *Real Analysis*, 2nd edition (John Wiley & Sons, Inc., 1999).

[Get90] R. K. Getoor. *Excessive Measures*, Probab. Appl. (Birkhäuser Boston, 1990).

[Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, Grad. Texts in Math., **156** (Springer, 1995).

[Kim25] Minhyong Kim. History, identity, and ownership in mathematics. *Notices Amer. Math. Soc.* **72** (2025) 846–853.

[Lit44] J. E. Littlewood. *Lectures On The Theory Of Functions* (Oxford University Press, 1944).

[Net25] Reviel Netz. History, identity, and ownership in mathematics: a response to Minhyong Kim. *Notices Amer. Math. Soc.* **72** (2025) 854–857.

[Oxt80] John C. Oxtoby. *Measure and Category*, 2nd edition, Grad. Texts in Math., **2** (Springer, 1980).

[Rud87] Walter Rudin. *Real and Complex Analysis*, 3rd edition (McGraw-Hill Book Co., 1987).

[Sol70] Robert M. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. Math. (2)* **92** (1970) 1–56.

[SS05] Elias M. Stein and Rami Shakarchi. *Real Analysis* (Princeton University Press, 2005).

[Tao11] Terence Tao. *An Introduction to Measure Theory*, Grad. Stud. Math., **126** (American Mathematical Society, 2011). Preliminary version available online at https://terrytao.wordpress.com/wp-content/uploads/2012/12/gsm-126-tao5-measure-book.pdf

[Tse22] Anush Tserunyan. *Introduction to Descriptive Set Theory.* Online lecture notes (2022). Available at: https://www.math.mcgill.ca/atserunyan/Teaching_notes/dst_lectures.pdf

Appendix

APPENDIX A

Metric Spaces and Topological Spaces

This appendix includes a very brief overview of concepts from point set topology. It is not expected for this course that you have a complete mastery of the subject, but we will use various concepts covered herein throughout the course. You may which to return to this appendix when relevant as the semester goes along.

You should be aware that these notes are light on exposition and heavy on definitions. I have tried to include interesting and illustrative examples and exercises to help with understanding the definitions. However, if you are struggling with any of the concepts, you may wish to consult a longer reference on metric spaces or topology.

1. Metric spaces

Metric spaces are an abstract mathematical object capturing the essential features of the notion of "distance."

Definition A.1

A *metric* (or *distance*) on a set X is a function $d : X \times X \to [0, \infty)$ satisfying the following properties for all $x, y, z \in X$:

- POSITIVE DEFINITE: $d(x, y) = 0 \iff x = y$;
- SYMMETRIC: $d(x, y) = d(y, x)$; and
- TRIANGLE INEQUALITY: $d(x, z) \leq d(x, y) + d(y, z)$.

If d is a metric on X, we say that the pair (X, d) is a *metric space*.

1.1. Examples of metric spaces.

Example A.2: Real numbers as a metric space

The real numbers are a metric space with the metric $d(x, y) = |x - y|$.

Example A.3: Metrics on Euclidean spaces

Let $k \in \mathbb{N}$. The Euclidean space $\mathbb{R}^k$ can be equipped with several different metrics. These include:

- Euclidean distance: $d_2(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{j=1}^{k} (x_j - y_j)^2}$
- Manhattan or taxicab distance: $d_1(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{k} |x_j - y_j|$
- Chebyshev distance: $d_\infty(\boldsymbol{x}, \boldsymbol{y}) = \max_{1 \leq j \leq k} |x_j - y_j|$

Example A.4: Discrete metric

Every set can be made into a metric space with the discrete metric

$$d(x, y) = \begin{cases} 0, & \text{if } x = y; \\ 1, & \text{if } x \neq y. \end{cases}$$

Example A.5: Graph distance

Every connected graph can be viewed as a metric space by defining, for pairs of vertices u, v,

$$d(u, v) = \# \text{ of edges in the shortest path from } u \text{ to } v.$$

Exercise A.1

Check that the spaces in the above examples are metric spaces.

1.2. Limits and continuity. Since metric spaces come with a notion of points being "close" to one another, they are a natural setting for dealing with familiar notions from analysis such as limits, continuity, etc.

