Analysis, Measure, and Probability: A visual introduction

Marcus Pivato

April 13, 2004

Contents

List of Figures

Preface

These lecture notes are written with four principles in mind:

1. **You learn by doing, not by watching.** Because of this, many of the routine or technical aspects of proofs and examples have been left as **exercises**, which are sprinkled through the text. Most of these really aren't that hard; indeed, it often actually *easier* to figure them out yourself than to pick through the details of someone else's explanation. It is also more fun. And it is definitely a better way to learn the material. I suggest you do as many of these exercises as you can. Don't cheat yourself.

2. **Pictures often communicate better than words.** Analysis is a geometrically and physically motivated subject. Algebraic formulae are just a language used to communicate visual/physical ideas in lieu of pictures, and they generally make a poor substitute. I've included as many pictures as possible, and I suggest that you look at the pictures *before* trying to figure out the formulae; often, once you understand the picture, the formula becomes transparent.

3. **Learning proceeds from the concrete to the abstract.** Thus, I begin each discussion with specific examples and only later proceed to a more general/abstract idea. This introduces a lot of "redundancy" into the text, in the sense that later formulations subsume the earlier ones. So the exposition is not as "efficient" as it could be. This is a good thing. Efficiency makes for good reference books, but lousy texts.

4. **Mathematics is a Lattice, not a Ladder.** Mathematical knowledge forms a heirarchy: mastery of simpler concepts is necessary to learn advanced material. But this heirarchy is not a linearly ordered 'ladder'; it is a *lattice*, like a ramified network of ivy ascending a brick wall.

   To learn an advanced topic appearing late in this book, it is usually only necessary to understand *some* previous material. Thus, each section heading is followed by a list of *prerequisites*: the previous sections which have the material logically necessary for that section. Some sections also list *recommended* material: this is not logically necessary, but may be intuitively helpful.

   If you are interested in a topic on page 300, you must first read the prerequisites to that topic (and before that, the prerequisites to the prerequisites, etc.). But you need not read the entire previous 300 pages. By utilizing this lattice of prerequisites, you can define many different itineraries of self-study.

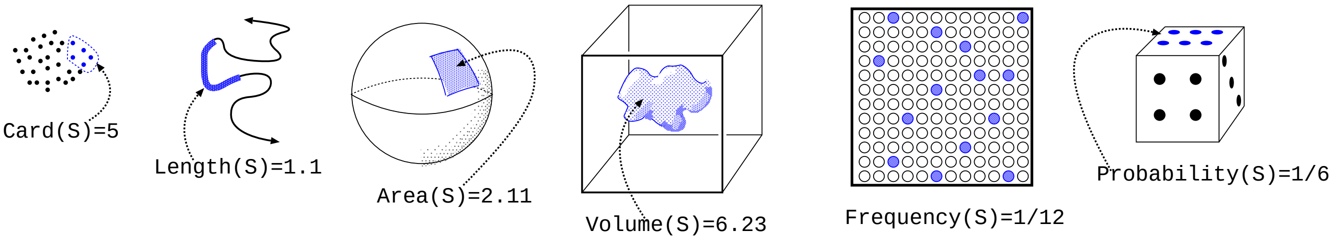


Figure 1.1: The notions of *cardinality*, *length*, *area*, *volume*, *frequency*, and *probability* are all examples of measures.

1 Basic Ideas

1.1 Preliminaries

1.1(a) The concept of size

Let $\mathbf{X}$ be a set, with subsets $\mathbf{U}, \mathbf{V} \subset \mathbf{X}$ Suppose we want to comparate the 'sizes' of $\mathbf{U}$ and $\mathbf{V}$. Perhaps we want a rigorous mathematical way to say that $\mathbf{U}$ is 'bigger' than $\mathbf{V}$. We might also want this concept of size to have an *additive* property; if $\mathbf{U}$ and $\mathbf{V}$ are disjoint, we might want to speak of the 'size' of $\mathbf{U} \sqcup \mathbf{V}$ as being the *sum* of the sizes of $\mathbf{U}$ and $\mathbf{V}$ separately.

Many notions of 'quantity' behave in this manner.

Cardinality, for discrete sets. For example, the *population* of a region or the number of marbles in a bag.

Length, for one-dimensional objects like ropes or roads.

Area, for two-dimensional surfaces like meadows or carpets.

Volume, for three-dimensional regions: the capacity of a bottle, or a quantity of wine.

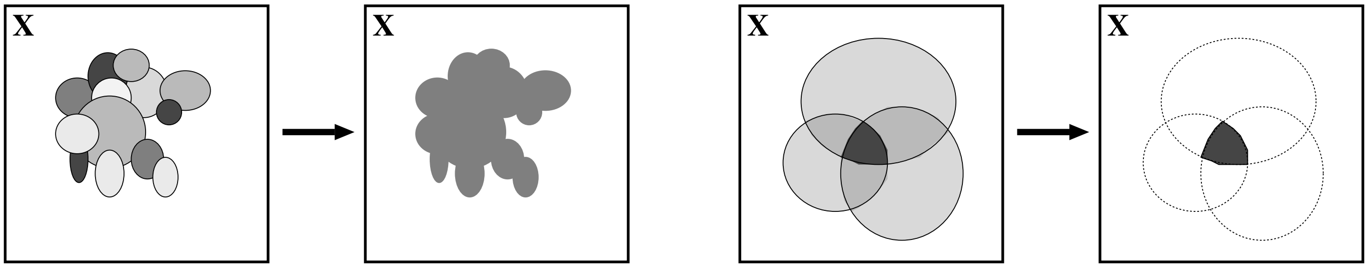
Mass, for quantities of matter.

Charge, for electrostatic objects.

Average Frequency —ie. how often, on average, a certain event occurs. Such frequencies are additive. For example, the number of days during the year when it is overcast can be written as a sum of:

- The number of days it is overcast and raining.
- The number of days it is overcast but *not* raining.

Probability: We often speak of the *odds* of certain events occuring; these are also additive. For example, in a game of dice, the chance of rolling less than *four* is a sum of:



(A) Closure under countable union **(B)** Closure under countable intersection

Figure 1.2: **(A)** A sigma algebra is closed under countable unions; **(B)** A sigma algebra is closed under countable intersections.

- The chance of rolling a *one*.
- The chance of rolling a *two*.
- The chance of rolling a *three*.

A **measure** is mathematical device which reflects this notion of 'quantity': each subset $\mathbf{U} \subset \mathbf{X}$, is assigned a positive real number $\mu[\mathbf{U}]$ which reflects the 'size' of $\mathbf{U}$. Because of this, one would expect that μ would be a function

$$\mu : \mathcal{P}(\mathbf{X}) \longrightarrow \mathbb{R}$$

where $\mathcal{P}(\mathbf{X}) := \{\mathbf{S} \subset \mathbf{X}\}$ is the **power set** of $\mathbf{X}$. Unfortunately, it is usually impossible to define a satisfactory notion of quantity for *all* subsets of $\mathbf{X}$; instead, we must isolate a smaller domain, where the measure will be well-defined. Subsets which are elements of this smaller domain are called **measurable**, and they become our domain of discourse. Subsets not in the domain are called **non-measurable**, and we try to ignore their existence whenever possible.

To understand this, suppose that, like the ancient Greeks, you wanted to represent all real numbers as fractions of integers. Of course, we now know that only *rational* numbers can be accurately represented in this way; irrational numbers are 'non-measurable' in terms of integer fractions. In a sense, the set of rational numbers is 'too small' to completely describe all real numbers. In the same way, the set of real numbers is too small to completely rank all subsets of $\mathbf{X}$; nonmeasurable sets are, for us, somewhat analogous to what irrational numbers were for the ancient Greeks.

Hence, before we can define a measure, we must describe a suitable domain for it. This domain will be a collection of subsets of the space $\mathbf{X}$, called a **sigma-algebra**.

1.1(b) Sigma Algebras

Definition 1 Sigma-algebra, Measurable Space

Let $\mathbf{X}$ *be a set. A* **sigma algebra** *over* $\mathbf{X}$ *is a collection* $\mathcal{X}$ *of subsets of* $\mathbf{X}$ *with the following properties:*

1. *$\mathcal{X}$ is closed under* **countable unions.** *In other words, if $\mathbf{U}_1, \mathbf{U}_2, \ldots$, are in $\mathcal{X}$, then their union $\bigcup_{n=1}^{\infty} \mathbf{U}_n$ is also in $\mathcal{X}$ (Figure 1.2A).*

2. *$\mathcal{X}$ is closed under* **countable intersections.** *If $\mathbf{U}_1, \mathbf{U}_2, \ldots$, are in $\mathcal{X}$, then their intersection $\bigcap_{n=1}^{\infty} \mathbf{U}_n$ is also in $\mathcal{X}$ (Figure 1.2B).*

3. *$\mathcal{X}$ is closed under* **complementation**; *if $\mathbf{U}$ is in $\mathcal{X}$, then $\mathbf{U}^{\complement} = (\mathbf{X} \setminus \mathbf{U})$ is also in $\mathcal{X}$.*

A **measurable space** *is an ordered pair $(\mathbf{X}, \mathcal{X})$, where $\mathbf{X}$ is a set, and $\mathcal{X}$ is a sigma-algebra on $\mathbf{X}$.*

Example 2: Trivial Sigma Algebras

For any set $\mathbf{X}$,

1. The collection $\{\emptyset, \mathbf{X}\}$ is a sigma-algebra. 2. The power set $\mathcal{P}(\mathbf{X})$ is a sigma-algebra.

The first of these is clearly far too small to do anything useful; the second is generally too large to be manageable.

One way to generate a 'manageable' sigma-algebra is to start with some collection $\mathcal{M}$ of 'manageable' sets, and then find the smallest sigma-algebra which contains all elements of $\mathcal{M}$. This is called the sigma-algebra **generated** by $\mathcal{M}$, and sometimes denoted "$\sigma(\mathcal{M})$".

Example 3: (co)Countable sets

The most conservative collection of 'manageable' sets is

$$\mathcal{M} \quad := \quad \{\{x\} \; ; \; x \in \mathbf{X}\},$$

the singleton subsets of $\mathbf{X}$. Then $\mathcal{C} = \sigma(\mathcal{M})$ is the sigma-algebra of **countable and co-countable sets**. That is,

$$\mathcal{C} \quad := \quad \{\mathbf{C} \subset \mathbf{X} \; ; \; \text{either } \mathbf{C} \text{ is countable, or } \mathbf{X} \setminus \mathbf{C} \text{ is countable}\}.$$

Exercise 1 Verify this.

Exercise 2 Verify: If $\mathbf{X}$ itself is finite or countable, then $\mathcal{C} = \mathcal{P}(\mathbf{X})$.

Example 4: Partition Algebras

Let $\mathbf{X}$ be a set. Figure 1.3 shows a **partition** of $\mathbf{X}$: a collection $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ of disjoint subsets, such that $\mathbf{X} = \bigsqcup_{n=1}^{N} \mathbf{P}_n$. The sets $\mathbf{P}_1, \ldots \mathbf{P}_N$ are called the **atoms** of the partition. Figure 1.4 shows the sigma-algebra *generated* by $\mathcal{P}$: the collection of all possible unions of $\mathcal{P}$-atoms:

$$\sigma(\mathcal{P}) \;=\; \{\mathbf{P}_{n_1} \sqcup \mathbf{P}_{n_2} \sqcup \ldots \sqcup \mathbf{P}_{n_k} \; ; \; n_1, n_2, \ldots, n_k \in [1..N]\}$$

Thus, if $\mathsf{card}\,[\mathbf{P}] = N$, then $\mathsf{card}\,[\sigma(\mathbf{P})] = 2^N$.

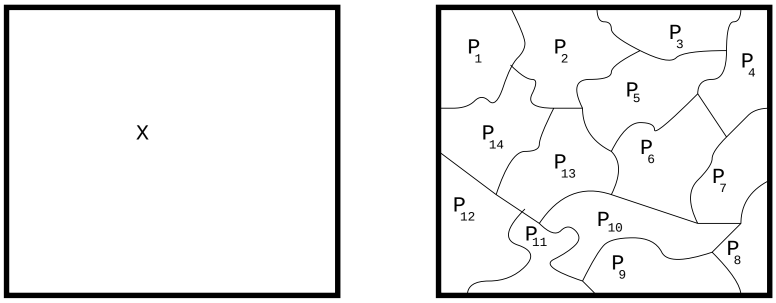


Figure 1.3: $\mathcal{P}$ is a partition of $\mathbf{X}$.

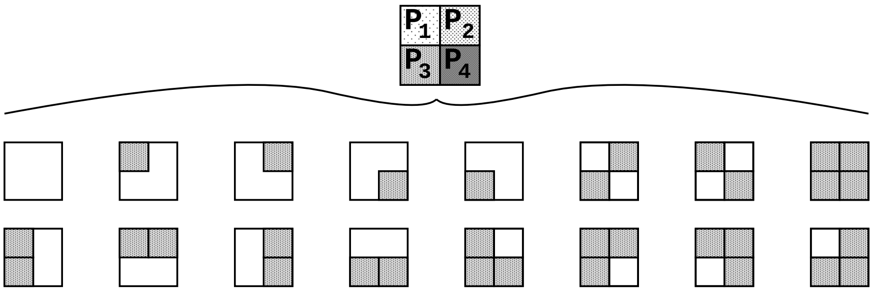


Figure 1.4: **The sigma-algebra generated by a partition:** Partition the square into four smaller squares, so $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4\}$. The corresponding sigma-algebra contains 16 elements.

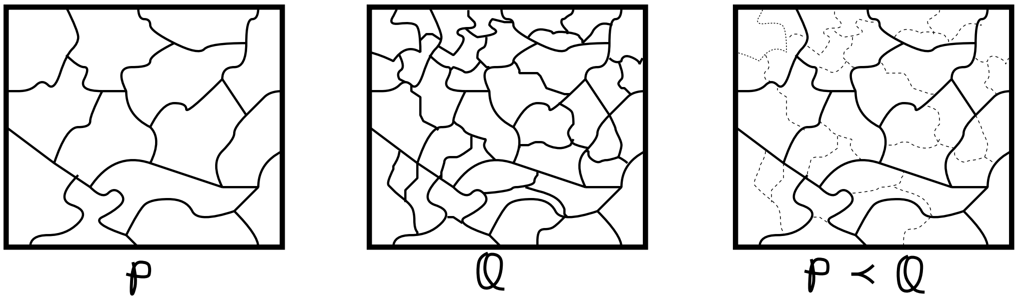


Figure 1.5: Partition $\mathcal{Q}$ refines $\mathcal{P}$ if every element of $\mathcal{P}$ is a union of elements in $\mathcal{Q}$.

If $\mathcal{Q}$ is another partition, we say that $\mathcal{Q}$ **refines** $\mathcal{P}$ if, for every $\mathbf{P} \in \mathcal{P}$, there are $\mathbf{Q}_1, \ldots, \mathbf{Q}_N \in \mathcal{Q}$ so that $\mathbf{P} = \bigsqcup_{n=1}^{N} \mathbf{Q}_P$; see Figure 1.5. We then write "$\mathcal{P} \prec \mathcal{Q}$" It follows (**Exercise 3**) that

$$\left(\mathcal{P} \prec \mathcal{Q} \right) \iff \left(\sigma(\mathcal{P}) \subset \sigma(\mathcal{Q}) \right)$$

Example 5: Borel Sigma-algebra of $\mathbb{R}$

Let $\mathbf{X} = \mathbb{R}$ be the real numbers, and let $\mathcal{M}$ be the set of all open intervals in $\mathbb{R}$:

$$\mathcal{M} = \{(a, b) \; ; \; -\infty \leq a < b \leq \infty\}$$

Then the sigma algebra $\mathcal{B} = \sigma(\mathcal{M})$ contains all open subsets of $\mathbb{R}$, all closed subsets, all countable intersections of open subsets, countable unions of closed subsets, etc. For example, $\mathcal{B}$ contains, as elements, the set $\mathbb{Z}$ of integers, the set $\mathbb{Q}$ of rationals, and the set $\not{\mathbb{Q}}$ of irrationals. $\mathcal{B}$ is called the **Borel sigma algebra** of $\mathbb{R}$.

Exercise 4 Verify these claims about the contents of $\mathcal{B}$.