Definition A.6

Let (X, d) be a metric space. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X and $x \in X$. We say that $(x_n)_{n\in\mathbb{N}}$ *converges* to x, written $\lim_{n\to\infty} x_n = x$, if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$n \geq N \implies d(x_n, x) < \varepsilon.$$

Definition A.7

Let (X, d) and (Y, ρ) be metric spaces, and let $f : X \to Y$.

- f is *continuous at a point* $x \in X$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that if $x' \in X$ and $d(x', x) < \delta$, then $\rho(f(x'), f(x)) < \varepsilon$.
- f is *continuous* if it is continuous at every point $x \in X$.
- f is *uniformly continuous* if for every $\varepsilon > 0$, there exists $\delta > 0$ such that if $x_1, x_2 \in X$ with $d(x_1, x_2) < \delta$, then $\rho(f(x_1), f(x_2)) < \varepsilon$.

Remark. Note that this agrees with the usual definition of limits and continuity in the real numbers using the metric from Example A.2.

Exercise A.2

Show that every uniformly continuous function is continuous. Give an example of a function that is continuous but not uniformly continuous.

Exercise A.3

Let (X, d) and (Y, ρ) be metric spaces, let $f : X \to Y$, and let $x \in X$. Show that f is continuous at x if and only if for every sequence $(x_n)_{n\in\mathbb{N}}$, if $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} f(x_n) = f(x)$.

1.3. Open and closed sets.

DEFINITION A.8

Let (X, d) be a metric space.

- The *open ball* of radius $r > 0$ centered at a point $x \in X$ is the set $B(x, r) = \{y \in X : d(x, y) < r\}$.
- The *closed ball* of radius $r > 0$ centered at a point $x \in X$ is the set $B[x, r] = \{y \in X : d(x, y) \leq r\}$.
- A set $U \subseteq X$ is *open* if for every $x \in U$, there exists $r > 0$ such that $B(x, r) \subseteq U$.
- A set $C \subseteq X$ is *closed* if every point approximable by elements of C belongs to C. That is, if $x \in X$ and $B(x, r) \cap C \neq \emptyset$ for every $r > 0$, then $x \in C$.
- A set $D \subseteq X$ is *dense* if every point can be approximated arbitrarily well by elements of D. That is, for every $x \in X$ and every $\varepsilon > 0$, there exists $y \in D$ such that $d(x, y) < \varepsilon$.
- The *closure* of a set $E \subseteq X$, denoted by $\overline{E}$, is the smallest closed set containing E.
- The *interior* of a set $E \subseteq X$, denoted by E°, is the largest open set contained in E.

EXERCISE A.4

Prove the following properties of open and closed sets:

(a) For every $x \in X$ and $r > 0$, the open ball $B(x, r)$ is open.
(b) For every $x \in X$ and $r > 0$, the closed ball $B[x, r]$ is closed.
(c) There exists a metric space (X, d), a point $x \in X$, and a radius $r > 0$ such that $\overline{B(x, r)} \neq B[x, r]$.
(d) A set is open if and only if its complement is closed.
(e) An arbitrary intersection of closed sets is closed.
(f) An arbitrary union of open sets is open.
(g) The closure and interior of a set are well-defined. (Hint: use (b) and (c).)
(h) A set D is dense in X if and only if $\overline{D} = X$.

1.4. Properties of metric spaces. We now discuss different properties that metric spaces may possess. To enable this discussion, we need another definition.

DEFINITION A.9

Let (X, d) be a metric space. A sequence $(x_n)_{n \in \mathbb{N}}$ is a *Cauchy sequence* if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$n, m \geq N \implies d(x_n, x_m) < \varepsilon.$$

EXERCISE A.5

Show that every convergent sequence in a metric space is Cauchy.

In a Cauchy sequence, the terms cluster together and appear to *want* to converge. However, convergence is not guaranteed.

Example A.10

Consider $\mathbb{Q}$ as a metric space with metric $d(x,y) = |x-y|$. Define a sequence $(x_n)_{n\in\mathbb{N}}$ recursively as follows:

$$x_0 = 2$$
$$x_n = \frac{x_{n-1}}{2} + \frac{1}{x_{n-1}}.$$

Since we start with a rational value for x_0, it is clear from the recurrence relation that x_n is rational for every $n \in \mathbb{N}$. By the arithmetic mean-geometric mean inequality,

$$x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) > \sqrt{x_{n-1} \cdot \frac{2}{x_{n-1}}} = \sqrt{2}.$$

It follows that $(x_n)_{n\in\mathbb{N}}$ is a decreasing sequence, since

$$x_n = x_{n-1} + \frac{1}{x_{n-1}} \underbrace{\left(-\frac{x_{n-1}^2}{2} + 1\right)}_{<0} < x_{n-1}.$$

Thus, $(x_n)_{n\in\mathbb{N}}$ is a strictly decreasing sequence in the interval $\mathbb{Q} \cap (\sqrt{2}, 2]$. One can check that bounded monotone sequences are always Cauchy in $\mathbb{Q}$.