Exercise 5 Show that $\mathcal{B}$ is also generated by any of the following collections of subsets:

$$\begin{aligned}
\mathcal{M} &= \{[a, b] \; ; \; -\infty \leq a < b \leq \infty\} && \text{(all closed intervals)} \\
\mathcal{M} &= \{[a, b) \; ; \; -\infty \leq a < b \leq \infty\} && \text{(right-open intervals)} \\
\mathcal{M} &= \{(\infty, b] \; ; \; -\infty < b \leq \infty\} && \text{(half-infinite closed intervals)} \\
\mathcal{M} &= \{[a, b) \; ; \; -\infty < a < b < \infty\} && \text{(strictly finite intervals)}
\end{aligned}$$

Example 6: Borel Sigma-algebras

Let $\mathbf{X}$ be any topological space, and let $\mathcal{M}$ be the set of all open subsets of $\mathbf{X}$. The sigma algebra $\sigma(\mathcal{M})$ is the **Borel sigma algebra** of $\mathbf{X}$, and denote $\mathcal{B}(\mathbf{X})$. It contains all open sets and closed subsets of $\mathbf{X}$, all countable intersections of open sets (called $G\delta$ **sets**), all countable unions of closed sets (called $F\sigma$ **sets**), etc. For example, $\mathbf{X}$ is Hausdorff, then $\mathcal{B}(\mathbf{X})$ contains all countable and cocountable sets.

Example 7: Product Sigma algebras

Suppose $(\mathbf{X}, \mathcal{X})$ and $(\mathbf{Y}, \mathcal{Y})$ are two measurable spaces, and consider the Cartesian product $\mathbf{X} \times \mathbf{Y}$. Let

$$\mathcal{M} \;=\; \{\mathbf{U} \times \mathbf{V} \; ; \; \mathbf{U} \in \mathcal{X}, \; \mathbf{V} \in \mathcal{Y}\} \qquad \text{(see Figure 1.6A)}$$

be the set of all 'rectangles' in $\mathbf{X} \times \mathbf{Y}$. Then $\sigma(\mathcal{M})$ is called the **product sigma-algebra**, and denoted $\mathcal{X} \otimes \mathcal{Y}$. Thus, $\mathcal{X} \otimes \mathcal{Y}$ contains all rectangles, all countable unions of rectangles, all countable intersections of unions, etc. (see Figure 1.6B).

For example: if $\mathbf{X}$ and $\mathbf{Y}$ are topological spaces with Borel sigma algebras $\mathcal{X}$ and $\mathcal{Y}$, and we endow $\mathbf{X} \times \mathbf{Y}$ with the product topology, then $\mathcal{X} \otimes \mathcal{Y}$ is the Borel sigma algebra of $\mathbf{X} \times \mathbf{Y}$. (**Exercise 6**)

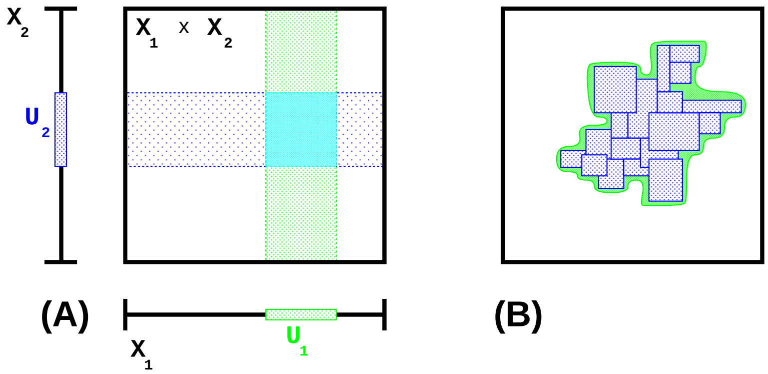


Figure 1.6: The product sigma-algebra.

We will discuss the product sigma algebra further in §2.1(b), where we construct the **product measure**; see Examples (59b) and (61b).

Example 8: Cylinder Sigma algebras

Suppose $(\mathbf{X}_\lambda, \mathcal{X}_\lambda)$ be measurable spaces for all $\lambda \in \Lambda$, where Λ is some (possibly uncountably infinite) indexing set. Consider the Cartesian product $\mathbf{X} = \prod_{\lambda\in\Lambda} \mathbf{X}_\lambda$. Let

$$\mathcal{M} \;=\; \left\{ \prod_{\lambda\in\Lambda} \mathbf{U}_\lambda \;;\; \forall\lambda\in\Lambda, \quad \mathbf{U}_\lambda \in \mathcal{X}_\lambda, \quad \text{and } \mathbf{U}_\lambda = \mathbf{X}_\lambda \text{ for all but finitely many } \lambda \right\}.$$

Such subsets are called **cylinder sets** in $\mathbf{X}$, and $\sigma(\mathcal{M})$ is called the **cylinder sigma-algebra**, and denoted $\bigotimes_{\lambda\in\Lambda} \mathcal{X}_\lambda$.

If $\mathbf{X}_\lambda$ are topological spaces with Borel sigma algebras $\mathcal{X}_\lambda$, and we endow $\mathbf{X}$ with the (Tychonoff) product topology, then $\bigotimes_{\lambda\in\Lambda} \mathcal{X}_\lambda$ is the Borel sigma algebra of $\mathbf{X}$.

Exercise 7 Verify this.

1.1(c) Measures

Prerequisites: §1.1(b) **Recommended:** §1.1(a)

Definition 9 Measure, Measure Space

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space. A* **measure** *on* $\mathcal{X}$ *is a map* $\mu : \mathcal{X} \longrightarrow [0, \infty]$ *which is* **countably additive**, *in the sense that, if* $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ *are all elements of* $\mathcal{X}$, *and are disjoint, then:*

$$\mu\left[\bigsqcup_{n=1}^{\infty} \mathbf{Y}_n\right] \;=\; \sum_{n=1}^{\infty} \mu\left[\mathbf{Y}_n\right]$$

A **measure space** *is an ordered triple* $(\mathbf{X}, \mathcal{X}, \mu)$*, where* $\mathbf{X}$ *is a set,* $\mathcal{X}$ *is a sigma-algebra and* μ *is a measure on* $\mathcal{X}$*.*

Thus, μ assigns a 'size' to the $\mathcal{X}$-measurable subsets of $\mathbf{X}$. Unfortunately, a rigorous construction of a nontrivial example of a measure is quite technically complicated. Therefore, we will first provide a *nonrigorous* overview of the most important sorts of measures.

$\langle$**i**$\rangle$ **The Counting Measure:** The **counting measure** assigns, to any set, the **cardinality** of that set:

$$\mu[\mathbf{S}] \quad := \quad \mathsf{card}[\mathbf{S}].$$

Of course, the counting measure of any infinite set is simply "∞"; the counting measure provides no means of distinguishing between 'large' infinite sets and small ones. Thus, it is only really very useful in finite measure spaces.

$\langle$**ii**$\rangle$ **Finite Measure spaces:** Suppose $\mathbf{X}$ is a finite set, and $\mathcal{X} = \mathcal{P}(\mathbf{X})$. Then a measure μ on $\mathbf{X}$ is entirely defined by some function $f : \mathbf{X} \longrightarrow [0, \infty]$; for any subset $\{x_1, x_2, \ldots, x_N\}$, we then define

$$\mu\{x_1, x_2, \ldots, x_N\} \quad = \quad \sum_{n=1}^{N} f(x_n)$$

Exercise 8 Show that every measure on $\mathbf{X}$ arises in this manner.

$\langle$**iii**$\rangle$ **Discrete Measures:** If $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, then an **atom** of μ is a subset $\mathbf{A} \in \mathcal{X}$ such that:

1. $\mu[\mathbf{A}] \; = \; A > 0$;

2. For any $\mathbf{B} \subset \mathbf{A}$, either $\mu[\mathbf{B}] = A$ or $\mu[\mathbf{B}] = 0$.

For example, in the finite measure space above, the singleton set $\{x_n\}$ is an atom if $f(x_n) > 0$.

The measure space $(\mathbf{X}, \mathcal{X}, \mu)$ is called **discrete** if we can write:

$$\mathbf{X} \quad = \quad \mathbf{Z} \sqcup \bigsqcup_{n=1}^{\infty} \mathbf{A}_n$$

where $\mu[\mathbf{Z}] = 0$ and where $\{\mathbf{A}_n\}_{n=1}^{\infty}$ is a collection of atoms.

Example 10:

(a) The set $\mathbb{Z}$, endowed with the counting measure, is discrete.

(b) Any *finite* measure space is discrete.

(c) If $\mathbf{X}$ is a countable set and μ is any measure, then $(\mathbf{X}, \mathcal{X}, \mu)$ must be discrete (**Exercise 9**).

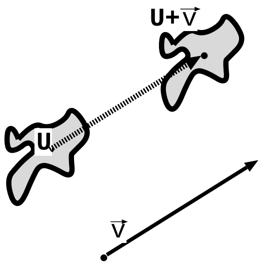


Figure 1.7: **The Haar Measure:** The Haar measure of $\mathbf{U}$ is the same as that of $\mathbf{U} + \vec{v}$

$\langle$**iv**$\rangle$ **The Lebesgue measure:** The **Lebesgue Measure** on $\mathbb{R}^n$ is the model of "length" (when $n = 1$), "area" (when $n = 2$), "volume" (when $n = 3$), etc. We will construct this measure in §2.1. For the most part, your geometric intuitions about lengths and volumes suffice to understand the Lebesgue measure, but certain 'pathological' subsets can have counterintuitive properties.

$\langle$**v**$\rangle$ **Haar Measures:** The Lebesgue measure has the extremely important property of **translation invariance**; that is, for any set $\mathbf{U} \subset \mathbb{R}^n$, and any element $\vec{v} \in \mathbb{R}^n$, we have

$$\mu[\mathbf{U}] \quad = \quad \mu[\mathbf{U} + \vec{v}].$$

(see Figure 1.7.)

We can generalise this property to any topological group, $\mathbf{G}$. Let $\mathbf{G}$ have Borel sigma-algebra $\mathcal{B}$, and suppose η is a measure on $\mathcal{B}$ so that, for any $\mathbf{B} \in \mathcal{B}$, and $g \in \mathbf{G}$,

$$\eta[\mathbf{B}.g] \quad = \quad \eta[\mathbf{B}].$$

This is called **right translation invariance**[1]. If $\mathbf{G}$ is *locally compact* and *Hausdorff*, then there is a measure η satisfying this property, and η is *unique* (up to multiplication by some constant, which can be thought of as a choice of 'scale'). We call η the right **Haar measure** on $\mathbf{G}$.

Now consider *left*-**translation invariance**: For any $g \in \mathbf{G}$ and $\mathbf{B} \in \mathcal{B}$,

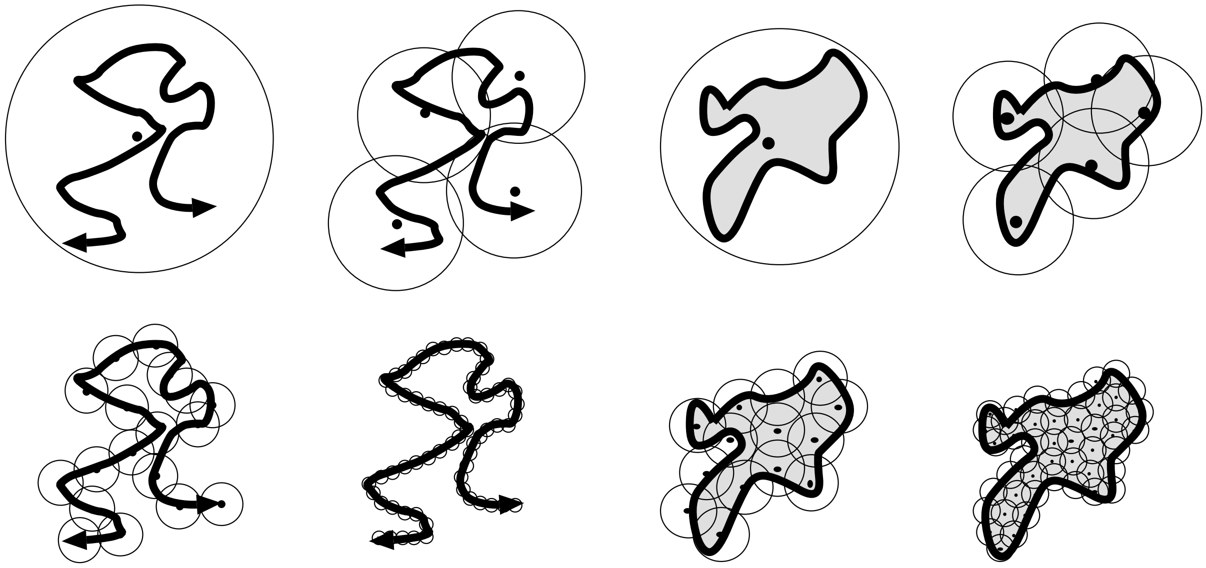
$$\eta[g.\mathbf{B}] \quad = \quad \eta[\mathbf{B}].$$

Again, there is a unique measure with this property; the *left* **Haar measure** on $\mathbf{G}$.

If $\mathbf{G}$ is abelian, then left-invariance and right-invariance are equivalent, and therefore, the two Haar measures agree. If $\mathbf{G}$ is nonabelian, however, the two measures may disagree. When the left- and right- Haar measures are equal, $\mathbf{G}$ is called **unimodular**.

Example 11:

[1]Remember: $\mathbf{G}$ may not be abelian, so left- and right- multiplication may behave differently.



(A) As we attempt to cover the curve with balls of smaller and smaller radius, we achieve a better and better approximation of its length. Note that the number of balls of radius R required for a covering grows approximately at a rate of $\frac{1}{R}$. Thus, the Hausdorff dimension of the curve is 1.

(B) As we attempt to cover the blob with balls of smaller and smaller radius, we achieve a better and better approximation of its area. Note that the number of balls of radius R required for a covering grows approximately at a rate of $\frac{1}{R^2}$. Thus, the Hausdorff dimension of the blob is 2.

Figure 1.8: The Hausdorff measure.

(a) Let $\mathbf{G}$ be a finite group, of cardinality G. Define $\eta\{g\} = 1/G$ for all $g \in \mathbf{G}$. Then η is a Haar measure on $\mathbf{G}$, such that $\eta[\mathbf{G}] = 1$ (**Exercise 10**).

(b) If $\mathbf{G} = \mathbb{Z}$, then the Haar measure is just the counting measure (**Exercise 11**).

(c) Suppose $\mathbf{G} = \mathbb{R}^n$, treated as a topological group under addition. Then the Haar measure of $\mathbb{R}^n$ is just the Lebesgue measure.

(d) Suppose $\mathbf{G} = \mathbb{S}^1 = \{z \in \mathbb{C} \,;\, |z| = 1\}$, treated as a topological group under complex multiplication. The Haar measure on $\mathbb{S}^1$ is just the natural measure of arc-length on a circle.

We will formally construct the Haar measure in §2.1(c) (Proposition 66 part 2), and discuss its properties in §**??**.

⟨vi⟩ Hausdorff Measure: The Lebesgue measure is a special case of another kind of measure. Instead of treating $\mathbb{R}^n$ as a topological group, regard $\mathbb{R}^n$ as a *metric space.* On any metric space, there is a natural measure called the **Hausdorff measure**.

Heuristically, the Hausdorff measure of a set $\mathbf{U}$ is determined by counting the number of open balls of small radius needed to *cover* $\mathbf{U}$. The more balls we need, the larger $\mathbf{U}$ must be.

However, for any nonzero radius R, a covering with balls of size R produces only an *approximate* measure of the size of $\mathbf{U}$, because any features of $\mathbf{U}$ which are much smaller than R are not 'detected' by such a covering. The Hausdorff measure is determined by looking at the *limit* of the number of balls needed, as R goes to zero. (see Figure 1.8(A))

It is possible to define a Hausdorff measure μ_d for any **dimension** $d \in (0, \infty)$. For example: the Lebesgue measure on $\mathbb{R}^n$ is a Hausdorff measure of dimension n. Note, that the dimension parameter d is allowed to take on *non-integer values*.

The dimension d describes how rapidly the measure of a ball of radius ϵ grows as a function of ϵ; we expect that, for any point x in our space,

$$\mu[\mathbf{B}(x, \epsilon)] \quad \sim \quad \epsilon^d.$$

For any metric space $\mathbf{X}$, there is a unique choice of dimension d_0 that yields a nontrivial Hausdorff dimension. For any value of $d > d_0$, the measure μ_d will assign every set measure *zero*, and for any value of $d < d_0$, the measure μ_d will assign every open set *infinite* measure.

The unique value d_0 is called the **Hausdorff Dimension** of the space $\mathbf{X}$, and carries important information about the geometry of $\mathbf{X}$. For example: the Hausdorff measure of $\mathbb{R}^n$ is n. Hence, if we tried to measure the 'volume' of subsets of $\mathbb{R}^2$, we would expect to get the value zero. Conversely, if we tried to measure the 'area' of a nontrivial subset of $\mathbb{R}^3$, the only sensible value to expect would be ∞. (see Figure 1.8 on the page before(B)).