However, the sequence $(x_n)_{n\in\mathbb{N}}$ does not converge in $\mathbb{Q}$. To see this, suppose for contradiction that $x = \lim_{n\to\infty} x_n$. Then

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty}\left(\frac{x_{n-1}}{2} + \frac{1}{x_{n-1}}\right) = \frac{x}{2} + \frac{1}{x}.$$

Solving for x gives $x = \sqrt{2}$, but $\sqrt{2}$ is an irrational number, so $(x_n)_{n\in\mathbb{N}}$ does not converge in $\mathbb{Q}$.

The failure of convergence in Example A.10 comes from the fact that the rational numbers have "holes" that are filled in with irrationals in the reals. Spaces without such "holes" are called complete.

Definition A.11

A metric space is *complete* if every Cauchy sequence converges.

Theorem A.12

Given a metric space (X, d), there exists a space $(\widetilde{X}, \widetilde{d})$ such that

- $(\widetilde{X}, \widetilde{d})$ is complete;
- there is an isometry[a] $\iota : X \to \widetilde{X}$; and
- $\iota(X)$ is dense in $\widetilde{X}$.

[a] An isometry is a distance-preserving map. Saying that ι is an isometry means $\widetilde{d}(\iota(x_1), \iota(x_2)) = d(x_1, x_2)$ for all $x_1, x_2 \in X$.

Remark. The space $(\widetilde{X}, \widetilde{d})$ is called the *completion* of (X, d).

Example A.13

The real numbers are the completion of the rational numbers.

Proof of Theorem A.12. (Sketch) Let Y be the space of Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ in X. Define an equivalence relation on Y by

$$(x_n)_{n\in\mathbb{N}} \sim (x'_n)_{n\in\mathbb{N}} \iff \lim_{n\to\infty} d(x_n, x'_n) = 0.$$

We then let $\widetilde{X} = Y/\sim$ be the space of equivalence class of sequences with the metric

$$\widetilde{d}\left([(x_n)_{n\in\mathbb{N}}], [(x'_n)_{n\in\mathbb{N}}]\right) = \lim_{n\to\infty} d(x_n, x'_n).$$

The map $\iota : X \to \widetilde{X}$ is defined by $\iota(x) = (x, x, x, \dots)$ for $x \in X$. Given a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$, the corresponding sequence $(\iota(x_n))_{n\in\mathbb{N}}$ in $\widetilde{X}$ now converges to the point $[(x_n)_{n\in\mathbb{N}}] \in \widetilde{X}$. □

One of the main theorems concerning convergence of sequences in the real numbers is the Bolzano–Weierstrass theorem, which says that every bounded sequence of real numbers has a convergent subsequence. This motivates the following definition.

Definition A.14

Let (X, d) be a metric space. A set $K \subseteq X$ is *sequentially compact* if every sequence in K has a convergent subsequence in K. That is, for every $(x_n)_{n\in\mathbb{N}}$ in K, there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} x_{n_k} \in K$.

It turns out that for metric spaces, a set being sequentially compact is equivalent to another notion of compactness given in the next definition.

Definition A.15

Let (X, d) be a metric space. A set $K \subseteq X$ is *compact* if every open cover of K has a finite subcover. That is, if $(U_i)_{i\in I}$ is a family of open sets and $K \subseteq \bigcup_{i\in I} U_i$, then there exists a finite set $\{i_1, \dots, i_N\} \subseteq I$ such that $K \subseteq \bigcup_{j=1}^{N} U_{i_j}$.

Theorem A.16

Let (X, d) be a metric space. The following are equivalent:

(i) X is sequentially compact;
(ii) X is compact;
(iii) X is complete and totally bounded (i.e., for every $\varepsilon > 0$, there is a finite set $\{x_1, \dots, x_N\} \subseteq X$ such that $X = \bigcup_{j=1}^{N} B(x_j, \varepsilon)$).

Exercise A.6

Prove Theorem A.16.

Compact subsets of the real line are characterized by the Heine–Borel theorem.

Theorem A.17: Heine–Borel

A subset $K \subseteq \mathbb{R}$ is compact if and only if it is closed and bounded.

2. Topological spaces

Topological spaces involve another level of abstraction, dispensing with the notion of distance but retaining a notion of open sets.

Definition A.18

A *topology* on a set X is a collection of subsets of X, $\tau \subseteq \mathscr{P}(X)$, such that

- $\emptyset, X \in \tau$;
- if $\{U_i\}_{i \in I} \subseteq \tau$, then $\bigcup_{i \in I} U_i \in \tau$;
- if $U, V \in \tau$, then $U \cap V \in \tau$.

Elements of τ are called *open sets*, and the pair (X, τ) is called a *topological space*.

Example A.19: Discrete and anti-discrete topology

Let X be a set. The *discrete topology* on X is $\tau_{\text{disc}} = \mathscr{P}(X)$, and the *anti-discrete topology* on X is $\tau_{\text{anti-disc}} = \{\emptyset, X\}$.

Example A.20: Metric spaces

Let (X, d) be a metric space. The open sets (as defined in Definition A.8) form a topology on X.

Exercise A.7

Check that the above examples satisfy the definition of a topology.

Remark. When discussing topological spaces, it is typical to omit explicit reference to τ and simply refer to its elements as open sets. In line with this standard mathematical practice, for the remainder of this document we will write, "X is a topological space," with the understanding that this entails an implicit collection of open sets.

2.1. Subsets of topological spaces. Motivated by the corresponding definitions for metric spaces, we introduce terminology for special classes of subsets of a topological space.

Definition A.21

Let X be a topological space.

- A set is *closed* if its complement is open.
- The *closure* of a set E is the smallest closed set containing E.
- The *interior* of E is the largest open set contained in E.
- A set D is *dense* in X if every nonempty open set in X intersects D.
- A set $K \subseteq X$ is *compact* if every open cover of K has a finite subcover.

2.2. Bases. It is often impractical to specify a topology by listing every possible open set. One instead frequently defines a topology in terms of a basis (or base).

Definition A.22

A *basis* (or *base*) for a topological space X is a family $\mathcal{B}$ of open sets such that

- for every point $x \in X$, there exists $B \in \mathcal{B}$ such that $x \in B$,
- for every $B_1, B_2 \in \mathcal{B}$ and every point $x \in B_1 \cap B_2$, there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

One should think of the basis axioms as corresponding to the assumption that X is open and that the intersection of two open sets is again an open set. The condition that arbitrary unions of open sets are open does not appear anywhere in the definition of a basis; this is because the topology generated by the basis consists precisely of arbitrary unions of basis elements.

Example A.23

In a metric space, the family of open balls forms a basis.

Example A.24: Furstenberg topology on the integers

For $a, b \in \mathbb{Z}$ with $a \neq 0$, let $S(a,b) = \{an + b : n \in \mathbb{Z}\} = a\mathbb{Z} + b$. Then $(S(a,b))_{a,b\in\mathbb{Z},a\neq 0}$ is the basis for a topology on $\mathbb{Z}$.

Exercise A.8

Show that each basis element $S(a,b)$ is also closed in the Furstenberg topology. Use the fundamental theorem of arithmetic to find an expression for the set $\mathbb{Z}\backslash\{-1,+1\}$ and conclude that there are infinitely many prime numbers.

2.3. Limits and continuity. We now extend notions of convergence and continuous functions to the setting of topological spaces.

Definition A.25

A sequence $(x_n)_{n\in\mathbb{N}}$ in a topological space X *converges* to $x \in X$ if for every open neighborhood $U \ni x$, there exists $N \in \mathbb{N}$ such that

$$n \geq N \implies x_n \in U.$$

Definition A.26

Let X and Y be topological spaces. A function $f : X \to Y$ is *continuous* if for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in X.

Exercise A.9

Check that these definitions agree with the corresponding definitions for metric spaces.

Exercise A.10

Suggest a definition for what it means for f to be *continuous at a point* $x \in X$. The definition should generalize the definition for metric spaces.

2.4. Properties of topological spaces. There are many properties that topological spaces may or may not have. We give a list of various properties that will play a role in this course.