We can construct a Haar measure on any subset $\mathbf{U}$ of $\mathbb{R}^n$, by treating $\mathbf{U}$ as a metric space under the restriction of the natural metric on $\mathbb{R}^n$. For example: if $\mathbf{U}$ is an embedded k-dimensional **manifold**, then the Hausdorff dimension of $\mathbf{U}$ is, k. However, there are also 'pathological' subsets of $\mathbb{R}^n$ which possess *non-integer* Hausdorff dimension. Benoit Mandelbrot has coined the term **fractal** to describe such strange objects, and the Hausdorff dimension is only one of many **fractal dimensions** which are used to characterise these objects.

We will formally construct the Hausdorff measure in §2.1(c) (Proposition 66 part 1), and discuss its properties in §**??**.

⟨vii⟩ Stieltjes Measures: If we see $\mathbb{R}$ as a *group*, then the Lebesgue measure is the *Haar* measure; if we see $\mathbb{R}$ as a *metric space*, then the Lebesgue measure arises as a *Hausdorff* measure. If we instead treat $\mathbb{R}$ as an **ordered set**, then the Lebesgue measure arises as a particular kind of **Stieltjes measure**.

One way to imagine an Stieltjes measure is with an analogy from commerce. Suppose you wanted to determine the *net profit* made by a company during the time interval between January 5, 1998, and March 27, 2005. One simple way to do this would be to simply compare the *net worth* of the company on these two successive dates; the net profit of the company would simply be the *difference* between its net worth on March 27, 2005 and its net worth on January 5, 1998.

If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is the function describing the net worth of the company over time, then f is a continuous, nondecreasing function[2]. The net profit over any time interval $(a, b]$ is thus

[2] Assuming the company is profitable!

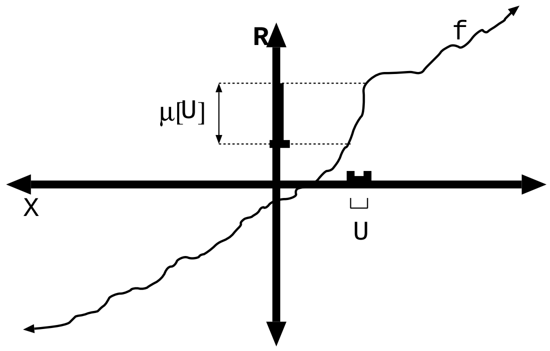


Figure 1.9: **Stieltjes measures:** The measure of the set $\mathbf{U}$ is the amount of height "accumulated" by f as we move from one end of $\mathbf{U}$ to the other.

$f(b) - f(a)$; this recipe defines a measure on $\mathbb{R}$.

More generally, given an ordered set $(\mathbf{X}, <)$, we can define a measure as follows. First we define our sigma-algebra $\mathcal{X}$ to be the sigma-algebra generated by all **left-open intervals** of the form $(a, b]$, where, for any a and b in $\mathbf{X}$,

$$(a, b] \quad := \quad \{x \in \mathbf{X} \; ; \; a < x \leq b\}.$$

Observe that, if $\mathbf{X} = \mathbb{R}$ with the usual linear ordering, then this $\mathcal{X}$ is just the usual Borel sigma-algebra from Example 5 on page 5.

Now suppose that $f : \mathbf{X} \longrightarrow \mathbb{R}$ is a right-continuous, nondecreasing function[3] Define the measure of any interval $(a, b]$ to be simply the *difference* between the value of f at the two endpoints a and b:

$$\mu_f(a, b] \quad := \quad f(b) - f(a)$$

(see Figure 1.9)

We then extend this measure to the rest of the elements of $\mathcal{X}$ by approximating them with disjoint unions of left-open intervals.

We call μ_f a **Stieltjes measure**, and call f the **accumulation function** or **cumulative distribution** of μ_f. Under suitable conditions, *every* measure on $(\mathbf{X}, \mathcal{X})$ can be generated in this way. Starting with an arbitrary measure μ, find a "zero point" $x_0 \in \mathbf{X}$, so that,

- $\mu(x_0, x]$ is finite for all $x > x_0$.
- $\mu(x, x_0]$ is finite for all $x < x_0$.

Then define the function $f : \mathbf{X} \longrightarrow \mathbb{R}$ by:

$$f(x) \quad := \quad \begin{cases} \mu(x_0, x] & \text{if} \quad x > x_0 \\ -\mu(x, x_0] & \text{if} \quad x < x_0 \end{cases}$$

[3]That is: if $x < y$, then $f(x) \leq f(y)$, and $\sup_{x<y} f(x) = f(y)$.

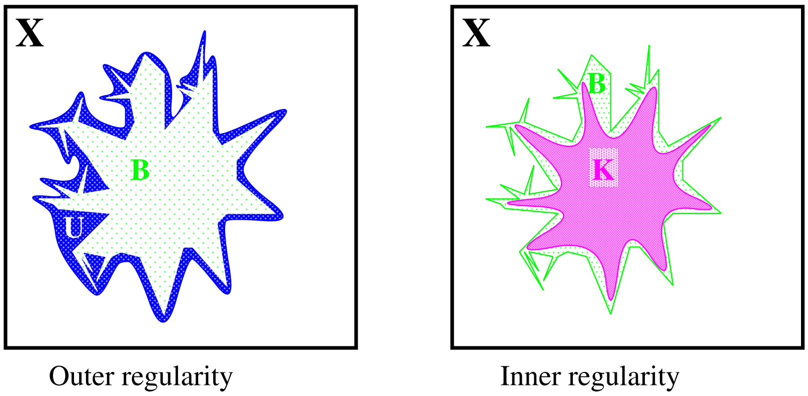


Figure 1.10: **(A)** Outer regularity: The measure of $\mathbf{B}$ is well-approximated by a slightly larger open set $\mathbf{U}$. **(B)** Inner regularity: The measure of $\mathbf{B}$ is well-approximated by a slightly smaller compact set $\mathbf{K}$.

For example, the Lebesgue Measure on $\mathbb{R}$ is obtained from accumulation function

$$f : \mathbb{R} \ni x \mapsto x \in \mathbb{R}.$$

We will formally construct the Stieltjes measure in §2.1(b), and discuss its properties in §2.2.

$\langle$**viii**$\rangle$ **Radon Measures:** The Lebesgue measure, the Haar measure, the Hausdorff measure, and accumulation measures are all measures defined on the Borel sigma-algebras of topological spaces, and are examples of *Radon Measures*. This means that, any measurable subset (no matter how pathological) can be well-approximated from above by slightly larger *open* sets containing $\mathbf{Y}$, and approximated from below by slightly smaller *compact* sets contained *within* $\mathbf{Y}$.

Formally, if $\mathbf{X}$ is a topological space with Borel sigma algebra $\mathcal{X}$, then μ is a **Radon measure** if it has two properties:

Outer regularity: For any $\mathbf{B} \subset \mathcal{B}$, and any $\epsilon > 0$, there is an open set $\mathbf{U} \supset \mathbf{B}$ so that $\mu[\mathbf{U} \setminus \mathbf{B}] < \epsilon$.

Inner regularity: For any $\mathbf{B} \subset \mathcal{B}$, and any $\epsilon > 0$, there is a compact set $\mathbf{K} \subset \mathbf{B}$ so that $\mu[\mathbf{B} \setminus \mathbf{K}] < \epsilon$.

Thus, we can understand the behaviour of μ by understanding how μ acts on 'nice' subsets of $\mathbf{X}$. For example, we can determine an accumulation measure from its values on half-open intervals.

If μ is a Radon measure on $\mathbf{X}$, then let $\mathbf{X}_0 \subset \mathbf{X}$ be the maximal open subset of $\mathbf{X}$ of measure zero. Formally:

$$\mathbf{X}_0 = \bigcup_{\substack{\text{open } \mathbf{U} \subset \mathbf{X} \\ \mu[\mathbf{U}]=0}} \mathbf{U}$$

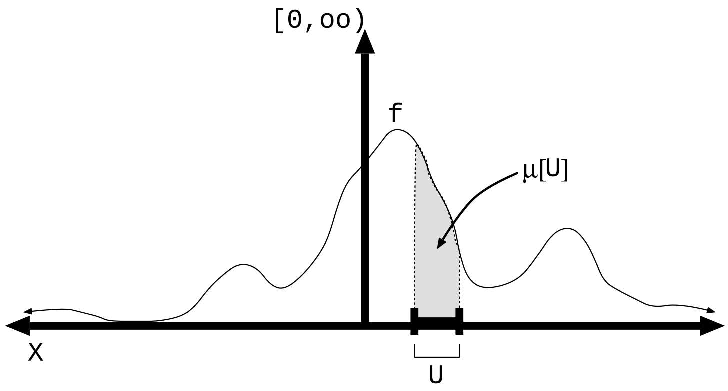


Figure 1.11: **Density Functions:** The measure of $\mathbf{U}$ can be thought of as the 'area under the curve' of the function f in the region over $\mathbf{U}$.

The **support** of μ is then the set $\mathsf{supp}\,[\mu] = \mathbf{X} \setminus \mathbf{X}_0$; thus, $\mathsf{supp}\,[\mu]$ is closed, and $\mu[\mathbf{U}] > 0$ for any open $\mathbf{U} \subset \mathsf{supp}\,[\mu]$. If $\mathsf{supp}\,[\mu] = \mathbf{X}$, then we say μ has **full support**. For example

1. The Lebesgue measure has full support on $\mathbb{R}$.

2. If $\mathbf{G}$ is a topological group, then any (nonzero) Haar measure has full support (**Exercise 12**).

3. Let μ be the measure on $\mathbb{R}$ such that $\mu\{n\} = 1$ for any $n \in \mathbb{Z}$, but $\mu(n, n+1) = 0$. Then $\mathsf{supp}\,[\mu] = \mathbb{Z}$, so that μ does *not* have full support on $\mathbb{R}$.

$\langle$**ix**$\rangle$ **Density Functions:** Recall the commercial analogy we used to illustrate *accumulation measures*. Suppose, once again, that you wanted to determine the *net profit* made by a company during the time interval between January 5, 1998, and March 27, 2005. The 'accumulation measure' method for doing this would be to subtract the *net worth* of the company on these two successive dates. Another approach would be to add up the company's daily net profits on *each* of the 2638 days between these two successive dates; the total amount would then be the net profit over the entire time interval. The mathematical model of this approach is the *density function*.

Let $\rho : \mathbb{R}^n \longrightarrow [0, \infty)$ be a positive, continuous[4] function on $\mathbb{R}^n$. For any $\mathbf{B} \in \mathcal{B}(\mathbb{R}^n)$, define

$$\mu_\rho(\mathbf{B}) := \int_{\mathbf{B}} \rho \qquad \text{(see Figure 1.11)}$$

Then μ_ρ is a measure. We call ρ the **density function** for μ. If we ρ describes the *density* of the distribution of matter in physical space, then μ_ρ measures the *mass* contained within

[4]Actually, we only need ρ to be *integrable*. We will formally define 'integrable' in §**??** it means what you think it means.

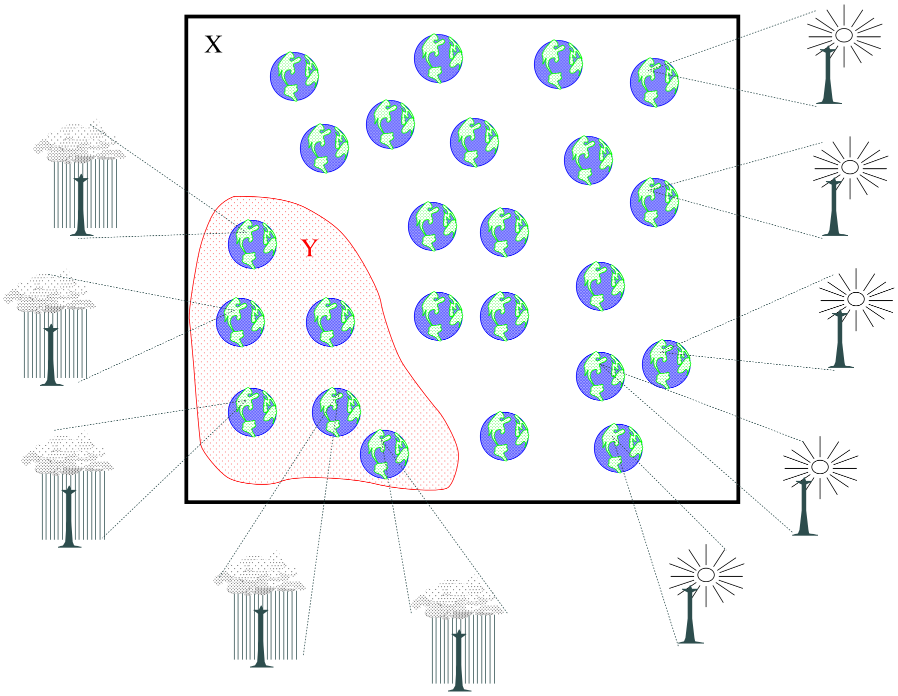


Figure 1.12: **Probability Measures: X** is the 'set of all possible worlds'. $\mathbf{Y} \subset \mathbf{X}$ is the set of all worlds where it is raining in Toronto.

any region of space. Alternately, if ρ describes *charge density* (ie. the physical distribution of charged particles), then μ_ρ measures the *charge* contained in a region of space.

It is important not to get *density functions* confused with *accumulation functions*: they determine measures in very different ways. Indeed, the Fundamental Theorem of Calculus basically says that a density functions are the 'derivatives' of an accumulation functions (see Theorem 74 on page 57).

$\langle$x$\rangle$ **Probability Measures:** A measure μ on **X** is a **probability measure** if $\mu[\mathbf{X}] = 1$ —ie. the 'total mass' of the space is 1. We say that $(\mathbf{X}, \mathcal{X}, \mu)$ is a **probability space.**

Imagine that **X** is the space of possible **states** of some "universe" $\mathcal{U}$. For example, in Figure 1.12, $\mathcal{U}$ is the weather over Toronto; thus, **X** is the set of all possible weather conditions. The elements of the sigma-algebra $\mathcal{X}$ are called **events**; each event corresponds to some assertion about $\mathcal{U}$. For example, in Figure 1.12, the assertion "It is raining in Toronto" corresponds to the event $\mathbf{Y} \subset \mathbf{X}$ of all states in **X** where it is, in fact, raining in the Toronto of the universe $\mathcal{U}$. The measure $\mu[\mathbf{Y}]$ is then the **probability** of this event being true (ie. the probability that it is, in fact, raining in Toronto).

Example 12:

(a) **Dice:** Imagine a six-sided die. In this case, the $\mathbf{X} = \{1, 2, 3, 4, 5, 6\}$, and $\mathcal{X} = \mathcal{P}(\mathbf{X})$. The probability measure μ is completely determined by the values of $\mu\{1\}$, $\mu\{2\}, \ldots, \mu\{6\}$.

For example, suppose:

$$\begin{array}{rclcrcl} \mu\{1\} &=& 1/12 & \quad & \mu\{4\} &=& 1/6 \\ \mu\{2\} &=& 1/12 & & \mu\{5\} &=& 1/6 \\ \mu\{3\} &=& 1/3 & & \mu\{6\} &=& 1/6 \end{array}$$

Then the probability of rolling a 2 *or* a 3 is $\mu\{2,3\} = 1/12 + 1/3 = 5/12$.

(b) **Urns:** Imagine a giant clay 'urn' full of 10 000 balls of various colours. You reach in and grab a ball randomly. What is the probability of getting a ball of a particular colour?

In this case, $\mathbf{X}$ is the set of balls. If we label the balls with numbers, then we can write $\mathbf{X} = \{1, 2, 3, \ldots, 10\ 000\}$. Again, $\mathcal{X} = \mathcal{P}(\mathbf{X})$. Now suppose that the balls come in colours red, green, blue, and purple. For simplicity, let

$$\begin{array}{rcll} \mathbf{R} &=& \{1, 2, 3, \ldots, 500\} & \text{be the set of all red balls;} \\ \mathbf{G} &=& \{501,\ 502,\ 503, \ldots, 1500\} & \text{be the set of all green balls;} \\ \mathbf{B} &=& \{1501,\ 1502,\ 1503, \ldots, 3000\} & \text{be the set of all blue balls;} \\ \mathbf{P} &=& \{3001,\ 3002,\ 3003, \ldots, 10\,00\} & \text{be the set of all purple balls;} \end{array}$$

Thus (assuming the urn is well-mixed and all balls are equally probable), the probability of getting a red ball is

$$\mu[\mathbf{R}] \quad = \quad \frac{500}{10000} = 0.05$$

while the probability of getting a green ball *or* a purple ball is

$$\mu[\mathbf{G} \sqcup \mathbf{P}] \quad = \quad \mu[\mathbf{G}] + \mu[\mathbf{P}] \quad = \quad \frac{1000}{10000} + \frac{7000}{10000} \quad = \quad 0.1 + 0.7 = 0.8$$

(c) **The unit interval:** Let $\mathbb{I} = [0,1]$ be the unit interval, with Borel sigma algebra $\mathcal{I}$, and let λ be the Lebesgue measure. Then $\lambda(\mathbb{I}) = 1$, so that $(\mathbb{I}, \mathcal{I}, \lambda)$ is a probability space.

(d) **The Gauss-Weierstrass distribution:** Define $g : \mathbb{R} \longrightarrow (0, \infty)$ by

$$g(x) \quad = \quad \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

and let μ be the probability measure on $\mathbb{R}$ with density f. In other words, for any measurable $\mathbf{U} \subset \mathbb{R}$, $\quad \mu[\mathbf{U}] \quad = \quad \int_{\mathbf{U}} g(x)\ dx$. This is called the **Gaussian** or **standard normal distribution**.

$\langle$**xi**$\rangle$ **Stochastic Processes:** A stochastic process is a particular kind of probability measure, which represents a system *randomly evolving in time.*

Imagine $\mathcal{S}$ is some complex system, evolving randomly. For example, $\mathcal{S}$ might be a die being repeatedly rolled, or a publically traded stock, a weather system. Let $\mathbf{X}$ be the set of all possible *states* of the system $\mathcal{S}$, and let $\mathbb{T}$ be a set representing *time*. For example:

- If $\mathcal{S}$ is a rolling die, then $\mathbf{X} = \{1, 2, 3, 4, 5, 6\}$, and $\mathbb{T} = \mathbb{N}$ indexes the successive dice rolls.
- If $\mathcal{S}$ is a publically traded stock, then its state is its price. Thus, $\mathbf{X} = \mathbb{R}$. Assume trading occurs continuously when the market is open, and let each trading 'day' have length $c < 1$. Then one representation of 'market time' is: $\mathbb{T} = \bigsqcup_{n=1}^{\infty} [n, n+c]$.
- If $\mathcal{S}$ is a weather system, then the 'state' of $\mathcal{S}$ can perhaps be described by a large array of numbers $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ representing the temperature, pressure, humidity, etc. at each point in space. Thus, $\mathbf{X} = \mathbb{R}^n$. Since the weather evolves continuously, $\mathbb{T} = \mathbb{R}$.

We represent the (random) evolution of $\mathcal{S}$ by assigning a probability to every possible *history* of $\mathcal{S}$. A history is an assignment of a state (in $\mathbf{X}$) to every moment in time (i.e. $\mathbb{T}$); in other words, it is a function $f : \mathbb{T} \longrightarrow \mathbf{X}$. The *set of all possible histories* is thus the space $\mathbf{H} = \mathbf{X}^{\mathbb{T}}$.

The sigma-algebra on $\mathbf{H}$ is usually a **cylinder algebra** of the type defined in Example 8. Suppose that $\mathbf{X}$ has sigma-algebra $\mathcal{X}$; then we give $\mathbf{H}$ the sigma-algebra $\mathcal{H} = \bigotimes_{t \in \mathbb{T}} \mathcal{X}_t$. An *event* —an element of $\mathcal{H}$ —thus corresponds to a cylinder set, a countable union of cylinder sets, etc. Suppose, for all $t \in \mathbb{T}$, that $\mathbf{U}_t \in \mathcal{X}$; with $\mathbf{U}_t = \mathbf{X}$ for all but finitely many t. The cylinder set $\mathbf{U} = \prod_{t \in \mathbb{T}} \mathbf{U}_t$ thus corresponds to the assertion, "For every $t \in \mathbb{T}$, at time t, the state of $\mathcal{S}$ was inside $\mathbf{U}_t$." A probability measure on $(\mathbf{H}, \mathcal{H})$ is then a way of assigning probabilities to such assertions.

Example 13:

Suppose that $\mathcal{S}$ was a (fair) six-sided die. Then $\mathbf{X} = \{1, 2, \ldots, 6\}$ and $\mathbb{T} = \mathbb{N}$. Thus, $\mathbf{H} = \{1, 2, \ldots, 6\}^{\mathbb{N}}$ is the set of all possible infinite sequences $\mathbf{x} = [x_1, x_2, x_3, \ldots]$ of elements $x_n \in \{1, 2, \ldots, 6\}$. Such a sequence represents a record of an infinite succession of dice throws. The sigma algebra $\mathcal{H}$ is generated by all cylinder sets of the form:

$$\langle y_1, y_2, \ldots, y_N \rangle \quad = \quad \left\{ \mathbf{x} \in \{1, 2, \ldots, 6\}^{\mathbb{N}} \; ; \; x_1 = y_1, \ldots, x_N = y_N \right\}$$

where $N \in \mathbb{N}$ and $y_1, y_2, \ldots, y_N \in \{1, 2, \ldots, 6\}$ are constants. For example,

$$\langle 3, 6, 2, 1 \rangle \quad = \quad \left\{ \mathbf{x} \in \{1, 2, \ldots, 6\}^{\mathbb{N}} \; ; \; x_1 = 3, \; x_2 = 6, \; x_3 = 2, \; x_4 = 1 \right\}$$

This event corresponds to the assertion, "The first time, you roll a *three*; the next time; a *six*, the third time, a *two*; and the fourth time, a *one*".

Assuming the die is fair, this event should have probability $\frac{1}{6^4} = \frac{1}{1296}$. More generally, for any $y_1, y_2, \ldots, y_N \in \{1, 2, \ldots, 6\}$, we should have:

$$\mu\langle y_1, y_2, \ldots, y_N\rangle \quad = \quad \frac{1}{6^N}$$

If the probabilities deviate from these values, we conclude the dice are loaded.

We will discuss the construction of stochastic processes in §**??**

1.2 Basic Properties and Constructions

Prerequisites: §1.1(c)

1.2(a) Continuity/Monotonicity Properties

Throughout this section, let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space.

Lemma 14 *If $\mathbf{A} \subset \mathbf{B} \subset \mathbf{X}$ are measurable, then $\mu[\mathbf{A}] \leq \mu[\mathbf{B}]$.*

Proof: Let $\mathbf{C} = \mathbf{B} \setminus \mathbf{A}$; then $\mathbf{C}$ is also measurable, and $\mu[\mathbf{B}] = \mu[\mathbf{A} \sqcup \mathbf{C}] = \mu[\mathbf{A}] + \mu[\mathbf{C}] \geq \mu[\mathbf{A}]$. □

Proposition 15 *Any measure μ has the following properties:*

Subadditivity: *If $\mathbf{U}_n \in \mathcal{X}$ for all $n \in \mathbb{N}$, then* $\mu\left[\bigcup_{n=1}^{\infty} \mathbf{U}_n\right] \;\leq\; \sum_{n=1}^{\infty} \mu[\mathbf{U}_n]$.

Lower-Continuity: *If $\mathbf{U}_1 \subset \mathbf{U}_2 \subset \ldots$, then* $\mu\left[\bigcup_{n=1}^{\infty} \mathbf{U}_n\right] \;=\; \lim_{n\to\infty} \mu[\mathbf{U}_n] \;=\; \sup_{n\in\mathbb{N}} \mu[\mathbf{U}_n]$.

Upper-Continuity: *If $\mathbf{U}_1 \supset \mathbf{U}_2 \supset \ldots$, then* $\mu\left[\bigcap_{n=1}^{\infty} \mathbf{U}_n\right] \;=\; \lim_{n\to\infty} \mu[\mathbf{U}_n] \;=\; \inf_{n\in\mathbb{N}} \mu[\mathbf{U}_n]$.

Proof: (1) For all $N \in \mathbb{N}$, define $\mathbf{V}_N = \bigcup_{n=1}^{N} \mathbf{U}_n$. Then $\mathbf{V}_1 \subset \mathbf{V}_2 \subset \ldots$, and $\bigcup_{n=1}^{\infty} \mathbf{U}_n = \bigcup_{n=1}^{\infty} \mathbf{V}_n$, so (1) follows from (2).

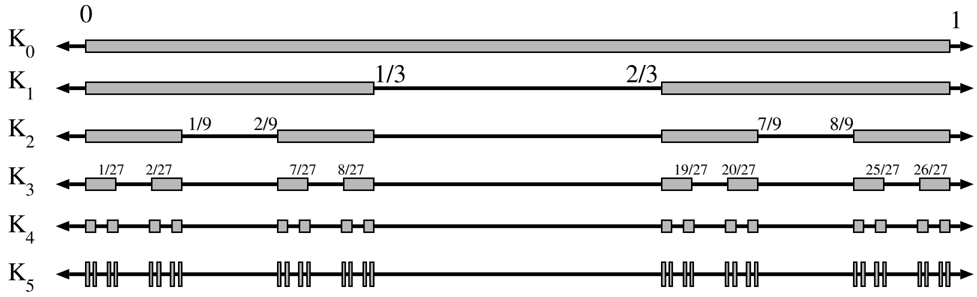


Figure 1.13: The Cantor set

(2) Define $\Delta_1 = \mathbf{U}_1$, and for all $n > 1$ defined $\Delta_n = \mathbf{U}_n \setminus \mathbf{U}_{n-1}$. Then the Δ_n are disjoint, and for any $N \in \mathbb{N}$, $\mathbf{U}_n = \bigsqcup_{n=1}^{N} \Delta_n$, while $\bigcup_{n=1}^{\infty} \mathbf{U}_n = \bigsqcup_{n=1}^{\infty} \Delta_n$. Thus

$$\mu\left[\bigcup_{n=1}^{\infty} \mathbf{U}_n\right] = \mu\left[\bigsqcup_{n=1}^{\infty} \Delta_n\right] = \sum_{n=1}^{\infty} \mu\left[\Delta_n\right] = \lim_{N\to\infty} \sum_{n=1}^{N} \mu\left[\Delta_n\right] = \lim_{N\to\infty} \mu\left[\bigsqcup_{n=1}^{N} \Delta_n\right] = \lim_{N\to\infty} \mu\left[\mathbf{U}_n\right]$$

(3) **Exercise 13** ______________________________ □

1.2(b) Sets of Measure Zero

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space. A **set of measure zero** is a subset $\mathbf{Z} \subset \mathbf{X}$ so that $\mu(\mathbf{Z}) = 0$.

Example 16:

(a) **Finite & Countable sets:** If $\mathbf{X} = \mathbb{R}$ and μ is the Lebesgue measure, then any finite or countable set has measure zero. In particular, the integers $\mathbb{Z}$ and the rational numbers $\mathbb{Q}$ have measure zero. However, the irrational numbers $\not{\mathbb{Q}}$ are *not* of measure zero.

Exercise 14 Verify these claims

(b) **The Cantor Set:** Any real number $\alpha \in [0, 1]$ has a unique[5] **trinary representation** $0.a_1a_2a_3a_4\ldots$ so that $\alpha = \sum_{n=0}^{\infty} \frac{a_n}{3^n}$. The **Cantor set** is defined (see Figure 1.13):

$$\mathbf{K} = \{\alpha \in [0, 1] \,;\, a_n = 0 \text{ or } 2, \quad \text{for all } n \in \mathbb{N}\}$$

This set is measurable and has measure zero.

Exercise 15 Verify this as follows:

[5]Well, *almost* unique. If $\alpha = 0.a_1a_2a_3\ldots a_{n-1}a_n000\ldots$, then we could also write $\alpha = 0.a_1a_2a_3\ldots a_{n-1}b_n2222\ldots$, where $b_n = a_n - 1$. This is analogous to the fact that $0.19999\ldots = 0.2$ in decimal notation.

Let $\mathbf{K}_0 = [0,1]$. Define $\mathbf{I}_1 = \left(\frac{1}{3}, \frac{2}{3}\right)$, and let $\mathbf{K}_1 = \mathbf{K}_0 \setminus \mathbf{I}_1$. Next let $\mathbf{I}_2 = \left(\frac{1}{9}, \frac{2}{9}\right) \sqcup \left(\frac{7}{9}, \frac{8}{9}\right)$, and let $\mathbf{K}_1 = \mathbf{K}_1 \setminus \mathbf{I}_2$. Proceed inductively, constructing $\mathbf{K}_{n+1}$ by deleting the 'middle thirds' from each of the 2^n intervals making up $\mathbf{K}_n$. Finally, define $\mathbf{K}_\infty = \bigcap_{n=1}^{\infty} \mathbf{K}_n$. Then verify:

1. Since $\mathbf{K}_\infty$ is a countable intersection of nested closed sets, it is closed and nonempty. In particular, it is measurable.
2. $\mathbf{K}_\infty = \mathbf{K}$, as defined above.
3. $\mu(\mathbf{K}_{n+1}) = \frac{2}{3}\mu(\mathbf{K}_n)$. Thus, $\mu(\mathbf{K}_\infty) = \lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$.

(c) **Submanifolds:** If $\mathbf{X} = \mathbb{R}^n$ and μ is the Lebesgue measure, then any submanifold of dimension less than n has measure zero. For example, curves in $\mathbb{R}^2$ have measure zero, and surfaces in $\mathbb{R}^3$ have measure zero.

(d) Let $\mathbf{X} = \mathbb{R}$ and let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a nondecreasing function, which we regard as the accumulation function for some measure. Then an interval $(a, b]$ has measure zero if and only if f is constant over $(a, b]$, so that $f(b) = \lim_{x \searrow a} f(x)$ (**Exercise 16**).

(e) Let $\mathbf{X} = \mathbb{R}$ and let $\rho : \mathbb{R} \longrightarrow [0, \infty)$ be a continuous, nonnegative function, which we regard as the density function for some measure. Then a subset $\mathbf{U} \subset \mathbf{X}$ has measure zero if and only if $\rho(u) = 0$ for all $u \in \mathbb{U}$ (exercise).

(f) **Closed subgroups:** Let $\mathbf{G}$ be a connected topological group, and $\mathbf{H} \subset \mathbf{G}$ be a closed proper subgroup. Then $\mathbf{H}$ has Haar measure zero (**Exercise 17**). For example, $\mathbb{Z}$ has Lebesgue measure zero in $\mathbb{R}$, and any vector subspace of $\mathbb{R}^n$ has Lebesgue measure zero. Note: although $\mathbb{Q} \subset \mathbb{R}$ also has Lebesgue measure zero, it is not for this reason (since $\mathbb{Q}$ is not a *closed* subgroup).

Definition 17 Complete Measure Space; Completion

A measure space $(\mathbf{X}, \mathcal{X}, \mu)$ is called **complete** *if, given any subset $\mathbf{Z} \in \mathcal{X}$ of measure zero,* all *subsets of $\mathbf{Z}$ are measurable. Formally:*

$$\left(\ \mu[\mathbf{Z}] = 0\ \right) \Longrightarrow \left(\ \mathcal{P}(\mathbf{Z}) \subset \mathcal{X}\ \right)$$

If $(\mathbf{X}, \mathcal{X}, \mu)$ is not complete, then the **completion** *of $\mathcal{X}$ is defined:*

$$\widetilde{\mathcal{X}} = \{\mathbf{U} \cup \mathbf{V}\ ;\ \mathbf{U} \in \mathcal{X},\ \textit{and}\ \ \mathbf{V} \subset \mathbf{Z}\ \ \textit{for some}\ \ \mathbf{Z} \in \mathcal{X}\quad \textit{with}\ \ \mu[\mathbf{Z}] = 0\}$$

We can then extend μ to a measure $\widetilde{\mu}$ on $\widetilde{\mathcal{X}}$ by defining $\widetilde{\mu}[\mathbf{U} \cup \mathbf{V}] = \mu[\mathbf{U}]$ for all such $\mathbf{U}$ and $\mathbf{V}$.

Exercise 18 *Verify that $\widetilde{\mathcal{X}}$ is a sigma-algebra, and that $\widetilde{\mu}$ is a measure.*

For example, the completion of the Borel sigma algebra on $\mathbb{R}^n$, with respect to the Lebesgue measure, is called the **Lebesgue sigma algebra**. Since we can complete any measure space in such a painless way, without affecting it's essential measure-theoretic properties, it is generally safe to assume that any measure space one works with is complete. This is analogous to the process of completing a metric space[6].

1.2(c) Measure Subspaces

Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, and $\mathbf{U} \subset \mathbf{X}$ is a measurable subset. If we define

$$\mathcal{U} \;=\; \{\mathbf{V} \in \mathcal{X} \;;\; \mathbf{V} \subset \mathbf{U}\}$$

then $\mathcal{U}$ is a sigma-algebra on $\mathbf{U}$, and $(\mathbf{U}, \mathcal{U})$ is a measurable space. Let $\mu\big|_{\mathbf{U}} := \mu\big|_{\mathcal{U}} : \mathcal{U} \longrightarrow [0, \infty]$ be the restriction of μ to $\mathcal{U}$; then $\mu\big|_{\mathbf{U}}$ is a measure on $(\mathbf{U}, \mathcal{U})$. We call $(\mathbf{U}, \mathcal{U}, \mu\big|_{\mathbf{U}})$ a **measure subspace** of $(\mathbf{X}, \mathcal{X}, \mu)$.