Definition A.27

Let X be a topological space. We say that X is

- *second countable* if there is a countable basis for the topology
- *separable* if there is a countable dense subset
- *locally compact* if for every $x \in X$, there is an open set $U \ni X$ such that $\overline{U}$ is compact
- *Hausdorff* if for every pair of distinct points $x, y \in X$, $x \neq y$, there are open sets $U \ni x$ and $V \ni y$ such that $U \cap V = \emptyset$
- *metrizable* if there is a metric inducing the topology
- *completely metrizable* if it is metrizable and there is a compatible metric for which X is complete
- *Polish* if it is separable and completely metrizable

Exercise A.11

Give an example of a non-metrizable topological space.

Exercise A.12

Show that every metric space is Hausdorff.

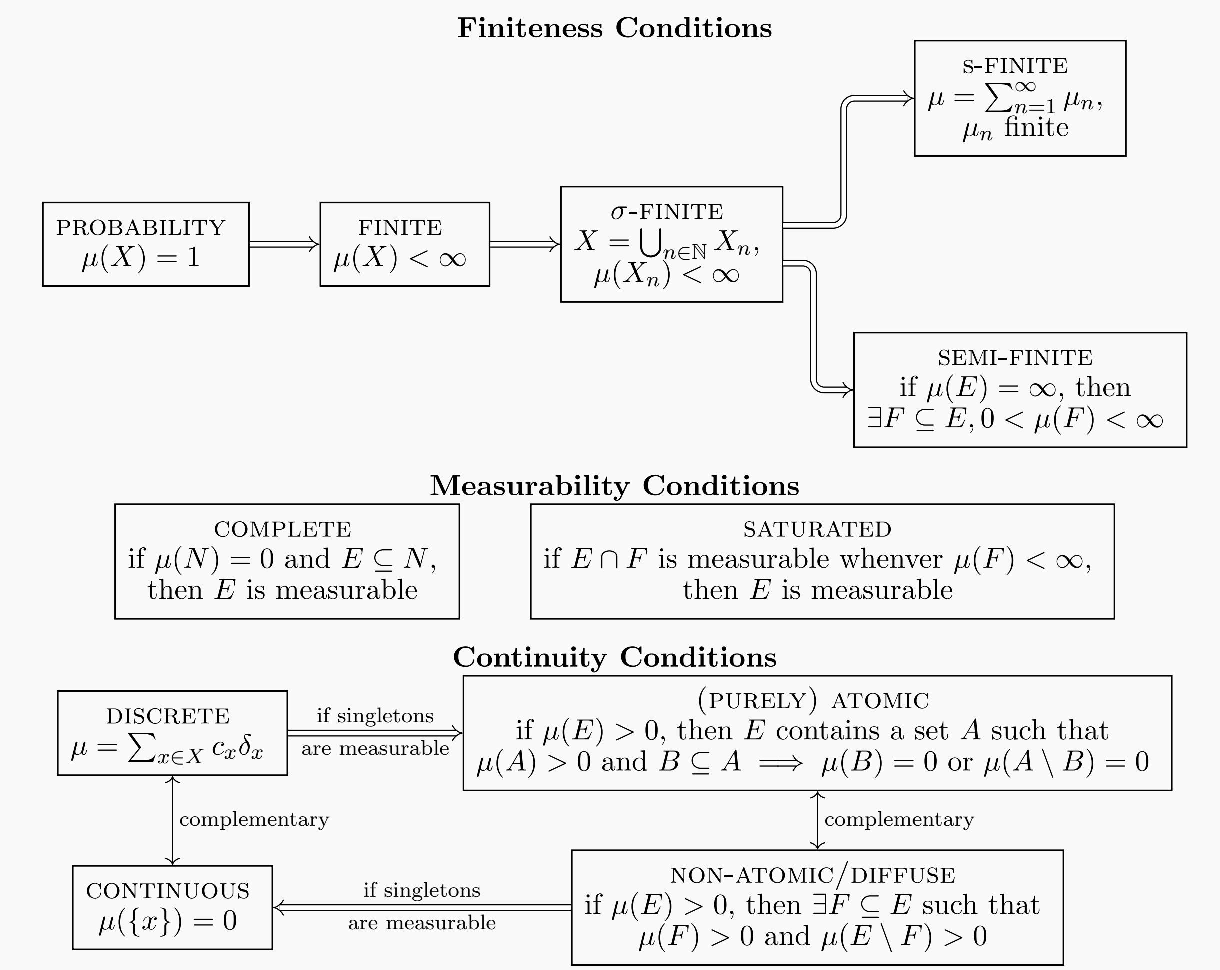
Exercise A.13

(a) Show that every second countable space is separable.
(b) Show that if X is a separable metric space, then it is second countable.
(c) Give an example of a separable topological space that is not second countable.