Example 18:

(a) Let $\mathbb{I} = [0, 1] \subset \mathbb{R}$, and let $\mathcal{I}$ be the Borel sigma-algebra of $\mathbb{I}$. Let λ be the restriction to $\mathcal{I}$ of the one-dimensional Lebesgue measure on $\mathbb{R}$; then $(\mathbb{I}, \mathcal{I}, \lambda)$ is a measure subspace of $\mathbb{R}$.

(b) Let $\mathbb{I}^2 = [0, 1] \times [0, 1] \subset \mathbb{R}^2$, and let $\mathcal{I}^2$ be the Borel sigma-algebra of $\mathbb{I}^2$. Let λ^2 be the restriction to $\mathcal{I}^2$ of the two-dimensional Lebesgue measure on $\mathbb{R}^2$; then $(\mathbb{I}^2, \mathcal{I}^2, \lambda^2)$ is a measure subspace of $\mathbb{R}^2$.

(c) Consider the rational numbers, $\mathbb{Q} \subset \mathbb{R}$; let $\mathcal{Q}$ be the Borel sigma-algebra of $\mathbb{Q}$, and let λ be the restriction of the Lebesgue measure to $\mathcal{Q}$. Then it is **Exercise 19** to verify: $\mathcal{Q} = \mathcal{P}(\mathbb{Q})$, and $\lambda \equiv 0$, so that $(\mathbb{Q}, \mathcal{P}(\mathbb{Q}), 0)$ is a measure subspace of $\mathbb{R}$

1.2(d) Disjoint Unions

Suppose $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are two measure spaces; assume that the sets $\mathbf{X}_1$ and $\mathbf{X}_2$ are disjoint. Let $\mathbf{X} = \mathbf{X}_1 \sqcup \mathbf{X}_2$. Then the collection

$$\mathcal{X} = \{\mathbf{U}_1 \sqcup \mathbf{U}_2 \;;\; \mathbf{U}_1 \in \mathcal{X}_1,\; \mathbf{U}_2 \in \mathcal{X}_2\};$$

is a sigma-algebra on $\mathbf{X}$ (**Exercise 20**). Define $\mu : \mathcal{X} \longrightarrow [0, \infty]$ by:

$$\mu[\mathbf{U}_1 \sqcup \mathbf{U}_2] \;=\; \mu_1[\mathbf{U}_1] + \mu_2[\mathbf{U}_2]$$

then μ is a measure on $\mathcal{X}$ (**Exercise 21**). The measure space $(\mathbf{X}, \mathcal{X}, \mu)$ is called the **disjoint union** of $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$.

[6] Indeed, as we will see, this analogy is very apt.

More generally, let Ω be a (poossibly infinite) indexing set, and for each $\omega \in \Omega$, let $(\mathbf{X}_\omega, \mathcal{X}_\omega, \mu_\omega)$ be a measure space. Define:

$$\mathbf{X} \;=\; \bigsqcup_{\omega\in\Omega} \mathbf{X}_\omega; \qquad \text{and } \mathcal{X} \;=\; \left\{ \bigsqcup_{\omega\in\Omega} \mathbf{U}_\omega \;;\; \mathbf{U}_\omega \in \mathcal{X}_\omega, \quad \text{for all } \;\omega \in \Omega \right\}$$

and define $\mu : \mathcal{X} \longrightarrow [0,\infty]$ by: $\mu \left[\bigsqcup_{\omega\in\Omega} \mathbf{U}_\omega \right] \;=\; \sum_{\omega\in\Omega} \mu_\omega \left[\mathbf{U}_\omega\right].$

If Ω is countable, then "$\sum_{\omega\in\Omega} \mu_\omega \left[\mathbf{U}_\omega\right]$" is defined the way you expect; if If Ω is *uncountable*, then we define:

$$\sum_{\omega\in\Omega} \mu_\omega \left[\mathbf{U}_\omega\right] \;:=\; \sup_{\Upsilon\subset\Omega,\;\; \Upsilon \text{ countable}} \;\; \sum_{\upsilon\in\Upsilon} \mu_\upsilon \left[\mathbf{U}_\upsilon\right]$$

Then $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space (**Exercise 22**).

1.2(e) Finite vs. sigma-finite

Prerequisites: §1.2(d)

The **total mass** of measure space $(\mathbf{X}, \mathcal{X}, \mu)$ is just the value $\mu[\mathbf{X}]$. If $\mu[\mathbf{X}] < \infty$, then we say $(\mathbf{X}, \mathcal{X}, \mu)$ is **finite**; otherwise $(\mathbf{X}, \mathcal{X}, \mu)$ is **infinite**.

Example 19:

(a) $(\mathbb{I}, \mathcal{I}, \lambda)$ is finite.

(b) $\mathbb{R}$ with the Lebesgue measure is infinite.

(c) $\mathbb{Z}$ with the counting measure is infinite.

(d) If $\mathbf{G}$ is a topological group, then the Haar measure is finite if and only if $\mathbf{G}$ is compact. (**Exercise 23**)

$(\mathbf{X}, \mathcal{X}, \mu)$ is called **sigma-finite** if it is a countable disjoint union of finite measure spaces. Formally: $\mathbf{X} \;=\; \bigsqcup_{n=1}^{\infty} \mathbf{X}_n$, where $(\mathbf{X}_n, \mathcal{X}_n, \mu_n)$ are finite.

Example 20:

(a) $\mathbb{R}$ with the Lebesgue measure is sigma-finite.

(b) $\mathbb{Z}$ with the counting measure is sigma-finite.

(c) Let Ω be an uncountably infinite set, and for each $\omega \in \Omega$, let $\mathbb{R}_\omega$ be a copy of the real line equipped with Lebesgue measure. Then the measure space $\mathbf{X} = \bigsqcup_{\omega\in\Omega} \mathbb{R}_\omega$ is *not* sigma-finite.

Exercise 24 Verify these three examples.

Virtually all measure spaces dealt with in practice are sigma-finite. Finiteness in measure-theory is analogous to *compactness* for topological spaces; sigma-finiteness is analogous to *sigma-compactness*.

1.2(f) Sums and Limits of Measures

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

Proposition 21

1. *If μ_1 and μ_2 are measures on $(\mathbf{X}, \mathcal{X})$, and $c_1, c_2 > 0$, then $\mu = c_1 \cdot \mu_1 + c_2 \cdot \mu_2$ is also a measure.*
2. *Suppose that $\{\mu_n\}_{n=1}^{\infty}$ is a sequence of measures, and that the limit $\mu[\mathbf{U}] = \lim_{n\to\infty} \mu_n[\mathbf{U}]$ exists for all $\mathbf{U} \in \mathcal{X}$. Then μ is also a measure.*
3. *Suppose $\{c_n\}_{n=1}^{\infty}$ are nonnegative real constants, and that the limit $\mu[\mathbf{U}] = \sum_{n=1}^{\infty} c_n \cdot \mu_n[\mathbf{U}]$ exists for all $\mathbf{U} \in \mathcal{X}$. Then μ is also a measure.*

Proof: **Exercise 25** □

1.2(g) Product Measures

Suppose $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are measure spaces, and consider the measurable space is $(\mathbf{X}_1 \times \mathbf{X}_2,\ \mathcal{X}_1 \otimes \mathcal{X}_2)$. The **product** of measures μ_1 and μ_2 is the unique measure $\mu_1 \times \mu_2$ on $\mathbf{X}_1 \times \mathbf{X}_2$ so that:

$$(\mu_1 \times \mu_2)\,[\mathbf{U}_1 \times \mathbf{U}_2] \quad = \quad \mu_1[\mathbf{U}_1] \cdot \mu_2[\mathbf{U}_2]$$

for any $\mathbf{U}_1 \in \mathcal{X}_1$ and $\mathbf{U}_2 \in \mathcal{X}_2$. This determines $\mu_1 \times \mu_2$ on all disjoint unions of rectangles by:

$$(\mu_1 \times \mu_2)\left(\bigsqcup_{n=1}^{\infty} \mathbf{U}_1^n \times \mathbf{U}_2^n\right) \quad = \quad \sum_{n=1}^{\infty} \mu_1\left(\mathbf{U}_1^n\right) \cdot \mu_2\left(\mathbf{U}_2^n\right)$$

We will formally construct the product measure in §2.1(a).

Example 22:

(a) Suppose $\mathbf{X}$ and $\mathbf{Y}$ are finite sets, with probability measures μ and ν. Then $\mathbf{X} \times \mathbf{Y}$ is a finite set, and the the product measure $\mu \times \nu$ is simply defined:

$$\mu\times\nu\left\{(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_N, y_N)\right\} \quad = \quad \mu\{x_1\}\cdot\nu\{y_1\} + \mu\{x_2\}\cdot\nu\{y_2\} + \ldots + \mu\{x_N\}\cdot\nu\{y_N\}.$$

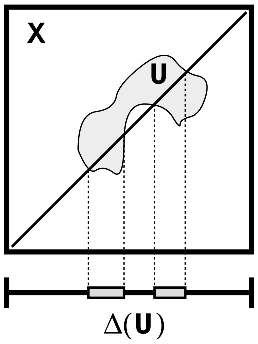


Figure 1.14: The diagonal measure

(b) If μ is the Lebesgue measure on $\mathbb{R}$, then $\mu\times\mu$ is the Lebesgue measure on $\mathbb{R}^2$ (**Exercise 26** Hint: It suffices to show that they agree on boxes).

(c) Suppose $\mathbf{G}_1$ and $\mathbf{G}_2$ are topological groups with left-Haar measures η_1 and η_2. If $\mathbf{G} = \mathbf{G}_1 \times \mathbf{G}_2$ is endowed with the product group structure and product topology, then $\eta = \eta_1 \times \eta_2$ is the left-Haar measure for $\mathbf{G}$ (**Exercise 27** Hint: It suffices to show that η is left-multiplication invariant). The same goes for right Haar measures, of course.

1.2(h) Diagonal Measures

Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, and consider the measurable space $(\mathbf{X}^2, \mathcal{X}^2) = (\mathbf{X} \times \mathbf{X},\ \mathcal{X} \otimes \mathcal{X})$. One possible measure on $(\mathbf{X}^2, \mathcal{X}^2)$ is the product measure $\mu^2 = \mu \times \mu$, but this is not the only candidate. Another valid measure is the **diagonal measure** μ_Δ, defined as follows. For any subset $\mathbf{U} \in \mathbf{X}^2$, define $\Delta(\mathbf{U}) = \{x \in \mathbf{X}\ ;\ (x,x) \in \mathbf{U}\}$, as shown in Figure 1.14. This represents the intersection of $\mathbf{U}$ with the 'diagonal set' of $\mathbf{X}^2$. Then define

$$\mu_\Delta(\mathbf{U}) \;=\; \mu\left[\Delta(\mathbf{U})\right]$$

Exercise 28 Verify that μ_Δ is a measure.

1.3 Mappings between Measure Spaces

1.3(a) Measurable Functions

Prerequisites: §1.1(b)

Definition 23 Measurable Function

Let $(\mathbf{X}_1, \mathcal{X}_1)$, and $(\mathbf{X}_2, \mathcal{X}_2)$, be measurable spaces. A function $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ is **measurable** *with respect to $\mathcal{X}_1$ and $\mathcal{X}_2$ if every $\mathcal{X}_2$-measurable set has an $\mathcal{X}_1$-measurable preimage under f. Formally:*

For all $\mathbf{A} \subset \mathbf{X}_2$, if $\mathbf{A} \in \mathcal{X}_2$, then $f^{-1}(\mathbf{A}) \in \mathcal{X}_1$

When unambiguous, we simply say 'f is **measurable**,' without explicit reference to $\mathcal{X}_1$ and $\mathcal{X}_2$.
Note that this definition depends only on the *sigma-algebras* $\mathcal{X}_1$ and $\mathcal{X}_2$, and has nothing to do with any actual *measures* on $\mathbf{X}_1$ or $\mathbf{X}_2$, *per se*. People often get confused and speak of a function being "measurable" with respect to some choice of measure; what they *mean* is that it is measurable with respect to the underlying sigma-algebras.

Example 24:

(a) **Borel Measurable functions on $\mathbb{R}$:** A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is **Borel measurable** if it is measurable with respect to the Borel sigma algebra; thus, if $\mathbf{B} \subset \mathbb{R}$ is a Borel set, then so is $f^{-1}(\mathbf{B})$. For example:

- Any continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is Borel-measurable.
- Any peicewise continuous function is Borel-measurable.
- If $\mathbb{1}_{\mathbb{Q}} : \mathbb{R} \longrightarrow \{0, 1\}$ is the characteristic function of the rationals, then $\mathbb{1}_{\mathbb{Q}}$ is Borel-measurable.
- f is Borel measurable if and only if $f^{-1}[a, b)$ is a Borel set for every $a, b \in \mathbb{R}$.
- In particular, any nonincreasing (or nondecreasing) function is Borel measurable.

Exercise 29 Verify these assertions.

(b) **Borel Measurable functions in General:**

Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are topological spaces; then a function $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ is **Borel measurable** if it is measurable with respect to the Borel sigma-algebras on $\mathbf{X}_1$ and $\mathbf{X}_2$. For example:

- Any continuous function is Borel-measurable.
- $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ is Borel measurable if and only if the preimage of every open subset of $\mathbf{X}_2$ is Borel-measurable in $\mathbf{X}_1$.

(**Exercise 30**)

(c) **Characteristic functions:** Let $\mathbf{U} \subset \mathbf{X}$, and let $\mathbb{1}_{\mathbf{U}} : \mathbf{X} \longrightarrow \{0, 1\}$ be it's **characteristic function:**

$$\mathbb{1}_{\mathbf{U}}(x) = \begin{cases} 1 & \text{if } x \in \mathbf{U} \\ 0 & \text{if } x \notin \mathbf{U} \end{cases}$$

Then $\mathbb{1}_{\mathbf{U}}$ is a measurable function if and only if $\mathbf{U}$ is a measurable subset of $\mathbf{X}$, where we suppose $\{0, 1\}$ has the power set sigma algebra $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. (**Exercise 31**)

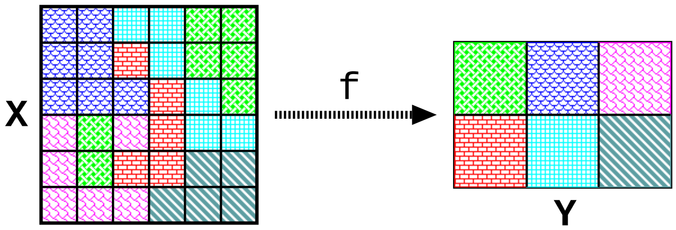


Figure 1.15: **Measurable mappings between partitions:** Here, $\mathbf{X}$ and $\mathbf{Y}$ are partitioned into grids of small squares. The function f is measurable if each of the six squares covering $\mathbf{Y}$ gets pulled back to a union of squares in $\mathbf{X}$.

(d) **Maps between partitions:** Let $\mathbf{X}, \mathbf{Y}$ be sets with partitions $\mathcal{P}$ and $\mathcal{Q}$, respectively, and let $\mathcal{X} = \sigma(\mathcal{Q})$ and $\mathcal{Y} = \sigma(\mathcal{Q})$. As shown in Figure 1.15, the function $f : \mathbf{X} \longrightarrow \mathbf{Y}$ is measurable with respect to $\mathcal{X}$ and $\mathcal{Y}$ if and only if, for every $\mathbf{Q} \in \mathcal{Q}$, there are $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_K \in \mathcal{P}$ so that $f^{-1}(\mathbf{Q}) = \bigsqcup_{k=1}^{K} \mathbf{P}_k$ (**Exercise 32**).

Proposition 25 (Closure Properties of Measurable Functions)

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

1. *Let $\mathbb{R}^N$ have the Borel sigma algebra. Suppose that $f, g : \mathbf{X} \longrightarrow \mathbb{R}$ are measurable. Then the functions $f + g$, $f \cdot g$, f^g, $\min\{f, g\}$, and $\max\{f, g\}$ are all measurable.*

2. *If $(\mathbf{G}, \mathcal{G})$ a topological group with Borel sigma algebra, and $f, g : \mathbf{X} \longrightarrow \mathbf{G}$ are measurable, then so is $f \cdot g$.*

3. *Let $(\mathbf{T}, \mathcal{T})$ be a topological space with Borel sigma-algebra, and let $f_n : \mathbf{X} \longrightarrow \mathbf{T}$ be a measurable function for all $n \in \mathbb{N}$. Suppose $f(x) = \lim_{n\to\infty} f_n(x)$ exists for all $x \in \mathbf{X}$ (with the limit taken in the $\mathbf{T}$-topology). Then $f : \mathbf{X} \longrightarrow \mathbf{T}$ is also measurable.*

4. *Let $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be a measurable functions for all $n \in \mathbb{N}$. Then the functions*

$$\begin{aligned} \overline{f}(x) &= \sup_{n\in\mathbb{N}} f_n(x) & \underline{f}(x) &= \inf_{n\in\mathbb{N}} f_n(x) \\ \overline{f}_\infty(x) &= \limsup_{n\to\infty} f_n(x) \quad \textit{and} & \underline{f}_\infty(x) &= \liminf_{n\to\infty} f_n(x) \end{aligned}$$

are all measurable. Also, if $F(x) = \sum_{n=1}^{\infty} f_n(x)$ exists for all $x \in \mathbf{X}$, then it defines a measurable function.

5. *If $(\mathbf{Y}, \mathcal{Y})$ and $(\mathbf{Z}, \mathcal{Z})$ are also measurable spaces, and $f : \mathbf{Y} \longrightarrow \mathbf{Z}$ and $g : \mathbf{X} \longrightarrow \mathbf{Y}$ are measurable, then $f \circ g : \mathbf{X} \longrightarrow \mathbf{Z}$ is also measurable.*

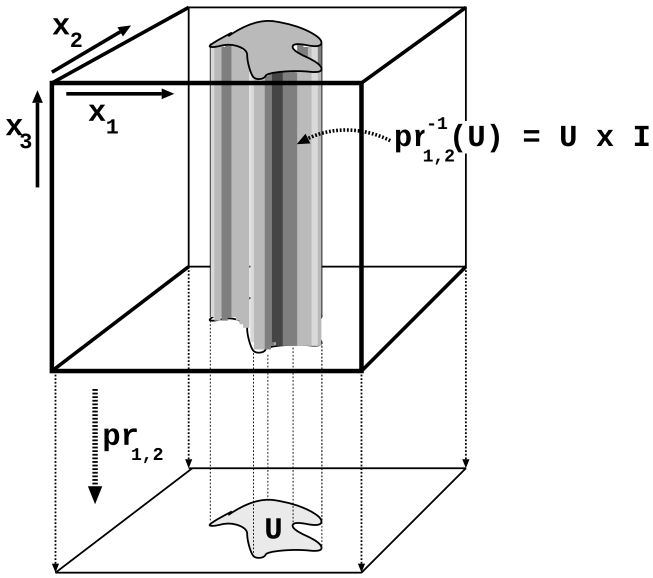


Figure 1.16: Elements of $\mathbf{pr}_{1,2}{}^{-1}(\mathcal{I}^2)$ look like"vertical fibres".

Proof: **Exercise 33** $\square$

Definition 26 Pulled-back Sigma-algebra

Suppose $(\mathbf{X}, \mathcal{X})$ *is a measurable space, and* $f : \mathbf{Y} \longrightarrow \mathbf{X}$ *is some function. Then we can use* f *to* **pull back** *the sigma-algebra on* $\mathbf{X}$ *to a sigma-algebra on* $\mathbf{Y}$*, defined:*

$$f^{-1}(\mathcal{X}) \quad := \quad \left\{ f^{-1}(\mathbf{U})\ ;\ \mathbf{U} \in \mathcal{X} \right\}$$

Exercise 34 Check that $f^{-1}(\mathcal{X})$ is a sigma-algebra

Lemma 27 $f : (\mathbf{X}, \mathcal{X}) \longrightarrow (\mathbf{Y}, \mathcal{Y})$ *is* **measurable** *if and only if* $f^{-1}(\mathcal{Y}) \subset \mathcal{X}$

Proof: **Exercise 35** $\square$

Example 28:

Consider the **unit cube**, $\mathbb{I}^3 := \mathbb{I} \times \mathbb{I} \times \mathbb{I}$, where $\mathbb{I} := [0, 1]$ is the unit interval. Let $\mathbb{I}^2$ be the unit square, and consider the projection onto the first two coordinates, $\mathbf{pr}_{1,2} : \mathbb{I}^3 \longrightarrow \mathbb{I}^2$; if $\mathbf{x} := (x_1, x_2, x_3) \in \mathbb{I}^3$, then $\mathbf{pr}_{1,2}(\mathbf{x}) = (x_1, x_2) \in \mathbb{I}^2$.

Consider the *pulled back* sigma algebra $\mathbf{pr}_{1,2}{}^{-1}(\mathcal{I}^2)$, (where $\mathcal{I}^2$ is the Borel sigma-algebra on $\mathbb{I}^2$). Elements of $\mathbf{pr}_{1,2}{}^{-1}(\mathcal{I}^2)$ look like "vertical fibres" in the cube (Figure 5.6). That is:

$$\mathbf{pr}_{1,2}{}^{-1}(\mathcal{I}^2) \quad = \quad \{\mathbf{pr}_1{}^{-1}(\mathbf{U})\ ;\ \mathbf{U} \in \mathcal{I}^2\} \quad = \quad \{\mathbf{U} \times \mathbb{I}\ ;\ \mathbf{U} \in \mathcal{I}^2\} \quad = \quad \mathcal{I}^2 \otimes \{\mathbb{I}\}.$$

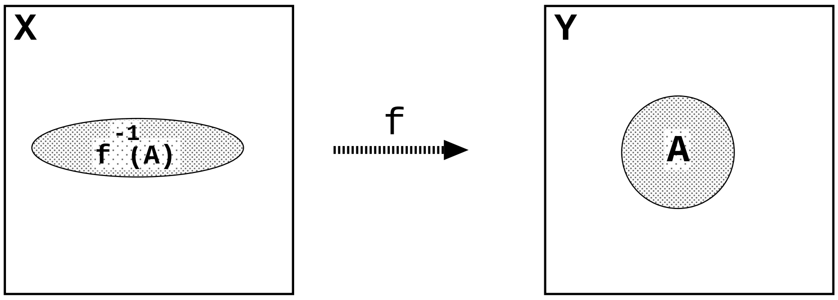


Figure 1.17: Measure-preserving mappings

Definition 29 Generating Sigma-algebras with Functions

Let $(\mathbf{X}_1, \mathcal{X}_1)$, $(\mathbf{X}_2, \mathcal{X}_2)$, $\ldots$, $(\mathbf{X}_N, \mathcal{X}_N)$ *be measurable space, and* $\mathbf{Y}$ *some set. For each* $n \in [1..N]$, *let* $f_n : \mathbf{Y} \longrightarrow \mathbf{X}_n$ *be some function.*

The **sigma algebra generated** *by* $f_1, \ldots, f_N$ *is the smallest sigma-algebra on* $\mathbf{Y}$ *so that* $f_1, \ldots, f_N$ *are all measurable with respect to the sigma-algebras of their respective target spaces.*

This sigma-algebra is denoted by "$\sigma(f_1, \ldots, f_N)$". Suppose that $\mathcal{Y}_1 = f_1^{-1}(\mathcal{X}_1)$, $\mathcal{Y}_2 = f_2^{-1}(\mathcal{X}_2), \ldots, \mathcal{Y}_N = f_N^{-1}(\mathcal{X}_N)$. Then (**Exercise 36**)

$$\sigma(f_1, \ldots, f_N) \quad = \quad \sigma\Big(\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \ldots \cup \mathcal{Y}_N\Big).$$

Remark 30:

- For a single function $f : \mathbf{Y} \longrightarrow \mathbf{X}$, it is clear that $\sigma(f) \quad = \quad f^{-1}(\mathcal{X})$ (**Exercise 37**).
- Given N functions $f_1, \ldots, f_N$, define $F : \mathbf{Y} \longrightarrow \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_N$ by:

$$F(y) = \Big(f_1(y),\ f_2(y),\ \ldots,\ f_N(y)\Big)$$

  Then $\sigma(f_1, \ldots, f_N) \quad = \quad F^{-1}(\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \ldots \mathcal{X}_N)$. (**Exercise 38**).

1.3(b) Measure-Preserving Functions

Prerequisites: §1.3(a)

Definition 31 Measure-Preserving Function

Let $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ *and* $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ *be measure spaces, and* $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ *some measurable function. Then* f *is* **measure-preserving** *if, for every* $\mathbf{A} \in \mathcal{X}_2$, $\mu_1[f^{-1}(\mathbf{A})] \quad = \quad \mu_2[\mathbf{A}]$. *See Figure 1.17.*

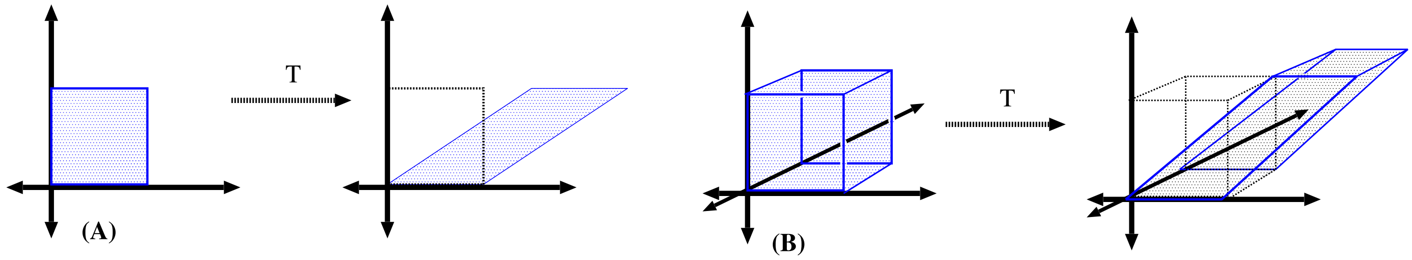


Figure 1.18: **(A)** A special linear transformation on $\mathbb{R}^2$ maps a rectangle to parallelogram with the same area. **(B)** A special linear transformation on $\mathbb{R}^3$ maps a box to parallelopiped with the same volume.

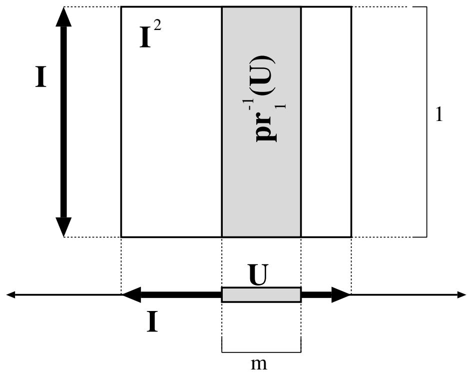


Figure 1.19: The projection from $\mathbb{I}^2$ to $\mathbb{I}$ is measure-preserving.

Example 32: $\mathbf{SL}(\mathbb{R}^n)$

Consider the group of special linear transformations of $\mathbb{R}^n$:

$$\mathbb{SL}\left[\mathbb{R}^n\right] \quad := \quad \{T : \mathbb{R}^n \longrightarrow \mathbb{R}^n \ ; \ T \ \text{ is linear, and } \ \det(T) = 1\}$$

As illustrated in Figure 1.18, an element $T \in \mathbb{SL}\left[\mathbb{R}^n\right]$ transforms cubes into parallellipipeds having the same volume. Thus, T transforms $\mathbb{R}^n$ in a manner which preserves the Lebesgue measure.

(Actually we don't need $\det(T) = 1$, but only that $|\det(T)| = 1$. Thus, for example, a map which "flips" the space is also measure-preserving.)

Example 33: Projection Maps

Let $\mathbb{I} = [0, 1]$ be the unit interval, and $\mathbb{I}^2 = [0, 1] \times [0, 1]$ be the unit square. Let $\mathbf{pr}_1 : \mathbb{I}^2 \longrightarrow \mathbb{I}$ to be the **projection map** of $\mathbb{I}^2$ onto the first coordinate; then $\mathbf{pr}_1$ is measure-preserving. To see this, consider Figure 1.19. Suppose $\mathbf{U} \subset \mathbb{I}$ has a Lebesgue measure (ie. a *length*) of m. Then its preimage is

$$\mathbf{pr}_1{}^{-1}[\mathbf{U}] \quad = \quad \mathbf{U} \times \mathbb{I} \quad \subset \quad \mathbb{I}^2,$$

and $\mathbf{U} \times \mathbb{I}$ has a Lebesgue measure (ie. an *area*) of $m \times 1 = m$.

Definition 34 Pushed Forward Measure

Let $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ *be a measure space, and* $(\mathbf{X}_2, \mathcal{X}_2)$ *a* measurable *space, and suppose that* $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ *is a measurable function. We use* f *to* **push forward** *the measure* μ_1 *to a measure,* $f^*\mu_1$*, defined on* $\mathbf{X}_2$ *as follows: For any subset* $A \in \mathcal{X}_2$,

$$f^*\mu[A] \quad := \quad \mu\left[f^{-1}(A)\right].$$

Exercise 39 *Prove that* $f^*\mu$ *is a measure.*

Note that, although we *pull back* sigma-algebras, we must *push forward* measures. We cannot, in general, "pull back" a measure; nor can we generally "push forward" a sigma-algebra.

Lemma 35 *Let* $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ *and* $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ *be measure spaces, and let* $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ *be some measurable function. Then* f *is measure-preserving if and only if* $f^*\mu_1 = \mu_2$.

Proof: **Exercise 40** □

1.3(c) 'Almost everywhere'

Prerequisites: §1.2(b),§1.3(a)

If a certain assertion is true (or a certain construction is well-defined) on the *complement* of a set of measure zero, we say that it is true (resp. well-defined) **almost everywhere**. Intuitively, this means there is a 'bad' set where the assertion/construction fails, but this bad set *doesn't matter*, because it is of measure zero. In the land of measure spaces, one can get away with working with such *almost everywhere* constructs and assertions. Indeed, it is often *not* possible, or even desirable, to define things everywhere on a space.

Synonymous with 'almost everywhere' are the terms **mod zero** and **essential**. In probability literature, the equivalent term is **almost surely**. Usually, 'almost everywhere' is abbreviated as **a.e.** (or **a.e.**$[\mu]$ if one wishes to specify the measure). Similarly, 'almost surely' is abbreviated as **a.s.** (or **a.s.**$[\mu]$).

Definition 36 Function mod Zero

Let $(\mathbf{X}, \mathcal{X}, \mu)$ *be a measure space, and let* $\mathbf{Y}$ *be some other set. A function* **mod zero** *is a function* f *defined* **almost everywhere** *on* $\mathbf{X}$. *In other words, there is some measurable subset* $\mathbf{X}_0 \subset \mathbf{X}$*, with* $\mu[\mathbf{X} \setminus \mathbf{X}_0] = 0$*, so that* $f : \mathbf{X}_0 \longrightarrow \mathbf{Y}$.

We often simply say that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ *is a 'function defined mod zero', or a 'function defined* **a.e.***'.*

Example: Let $f : \mathbb{Q} \longrightarrow \mathbb{Q}$ be defined $f(x) = x$. Then f is a mod zero function $f : \mathbb{R} \longrightarrow \mathbb{Q}$.

A function mod zero is can be 'completed' to a function defined 'everywhere' on $\mathbf{X}$ in an entirely arbitrary manner, without affecting its measure-theoretic properties at all. To be precise, if $\mathbf{X}_1$ is a *complete* measure space, then:

- Any given completion of f will be measurable if and only if *all* completions are measurable.
- Any given completion will be measure-preserving if and only if *all* completions are measure-preservng

  Exercise 41 Verify these assertions.

Lemma 37

1. *If $f, g : (\mathbf{X}, \mathcal{X}, \mu) \longrightarrow \mathbb{C}$ are defined* **a.e.**, *then the functions $f + g$, $f \cdot g$, and f^g, are also defined* **a.e.**.
2. *If $f_1 : (\mathbf{X}_1, \mathcal{X}_1, \mu_1) \longrightarrow \mathbf{Y}_1$ and $f_2 : (\mathbf{X}_2, \mathcal{X}_2, \mu_2) \longrightarrow \mathbf{Y}_2$ are defined* **a.e.**, *then $f_1 \times f_2 : (\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{X}_1 \otimes \mathcal{X}_2, \mu_1 \times \mu_2) \longrightarrow \mathbf{Y}_1 \times \mathbf{Y}_2$ is also defined* **a.e.**.
3. *If $f_1 : (\mathbf{X}_1, \mathcal{X}_1, \mu_1) \longrightarrow (\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ and $f_2 : (\mathbf{X}_2, \mathcal{X}_2, \mu_2) \longrightarrow \mathbf{Y}$ are defined* **a.e.**, *then $f_2 \circ f_1 : (\mathbf{X}_1, \mathcal{X}_1, \mu_1) \longrightarrow \mathbf{Y}$ is defined* **a.e.**.

Proof: **Exercise 42** □

Definition 38 Essential Equality

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathbf{Y}$ be any set. Let $f, g : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined **a.e.**.

f and g are **essentially equal** *(or* **equal almost everywhere***) if there is some $\mathbf{X}_0 \subset \mathbf{X}$ so that $\mu[\mathbf{X} \setminus \mathbf{X}_0] = 0$, and so that, for all $x \in \mathbf{X}_0$, $f(x_0) = g(x_0)$.*

Sometimes we will write this: "$f \;\; =_\mu \;\; g$."