APPENDIX B

Taxonomy of Measures

In this appendix, we collect various properties that measures may or may not have and that appear throughout the notes. The properties are arranged into a taxonomy in Figure B.1. Definitions of each property are given afterwards.



Regularity Conditions (for Borel Measures)

LOCALLY FINITE	$\forall x \in X, \exists U \ni x$ neighborhood with $\mu(U) < \infty$
OUTER REGULAR	$\mu(E) = \inf\{\mu(U) : U \supseteq E \text{ open}\}$
INNER REGULAR	$\mu(E) = \sup\{\mu(K) : K \subseteq E \text{ compact}\}$

FIGURE B.1. Taxonomy of Measures

1. Properties of Measures

Definition B.1

Let $(X, \mathcal{B}, \mu)$ be a measure space. A set E is

- *locally measurable* if $E \cap K \in \mathcal{B}$ for every set $K \in \mathcal{B}$ with $\mu(K) < \infty$;
- an *atom* if $\mu(E) > 0$ and every measurable subset $F \subseteq E$, $F \in \mathcal{B}$ satisfies either $\mu(F) = 0$ or $\mu(E \setminus F) = 0$.

Definition B.2

Let $(X, \mathcal{B})$ be a measurable space. A measure μ on $(X, \mathcal{B})$ is

- a *probability measure* if $\mu(X) = 1$;
- *finite* if $\mu(X) < \infty$;
- *σ-finite* if there is a countable sequence of measurable sets $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{B}$ such that $X = \bigcup_{n\in\mathbb{N}} X_n$ and $\mu(X_n) < \infty$ for each $n \in \mathbb{N}$;
- *s-finite* if μ is a countable sum $\mu = \sum_{n\in\mathbb{N}} \mu_n$ of finite measures $\mu_n : \mathcal{B} \to [0, \infty)$;
- *semi-finite* if every set of infinite measure contains a subset of positive finite measure, i.e. if $E \in \mathcal{B}$ and $\mu(E) = \infty$, then there exists $F \in \mathcal{B}$ with $F \subseteq E$ and $\mu(F) < \infty$;
- *complete* if every subset of every null set is measurable, i.e. if $E \subseteq X$ and there exists $N \in \mathcal{B}$ with $E \subseteq N$ and $\mu(N) = 0$, then $E \in \mathcal{B}$;
- *saturated* if every locally measurable set is measurable;
- *discrete* if μ is a combination of Dirac measures, $\mu = \sum_{x\in X} c_x \delta_x$ for some coefficients $c_x \in [0, \infty]$;
- *continuous* if μ has no point masses, i.e. $\mu(\{x\}) = 0$ for every $x \in X$;
- *(purely) atomic* if every set of positive measure contains an atom;
- *non-atomic* or *diffuse* if there are no atoms.

If X is a topological space and $\mathcal{B} = \mathrm{Borel}(X)$, then μ is

- *locally finite* if every point has a neighborhood of finite measure;
- *outer regular* if $\mu(E) = \inf\{\mu(U) : U \supseteq E \text{ open}\}$ for every $E \in \mathcal{B}$;
- *inner regular* if $\mu(E) = \sup\{\mu(K) : K \subseteq E \text{ compact}\}$ for every $E \in \mathcal{B}$.

The relationships between the various properties in Definition B.2 are displayed in Figure B.1.

2. Finiteness Properties

As shown in Figure B.1, every probability measure is finite, every finite measure is σ-finite, and every σ-finite measure is both s-finite and semi-finite. The next example shows that s-finiteness and semi-finiteness are rather different notions from one another, neither one implying the other.

Example B.3

An s-finite measure that is not semi-finite: Let X be a non-empty set, and let $x \in X$. Define $\mu(E) = \infty$ if $x \in E$ and $\mu(E) = 0$ if $x \notin E$. Then $\mu = \sum_{n=1}^{\infty} \delta_x$, so μ is s-finite. However, the set $\{x\}$ has infinite measure and no subsets of non-zero finite measure, so μ is not semi-finite.