Lemma 39 *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a complete measure space, and $(\mathbf{Y}, \mathcal{Y})$ a measurable space. Let $f, g : \mathbf{X} \longrightarrow \mathbf{Y}$ be arbitrary functions.*

1. *If f is measurable, and $g =_\mu f$, then g is also measurable.*
2. *In particular, if $\mathsf{supp}[f]$ is a set of measure zero (ie. $f =_\mu 0$), then f is measurable.*

Proof: **Exercise 43** Hint: Prove (2) first. Then (1) follows from Part 1 of Proposition 25 on page 25 □

Exercise 44 Find a counterexample to Lemma 39 when $(\mathbf{X}, \mathcal{X}, \mu)$ is not complete.

The next result says that completing a sigma-algebra does not add any 'essentially' new measurable functions. For this reason, we generally can assume that all measure spaces are complete (or have been completed) without compromising the generality of our results.

Lemma 40 *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\widetilde{\mathcal{X}}$ be the completion of $\mathcal{X}$. Let $(\mathbf{Y}, \mathcal{Y})$ be a measurable space, and let $\widetilde{f} : \mathbf{X} \longrightarrow \mathbf{Y}$ be $\widetilde{\mathcal{X}}$-measurable. Then there exists a function $f : \mathbf{X} \longrightarrow \mathbf{Y}$ such that* **(1)** *f is $\mathcal{X}$-measurable, and* **(2)** *$f =_\mu \widetilde{f}$.*

Proof: **Exercise 45** ______ □

Definition 41 Almost Everywhere Convergence

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathbf{Y}$ be a topological space. Let $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined **a.e.**, *for all $n \in \mathbb{N}$, and let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ also be defined* **a.e.**.

The sequence $\{f_n\}_{n=1}^{\infty}$ converges to f **almost everywhere** *if there is $\mathbf{X}_0 \subset \mathbf{X}$ so that $\mu[\mathbf{X} \setminus \mathbf{X}_0] = 0$, and so that, for all $x \in \mathbf{X}_0$, $\lim_{n\to\infty} f_n(x) = f(x)$.*

We write this: "$\lim_{n\to\infty} f_n = f$, **a.e.**"

Lemma 42 *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a complete measure space, and $\mathbf{Y}$ a topological space. Suppose that $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ are measurable functions for all $n \in \mathbb{N}$, and that $\lim_{n\to\infty} f_n = f$,* **a.e.**. *Then f is also measurable.*

Proof: **Exercise 46** ______ □

Exercise 47 Find a counterexample to Lemma 42 when $(\mathbf{X}, \mathcal{X}, \mu)$ is not complete.

Definition 43 Essentially Injective

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathbf{Y}$ be any set. Let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined **a.e.**.

f is **essentially injective** *if f is injective everywhere except on a subset of $\mathbf{X}$ of measure zero. In other words, there is some $\mathbf{X}_0 \subset \mathbf{X}$, so that $\mu[\mathbf{X} \setminus \mathbf{X}_0] = 0$, and so that the restriction $f_{|\mathbf{X}_0} : \mathbf{X}_0 \longrightarrow \widehat{X}$ is injective.*

Sometimes we say that f is **injective a.e.**.

Example 44:

(a) The map $f : [0,1] \longrightarrow \mathbb{C}$ defined $f(x) = e^{2\pi \mathbf{i} x}$ is essentially injective, since it is injective everywhere except at the points 0 and 1, which both map to the point $1 \in \mathbb{C}$.

(b) If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined: $f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \\ x; & \text{if } x \in \not{\mathbb{Q}}. \end{cases}$, then f is essentially injective. ___

Definition 45 Essentially Surjective

Let $(\mathbf{Y}, \mathcal{Y}, \nu)$ be a measure space, and let $\mathbf{X}$ be any set. Let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ be some function.
f is **essentially surjective** *if f is surjective onto a subset of $\mathbf{Y}$ whose complement has measure zero. In other words, the set $\mathbf{Y} \setminus \mathsf{image}[f] \subset \mathbf{Y}$ has measure zero.*

Example 46:

(a) The embedding $(0,1) \hookrightarrow [0,1]$ is essentially surjective, since $[0,1] \setminus \mathsf{image}[f] = \{0,1\}$ has measure zero.

(b) The embedding $\mathbb{Q} \hookrightarrow \mathbb{R}$ is essentially surjective.

(c) If $f : (\mathbf{X}, \mathcal{X}, \mu) \longrightarrow (\mathbf{Y}, \mathcal{Y}, \nu)$ is measure-preserving, it is essentially surjective (**Exercise 48**).

Definition 47 Essential Inverse

Let $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ be measure spaces, and let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined **a.e.**$[\mu]$. *An* **essential inverse** *for f is a function, $g : \mathbf{Y} \longrightarrow \mathbf{X}$, defined* **a.e.**$[\nu]$, *so that $f \circ g \ =_{\mu}\ \mathbf{Id}_{\mathbf{Y}}$ and $g \circ f \ =_{\mu}\ \mathbf{Id}_{\mathbf{X}}$.*
We then say f is **essentially invertible**.

Example: Define $g : [0,1] \longrightarrow (0,1)$ by: $f(x) = \begin{cases} x & \text{if } 0 < x < y \\ \frac{1}{2} & \text{if } x = 0 \text{ or } x = 1 \end{cases}$. Then g is an essential inverse of the embedding $(0,1) \hookrightarrow [0,1]$.

Lemma 48 *Let $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ be measure spaces, and let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined* **a.e.**.
f is essentially injective and essentially surjective if and only if f is essentially invertible.

Proof: **Exercise 49** □

Definition 49 Isomorphism mod Zero

Let $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ be measure spaces, and let $f : \mathbf{X} \longrightarrow \mathbf{Y}$ be defined mod zero.
f is an **isomorphism mod zero** *if:*

1. *f is measurable with respect to $\mathcal{X}$ and $\mathcal{Y}$.*
2. *f is measure-preserving.*
3. *f is essentially invertible, and the inverse function f^{-1} is also measurable and measure-preserving.*

1.3(d) A categorical approach

Prerequisites: §1.3(b), §1.3(c)

Recall that Example (46c) says that all measure-preserving maps must be surjective; this is to some extent an artifact of Definition 31 on page 27. We can weaken this definition as follows...

Definition 50 Weakly Measure-Preserving

Let $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ be measure spaces, and $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ some measurable function. Let $\mathbf{Y} = \mathsf{image}[f] \subset \mathbf{X}_2$. Then f is **weakly measure-preserving** *if it is measure-preserving as a function $f : \mathbf{X}_1 \longrightarrow \mathbf{Y}$. In other words, for every $\mathbf{A} \in \mathcal{X}_2$, $\mu_1[f^{-1}(\mathbf{A})] = \mu_2[\mathbf{A} \cap \mathbf{Y}]$.*

We now define the category, $\mathcal{M}eas$, of measure spaces and weakly measure-preserving functions mod zero, as follows:

- The objects of $\mathcal{M}eas$ are measure spaces.
- The morphisms of $\mathcal{M}eas$ are weakly measure-preserving functions, defined **a.e.**.

As mentioned above, the composition of two such morphisms is well-defined. Also,

1. The *monics* of $\mathcal{M}eas$ are the essentially injective maps (ie. *embeddings* of measure spaces).
2. The *epics* of $\mathcal{M}eas$ are the essentially surjective maps (ie. (strongly) measure-preserving maps).
3. The *isomorphisms* of $\mathcal{M}eas$ are essential isomorphisms.
4. The *product* of two measure spaces $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ is $(\mathbf{X} \times \mathbf{Y},\ \mathcal{X} \otimes \mathcal{Y},\ \mu \times \nu)$.
5. The *coproduct* of two measure spaces $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ is $(\mathbf{X} \sqcup \mathbf{Y},\ \sigma(\mathcal{X} \sqcup \mathcal{Y}),\ \eta)$, where $\eta(\mathbf{U}) = \mu(\mathbf{U} \cap \mathbf{X}) + \nu(\mathbf{U} \cap \mathbf{Y})$.

Exercise 50 Verify these statements.

2 Construction of Measures

2.1 Outer Measures

Prerequisites: §1.1(c)

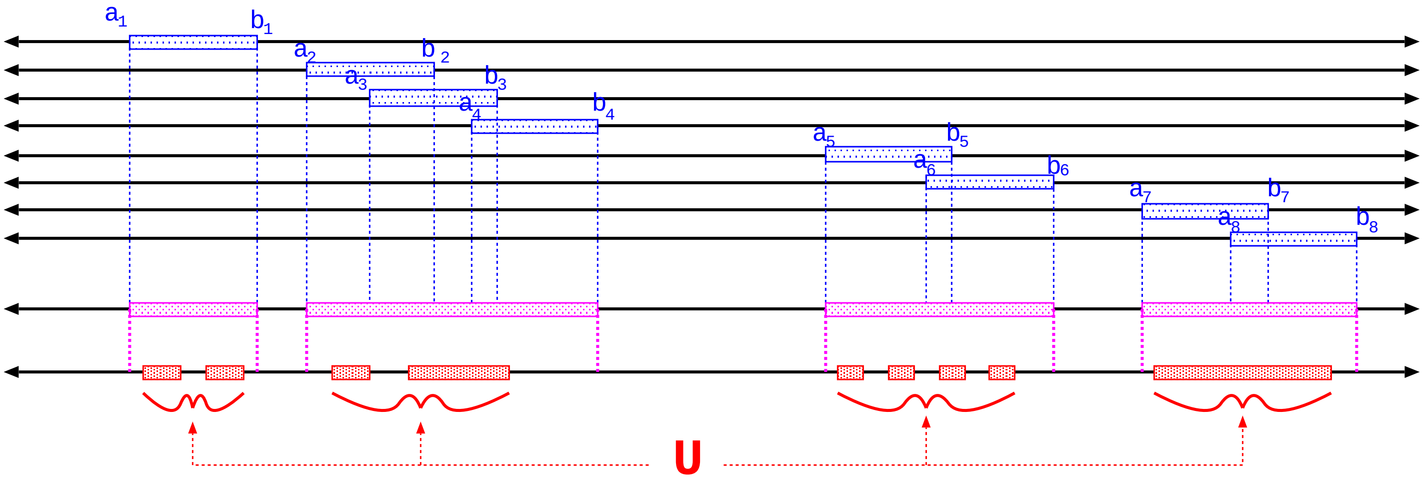


Figure 2.1: The intervals $(a_1, b_1]$, $(a_2, b_2], \ldots, (a_8, b_8]$ form a covering of $\mathbf{U}$.

The theory of outer measures is rather abstract, so concrete examples are essential. However, on a first reading, the multitude of examples may be confusing. I suggest that the reader initially concentrate on the Lebesgue measure (developed in Examples 51, (54a), (56a), (59a), (61a), (62a), and (65a)), and only skim the other examples. Then go back and study the other examples in detail.

Let $\mathbf{X}$ be a set, with power set $\mathcal{P}(\mathbf{X})$. An **outer measure** is a function $\widetilde{\mu} : \mathcal{P}(\mathbf{X}) \longrightarrow [0, \infty]$ having the following three properties:

(OM1) $\widetilde{\mu}(\emptyset) \;=\; 0.$

(OM2) If $\mathbf{U} \subset \mathbf{V}$, then $\widetilde{\mu}(\mathbf{U}) \leq \widetilde{\mu}(\mathbf{V})$

(OM3) $\widetilde{\mu}\left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right) \;\leq\; \sum_{n=1}^{\infty} \widetilde{\mu}(\mathbf{A}_n)$, for any subsets $\mathbf{A}_n \subset \mathbf{X}$.

Example 51: Lebesgue Outer Measure

Let $\mathbf{X} = \mathbb{R}$. If $\mathbf{U} \subset \mathbb{R}$, then a **covering** of $\mathbf{U}$ is a collection of *left-open* intervals $\left\{(a_n, b_n]\right\}_{n=1}^{\infty}$, (with $-\infty \leq a_n < b_n \leq \infty$ for all $n \in \mathbb{N}$) such that $\mathbf{U} \subset \bigcup_{n=1}^{\infty} (a_n, b_n]$ (see Figure 2.1). The **Lebesgue Outer Measure** is defined:

$$\widetilde{\mu}(\mathbf{U}) \;=\; \inf_{\substack{\left\{(a_n, b_n]\right\}_{n=1}^{\infty} \\ \text{a covering of } \mathbf{U}}} \sum_{n=1}^{\infty} (b_n - a_n)$$

We will prove that $\widetilde{\mu}$ satisfies properties **(OM1)**, **(OM2)**, and **(OM3)** in §2.1(a), and also construct other examples of outer measures there; the reader is encouraged to glance at §2.1(a) to get some intuition about outer measures before continuing this section.

For $\widetilde{\mu}$ to be a measure, we want **(OM3)** to be an *equality* in the case of a disjoint union:

$$\widetilde{\mu}\left(\bigsqcup_{n=1}^{\infty} \mathbf{U}_n\right) \quad = \quad \sum_{n=1}^{\infty} \widetilde{\mu}\left(\mathbf{U}_n\right), \tag{2.1}$$

But (2.1) fails in general. Indeed, there may even be a *pair* of disjoint sets $\mathbf{U}$ and $\mathbf{V}$ such that

$$\widetilde{\mu}\left(\mathbf{U} \sqcup \mathbf{V}\right) \quad < \quad \widetilde{\mu}(\mathbf{U}) + \widetilde{\mu}(\mathbf{V}) \tag{2.2}$$

Intuitively, this is because $\mathbf{U}$ and $\mathbf{V}$ have 'fuzzy boundaries', which overlap in some weird way. If $\mathbf{Y} = \mathbf{U} \sqcup \mathbf{V}$, then of course $\mathbf{U} = \mathbf{U} \cap \mathbf{Y}$ and $\mathbf{V} = \mathbf{U}^{\complement} \cap \mathbf{Y}$. Thus, we can rewrite (2.2) as:

$$\widetilde{\mu}\left[\mathbf{Y}\right] \quad < \quad \widetilde{\mu}\left[\mathbf{U} \cap \mathbf{Y}\right] \;+\; \widetilde{\mu}\left[\mathbf{U}^{\complement} \cap \mathbf{Y}\right] \tag{2.3}$$

To obtain the additivity property (2.1), we must exclude sets with 'fuzzy boundaries'; hence we should exclude any subset $\mathbf{U}$ which causes (2.3). Indeed, we will require something even stronger. We say a subset $\mathbf{U} \subset \mathbf{X}$ is **$\widetilde{\mu}$-measurable** if it satisfies:

$$\text{For any } \mathbf{Y} \subset \mathbf{X}, \quad \widetilde{\mu}\left[\mathbf{Y}\right] \quad = \quad \widetilde{\mu}\left[\mathbf{U} \cap \mathbf{Y}\right] \;+\; \widetilde{\mu}\left[\mathbf{U}^{\complement} \cap \mathbf{Y}\right] \tag{2.4}$$

Remark: Suppose $\widetilde{\mu}[\mathbf{X}]$ is *finite*; Then another way to to think of this is as follows: for any $\mathbf{U} \subset \mathbf{X}$, we define the **inner measure** of $\mathbf{U}$:

$$\underline{\mu}[\mathbf{U}] \;=\; \widetilde{\mu}[\mathbf{X}] - \widetilde{\mu}\left[\mathbf{U}^{\complement}\right]$$

we then say that $\mathbf{U}$ is **measurable** if $\underline{\mu}[\mathbf{U}] = \widetilde{\mu}[\mathbf{U}]$. Observe that equation (2.4) implies this when $\mathbf{Y} = \mathbf{X}$.

If $\widetilde{\mu}[\mathbf{X}]$ is *infinite*, this obviously doesn't work. Instead, we 'approximate' $\mathbf{X}$ with an increasing sequence of finite subsets. Let $\mathcal{F} = \{\mathbf{F} \subset \mathbf{X}\ ;\ \widetilde{\mu}[\mathbf{F}] < \infty\}$. For any $\mathbf{U} \subset \mathbf{F}$, say that $\mathbf{U}$ is **measurable within F** if

$$\widetilde{\mu}[\mathbf{U}] \;=\; \underline{\mu}_{\mathbf{F}}[\mathbf{U}] \;:=\; \widetilde{\mu}[\mathbf{F}] - \widetilde{\mu}\left[\mathbf{F} \setminus \mathbf{U}\right]$$

If $\mathbf{U} \subset \mathbf{X}$, then say is **measurable** if $\mathbf{U} \cap \mathbf{F}$ is measurable inside $\mathbf{F}$ for all $\mathbf{F} \in \mathcal{F}$.

Exercise 51 Verify that this is equivalent to equation (2.4). Hint: Let $\mathbf{F}$ play the role of $\mathbf{Y}$.