A semi-finite measure that is not s-finite: Let X be an uncountable set, and let $\mu : \mathscr{P}(X) \to [0, \infty]$ be the counting measure on X. If $E \subseteq X$ and $\mu(E) = \infty$, then taking any point $x \in E$, we have $\mu(\{x\}) = 1 < \infty$, so μ is a semi-finite measure. However, since

X is uncountable, μ cannot be expressed as a countable sum of finite measures, so μ is not s-finite.

Most texts on measure theory focus on σ-finite measures and omit mention of the more general concept of s-finite measures. However, as we will see, s-finite measures are the natural class of measures for many important results in measure theory. One reason to appreciate the generality of s-finite measures is provided by the next theorem. We will encounter other advantages of working with s-finite measures later on in the course.

Theorem B.4

(1) Let $(X, \mathcal{B}, \mu)$ be an s-finite measure space, and let $(Y, \mathcal{C})$ be a measurable space. Suppose $\pi : X \to Y$ is a measurable map. Then the measure $\pi_*\mu : \mathcal{C} \to [0, \infty]$ defined by $\pi_*\mu(C) = \mu(\pi^{-1}(C))$ is s-finite.
(2) There exists a σ-finite measure space $(X, \mathcal{B}, \mu)$, a measurable space $(Y, \mathcal{C})$, and a measurable map $\pi : X \to Y$ such that $\pi_*\mu$ is not σ-finite.

Proof. **(1)** First note that the projection of a finite measure is finite. Indeed, $\pi_*\mu(Y) = \mu(\pi^{-1}(Y)) = \mu(X)$. Noting that $\pi_* \left(\sum_{n=1}^{\infty} \mu_n\right) = \sum_{n=1}^{\infty} \pi_*\mu_n$ then completes the proof.

(2) Let $X = \mathbb{Z}^2$, $\mathcal{B} = \mathscr{P}(\mathbb{Z}^2)$, and let $\mu : \mathcal{B} \to [0, \infty]$ be the counting measure. Let $Y = \mathbb{Z}$ and $\mathcal{C} = \mathscr{P}(\mathbb{Z})$, and let $\pi : \mathbb{Z}^2 \to \mathbb{Z}$ be the projection onto the first coordinate, i.e. $\pi(n, m) = n$ for $(n, m) \in \mathbb{Z}^2$. Then $\pi_*\mu(E)$ counts the number of points in $\mathbb{Z}^2$ whose first coordinate belongs to E. Hence, $\pi_*\mu(E) = \infty$ whenever $E \neq \emptyset$. Therefore, $\pi_*\mu$ is not σ-finite (nor even semi-finite). □

3. Decompositions of Measures

When we say that two notions are "complementary," we mean that they are mutually exclusive and every (σ-finite) measure can be decomposed into pieces satisfying one or the other property. Namely, for the complementary notions shown in Figure B.1, we have the following decomposition result:

Proposition B.5

(1) Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space. Then there is a unique decomposition $\mu = \mu_a + \mu_{na}$ as a sum of a purely atomic measure μ_a and a non-atomic measure μ_{na}.
(2) Let $(X, \mathcal{B}, \mu)$ be an s-finite measure space, and suppose $\{x\} \in \mathcal{B}$ for every $x \in X$. Then there is a unique decomposition $\mu = \mu_d + \mu_c$ as a sum of a discrete measure μ_d and a continuous measure μ_c.

In general, *atomic* and *discrete* are different notions.

Example B.6

Let X be an uncountable set, and let $\mathcal{B} = \{E \subseteq X : E \text{ is countable or } X \setminus E \text{ is countable}\}$. Define a probability measure $\mu : \mathcal{B} \to [0, 1]$ by $\mu(E) = 0$ if E is countable and $\mu(E) = 1$ is $X \setminus E$ is uncountable. Then E is atomic (each co-countable set is an atom) but also continuous.

However, in many frequently-encountered situations, atomic and discrete measures coincide.

Theorem B.7

Let X be a separable metric space. Suppose μ is a locally finite Borel measure on X. If $A \in \mathrm{Borel}(X)$ is an atom of μ, then there is a point $x \in A$ such that $\mu(\{x\}) = \mu(A) > 0$. Hence, every atomic locally finite Borel measure on X is discrete.

Proving the decomposition of a σ-finite measure into atomic and non-atomic components is a bit lengthy, so we will prove only part (2) of Proposition B.5. Because of Theorem B.7, the decomposition into discrete and continuous parts is sufficient for most purposes.

Proof of Proposition B.5(2).

Step 1. Existence.

Since μ is s-finite, we may write $\mu = \sum_{n=1}^{\infty} \mu_n$ for some finite measures $\mu_n : \mathcal{B} \to [0, \infty)$. For each $k \in \mathbb{N}$, let $X_{n,k} = \left\{x \in X : \mu_n(\{x\}) \geq \frac{1}{k}\right\}$. Note that $X_{n,k}$ has at most $k\mu_n(X)$ elements for each $n, k \in \mathbb{N}$. Therefore, $X_0 = \{x \in X : \mu(\{x\}) > 0\} = \bigcup_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} X_{n,k}$ is a countable set.

For $x \in X_0$, let $c_x = \mu(\{x\})$. Define $\mu_d = \sum_{x\in X_0} c_x \delta_x$, and let $\mu_c : \mathcal{B} \to [0, \infty)$ be the measure $\mu_c(E) = \mu(E \setminus X_0)$ for $E \in \mathcal{B}$. Then μ_d is manifestly a discrete measure. Moreover, for any $x \in X$,

$$\mu_c(\{x\}) = \mu(\{x\} \setminus X_0) = \begin{cases} \mu(\{x\}), & \text{if } x \notin X_0; \\ 0, & \text{if } x \in X_0. \end{cases}$$

Since X_0 is the set of all point masses for μ, it follows that $\mu_c(\{x\}) = 0$ for every $x \in X$; that is, μ_c is continuous. Finally, for any $E \in \mathcal{B}$,

$$\mu(E) = \mu(E \cap X_0) + \mu_c(E)$$

and

$$\mu(E \cap X_0) = \sum_{x\in E\cap X_0} \mu(\{x\}) = \sum_{x\in X_0} c_x \delta_x(E) = \mu_d(E).$$

Step 2. Uniqueness.

Let $\mu = \mu_d + \mu_c$ be the decomposition obtained in Step 1. Suppose $\mu = \mu'_d + \mu'_c$ is another decomposition into a discrete measure μ'_d and a continuous measure μ'_c. We want to show $\mu'_d = \mu_d$ and $\mu'_c = \mu_c$.

Let $x \in X_0$. Since μ'_c is continuous, we have $\mu'_c(\{x\}) = 0$, so $\mu'_d(\{x\}) = \mu(\{x\}) = c_x$. On the other hand, if $x \in X$ is any point and $\mu'_d(\{x\}) > 0$, then $\mu(\{x\}) \geq \mu'_d(\{x\}) > 0$, so $x \in X_0$. Therefore, the point masses of μ'_d are exactly the elements of X_0, and $\mu'_d(\{x\}) = c_x$ for $x \in X_0$. Since μ'_d is discrete, it can thus be represented as $\mu'_d = \sum_{x\in X_0} c_x\delta_x$. That is, $\mu'_d = \mu_d$, and it follows that we also have $\mu'_c = \mu_c$.

□

The condition of semi-finiteness also leads to a decomposition result.

Proposition B.8

Let $(X, \mathcal{B}, \mu)$ be a measure space. There exists a decomposition $\mu = \mu_{\mathrm{sf}} + \mu_{\mathrm{inf}}$ such that μ_{sf} is semi-finite and μ_{inf} takes only the values 0 and ∞.

Unlike the decompositions in Proposition B.5, the decomposition in Proposition B.8 is not unique in general. One way of obtaining the decomposition is to define

$$\mu_{\mathrm{sf}}(E) = \sup\left\{\mu(F) : F \in \mathcal{B}, F \subseteq E, \text{and } \mu(F) < \infty\right\},$$

and

$$\mu_{\mathrm{inf}}(E) = \begin{cases} 0, & \text{if } E \text{ is semi-finite;} \\ \infty, & \text{if } E \text{ is not semi-finite.} \end{cases}$$

Here, we say that a measurable set E is semi-finite if the measure $\mu_E : \mathcal{B} \to [0, \infty]$ defined by $\mu_E(A) = \mu(A \cap E)$ is a semi-finite measure. In other words, $E \in \mathcal{B}$ is semi-finite if every subset of E of infinite measure has a further subset of positive finite measure.