The sets satisfying (2.4) have 'crisp boundaries', and the outer measure $\widetilde{\mu}$ satisfies (2.1) on them. To be precise, we have:

Theorem 52 (Carathéodory)

Let $\widetilde{\mu}$ be an outer measure on $\mathbf{X}$, and let $\mathcal{X}$ be the set of $\widetilde{\mu}$-measurable subsets of $\mathbf{X}$. Then:

1. *$\mathcal{X}$ is a sigma-algebra.*
2. *$\widetilde{\mu} : \mathcal{X} \longrightarrow [0, \infty]$ is a complete measure.*

Proof: First we will simplify the condition for membership in $\mathcal{X}$:

Claim 1: *Let $\mathbf{U} \subset \mathbf{X}$. Then $\mathbf{U} \in \mathcal{X}$ if and only if:*
For any $\mathbf{Y} \subset \mathbf{X}$, with $\widetilde{\mu}[\mathbf{Y}] < \infty$, $\widetilde{\mu}[\mathbf{Y}] \;\geq\; \widetilde{\mu}[\mathbf{U} \cap \mathbf{Y}] + \widetilde{\mu}\left[\mathbf{U}^{\complement} \cap \mathbf{Y}\right]$.

Proof: Since $\mathbf{U} = (\mathbf{U} \cap \mathbf{Y}) \cup \left(\mathbf{U}^{\complement} \cap \mathbf{Y}\right)$, we have automatically

$$\widetilde{\mu}[\mathbf{Y}] \;\leq\; \widetilde{\mu}[\mathbf{U} \cap \mathbf{Y}] + \widetilde{\mu}\left[\mathbf{U}^{\complement} \cap \mathbf{Y}\right]$$

by property **(OM3)** of outer measures. Thus, we need only verify the reverse inequality in (2.4), which is automatically true whenever $\widetilde{\mu}[\mathbf{Y}] = \infty$. Thus, to get $\mathbf{U} \in \mathcal{X}$, we need only test the case when $\mu[\mathbf{Y}] < \infty$ .... □ [Claim 1]

Claim 2: *$\mathcal{X}$ is closed under complementation: if $\mathbf{U} \in \mathcal{X}$, then $\mathbf{U}^{\complement} \in \mathcal{X}$.*

Proof: Observe that the condition (2.4) is symmetric in $\mathbf{U}$ and $\mathbf{U}^{\complement}$: it is true for $\mathbf{U}$ if and only if it is true for $\mathbf{U}^{\complement}$. .... □ [Claim 2]

Claim 3: *$\mathcal{X}$ is closed under finite unions: if $\mathbf{A}, \mathbf{B} \in \mathcal{X}$, then $\mathbf{A} \cup \mathbf{B} \in \mathcal{X}$ also.*

Proof: Observe that

$$\mathbf{A} \cup \mathbf{B} \;=\; (\mathbf{A} \cap \mathbf{B}) \sqcup \left(\mathbf{A} \cap \mathbf{B}^{\complement}\right) \sqcup \left(\mathbf{A}^{\complement} \cap \mathbf{B}\right) \qquad \text{and} \quad (\mathbf{A} \cup \mathbf{B})^{\complement} \;=\; \mathbf{A}^{\complement} \cap \mathbf{B}^{\complement}.$$

Let $\mathbf{Y} \subset \mathbf{X}$ be arbitrary. It follows that

$$\mathbf{Y} \cap (\mathbf{A} \cup \mathbf{B}) \;=\; (\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}) \sqcup \left(\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}^{\complement}\right) \sqcup \left(\mathbf{Y} \cap \mathbf{A}^{\complement} \cap \mathbf{B}\right) \tag{2.5}$$

$$\text{and} \quad \mathbf{Y} \cap (\mathbf{A} \cup \mathbf{B})^{\complement} \;=\; \mathbf{Y} \cap \mathbf{A}^{\complement} \cap \mathbf{B}^{\complement}. \tag{2.6}$$

$$\begin{aligned}
\text{Thus,} \quad & \widetilde{\mu}\left[\mathbf{Y} \cap (\mathbf{A} \cup \mathbf{B})\right] + \widetilde{\mu}\left[\mathbf{Y} \cap (\mathbf{A} \cup \mathbf{B})^{\complement}\right] \\
=_{(a)} \; & \widetilde{\mu}\left[(\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}) \sqcup \left(\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}^{\complement}\right) \sqcup \left(\mathbf{Y} \cap \mathbf{A}^{\complement} \cap \mathbf{B}\right)\right] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A}^{\complement} \cap \mathbf{B}^{\complement}\right] \\
\leq_{(b)} \; & \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}\right] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A} \cap \mathbf{B}^{\complement}\right] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A}^{\complement} \cap \mathbf{B}\right] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A}^{\complement} \cap bB^{\complement}\right] \\
=_{(c)} \; & \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A}\right] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{A}^{\complement}\right] \quad =_{(d)} \quad \widetilde{\mu}[\mathbf{Y}].
\end{aligned}$$

(a) by (2.5) and (2.6); (b) by **(OM3)**; (c) Since $\mathbf{B} \in \mathcal{X}$, apply (2.4) to $\mathbf{Y} \cap \mathbf{A}$ and $\mathbf{Y} \cap \mathbf{A}^{\complement}$;
(d) Since $\mathbf{A} \in \mathcal{X}$, apply (2.4) to $\mathbf{Y}$.

Hence, by Claim 1, $\mathbf{A} \cup \mathbf{B}$ is also in $\mathcal{X}$. .... □ [Claim 3]

Claim 4: *$\mathcal{X}$ is closed under countable disjoint unions.*

Proof: Suppose $\mathbf{A}_1, \mathbf{A}_2, \ldots \in \mathcal{X}$ are disjoint and let $\mathbf{U} = \bigsqcup_{n=1}^{\infty} \mathbf{A}_n$. We want to show that $\mathbf{U}$ satisfies (2.4). For all N, let $\mathbf{U}_N = \bigsqcup_{n=1}^{N} \mathbf{A}_n$, and let $\mathbf{Y} \subset \mathbf{X}$ be arbitrary.

Claim 4.1: $\widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}_N] \;=\; \sum_{n=1}^{N} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n]$.

Proof: Observe that $\mathbf{U}_N \cap \mathbf{A}_N = \mathbf{A}_N$, while $\mathbf{U}_N \cap \mathbf{A}_N^{\complement} = \mathbf{U}_{N-1}$. Since $\mathbf{A}_N \in \mathcal{X}$ apply (2.4) to conclude:

$$\widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}_N] \;=\; \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}_N \cap \mathbf{A}_N] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}_N \cap \mathbf{A}_N^{\complement}\right] \;=\; \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_N] + \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}_{N-1}].$$

Now argue inductively. .... $\square$ `[Claim 4.1]`

Claim 4.2: $\widetilde{\mu}[\mathbf{Y}] \;=\; \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right] \;=\; \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right]$.

Proof: Observe that

$$\begin{aligned}
\widetilde{\mu}[\mathbf{Y}] \;&=_{(a)}\; \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}_N] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}_N^{\complement}\right] \;=_{(b)}\; \sum_{n=1}^{N} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}_N^{\complement}\right] \\
&\geq_{(c)}\; \sum_{n=1}^{N} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right].
\end{aligned}$$

(a) $\mathbf{U}_N \in \mathcal{X}$ by Claim 3, so apply (2.4). (b) By Claim 4.1.
(c) Because $\mathbf{U}_N \subset \mathbf{U}$, so $\mathbf{U}^{\complement} \subset \mathbf{U}_N^{\complement}$; apply **(OM2)**.

Thus, letting $N \to \infty$,

$$\begin{aligned}
\widetilde{\mu}[\mathbf{Y}] \;&\geq\; \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right] \;\geq_{(a)}\; \widetilde{\mu}\left[\bigcup_{n=1}^{\infty} (\mathbf{Y} \cap \mathbf{A}_n)\right] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right] \\
&=\; \widetilde{\mu}\left[\mathbf{Y} \cap \left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right)\right] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right] \;=\; \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right] \\
&\geq_{(b)}\; \widetilde{\mu}\left[(\mathbf{Y} \cap \mathbf{U}) \sqcup \left(\mathbf{Y} \cap \mathbf{U}^{\complement}\right)\right] \;=\; \widetilde{\mu}[\mathbf{Y}].
\end{aligned}$$

(a,b) By **(OM3)**.

We conclude that these inequalities are equalities. .... $\square$ `[Claim 4.2]`

In particular, Claim 2 implies: $\widetilde{\mu}[\mathbf{Y}] = \widetilde{\mu}[\mathbf{Y} \cap \mathbf{U}] \;+\; \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right]$. This is true for any $\mathbf{Y} \subset \mathbf{X}$. Thus, $\mathbf{U}$ satisfies (2.4), as desired. .... $\square$ `[Claim 4]`

Claim 5: *$\widetilde{\mu}$ is countably additive on elements of $\mathcal{X}$.*

Proof: Again, let $\mathbf{A}_1, \mathbf{A}_2, \ldots \in \mathcal{X}$ be disjoint and $\mathbf{U} = \bigsqcup_{n=1}^{\infty} \mathbf{A}_n$. From Claim 4.2, we have, for any $\mathbf{Y} \subset \mathbf{X}$, $\widetilde{\mu}[\mathbf{Y}] = \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}_n] + \widetilde{\mu}\left[\mathbf{Y} \cap \mathbf{U}^{\complement}\right]$. Set $\mathbf{Y} = \mathbf{U}$ to get:
$\widetilde{\mu}[\mathbf{U}] = \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{U} \cap \mathbf{A}_n] + \widetilde{\mu}\left[\mathbf{U} \cap \mathbf{U}^{\complement}\right] = \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{A}_n] + \widetilde{\mu}[\emptyset] = \sum_{n=1}^{\infty} \widetilde{\mu}[\mathbf{A}_n].$
□ [Claim 5]

Claim 6: *$\mathcal{X}$ is a sigma-algebra.*

Proof: $\mathcal{X}$ is closed under complementation by Claim 2, and under finite unions by Claim 3; thus, it is closed under finite intersections by application of de Morgan's law.

Suppose $\mathbf{A}_1, \mathbf{A}_2, \ldots$ are in $\mathcal{X}$; we want to show that $\mathbf{U} = \bigcup_{n=1}^{\infty} \mathbf{A}$ is also in $\mathcal{X}$. Let $\mathbf{U}_N = \bigcup_{n=1}^{N-1} \mathbf{A}_n$; then $\mathbf{U}_N \in \mathcal{X}$ by Claim 3. Thus, $\mathbf{U}_N^{\complement} \in \mathcal{X}$ by Claim 2. Thus, $\mathbf{B}_N = \mathbf{A}_N \cap \mathbf{U}_N^{\complement} \in \mathcal{X}$.

But observe that $\mathbf{B}_1, \mathbf{B}_2, \ldots$ are disjoint, and $\mathbf{U} = \bigsqcup_{n=1}^{\infty} \mathbf{B}_n$. Thus, $\mathbf{U} \in \mathcal{X}$ by Claim 4.

Finally, apply de Morgan's law once again to conclude that $\mathcal{X}$ is closed under countable intersection. .... □ [Claim 6]

From Claims 5 and 6, it follows that $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space. It remains only to show:

Claim 7: *$(\mathbf{X}, \mathcal{X}, \mu)$ is complete.*

Proof: Suppose $\mathbf{U} \in \mathcal{X}$ has measure zero; we want to show that all subsets $\mathbf{V} \subset \mathbf{U}$ are in $\mathcal{X}$. So, let $\mathbf{Y} \subset \mathbf{X}$ be arbitrary. Then $\mathbf{V} \cap \mathbf{Y} \subset \mathbf{V} \subset \mathbf{U}$, so that $\widetilde{\mu}[\mathbf{V} \cap \mathbf{Y}] \leq \widetilde{\mu}[\mathbf{U}] = 0$, so that $\widetilde{\mu}[\mathbf{V} \cap \mathbf{Y}] = 0$. Thus

$$\widetilde{\mu}[\mathbf{Y}] \geq \widetilde{\mu}[\mathbf{Y} \cap \mathbf{V}^{\complement}] = \widetilde{\mu}[\mathbf{Y} \cap \mathbf{V}^{\complement}] + 0 = \widetilde{\mu}[\mathbf{Y} \cap \mathbf{V}^{\complement}] + \widetilde{\mu}[\mathbf{V} \cap \mathbf{Y}].$$

Apply Claim 1 to conclude that $\mathbf{V} \in \mathcal{X}$. .... □ [Claim 7 & Theorem]

$\mathcal{X}$ is sometimes called the **Carathéodory sigma-algebra**, and the measure $\mu = \widetilde{\mu}|_{\mathcal{X}}$ is the **Carathéodory measure**.

Example 53: The Lebesgue Measure

If $\mathbf{X} = \mathbb{R}$ and $\mathcal{B}$ is the set of all finite left-open intervals, then the Lebesgue outer measure (Example 51) yields the **Lebesgue measure** λ, and $\mathcal{X}$ is the **Lebesgue sigma-algebra** of $\mathbb{R}$. We will show in §2.1(b) that $\mathcal{X}$ contains all elements of the Borel sigma algebra of $\mathbb{R}$.

2.1(a) Covering Outer Measures

Let $\mathcal{B} \subset \mathcal{P}(\mathbf{X})$ be a collection of 'basic sets'. Assume that $\emptyset$ and $\mathbf{X}$ are in $\mathcal{B}$. Let $\rho : \mathcal{B} \longrightarrow [0, \infty]$ be function gauging the 'size' of the basic subsets, and assume $\rho(\emptyset) = 0$. The idea of a *covering outer measure* is to measure arbitrary subsets of $\mathbf{X}$ by covering them with elements of $\mathcal{B}$.

Example 54:

(a) **The Lebesgue Gauge:** If $\mathbf{X} = \mathbb{R}$, let $\mathcal{B}$ be the set of all **left-open intervals** $(a, b]$, with $-\infty \leq a < b \leq \infty$. Define $\rho(a, b] \;=\; b - a$.

(b) **The Stieltjes Gauge:** Again, let $\mathbf{X} = \mathbb{R}$ and let $\mathcal{B}$ be the set of all left-open intervals. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a right-continuous, monotone-increasing function. Define $\rho(a, b] \;=\; f(b) - f(a)$.

For example, if $f(x) = x$, then $\rho(a, b] = b - a$, so we recover the Lebesgue Gauge of Example (54a).

(c) **Balls in $\mathbb{R}^2$:** If $\mathbf{X} = \mathbb{R}^2$, let $\mathcal{B}$ be the set of all **open balls** in $\mathbf{X}$. For any ball $\mathbf{B}$ of radius ϵ, let $\rho(\mathbf{B}) = \pi\epsilon^2$ be its area.

(d) **The Hausdorff Gauge:** Suppose $\mathbf{X}$ is a metric space. Let $\mathcal{B}$ be the set of all open balls in $\mathbf{X}$. For any $\mathbf{B} \in \mathcal{B}$, define $\mathsf{diam}\,[\mathbf{B}] = \sup_{a,b \in \mathcal{B}} d(a, b)$. Fix some $\alpha > 0$ (the 'dimension' of $\mathbf{X}$) then define $\rho(\mathbf{B}) = \mathsf{diam}\,[\mathbf{B}]^\alpha$.

When $\mathcal{B}$ are open balls in $\mathbb{R}^2$ and $\alpha = 2$, this agrees with Example (54c), modulo multiplication by $\pi/4$.)

(e) **The Product Gauge:** Suppose $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are measure spaces. Let $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$, and let $\mathcal{R}$ be the set of all **rectangles** of the form $\mathbf{U}_1 \times \mathbf{U}_2$, where $\mathbf{U}_1 \in \mathcal{X}_1$ and $\mathbf{U}_2 \in \mathcal{X}_2$. Then define $\rho(\mathbf{U}_1 \times \mathbf{U}_2) \;=\; \mu_1[\mathbf{U}_1] \cdot \mu_2[\mathbf{U}_2]$.

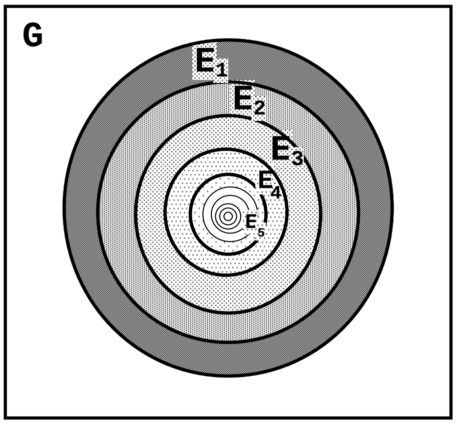
(f) **The Haar Gauge:** Let $\mathbf{G}$ be a locally compact topological group with Borel sigma-algebra $\mathcal{G}$. Let $e \in \mathbf{G}$ be the identity element, and let $\mathbf{E}_1 \supset \mathbf{E}_2 \supset \mathbf{E}_3 \supset \ldots \supset \{e\}$ be a descending sequence of open neighbourhoods of e (Figure 2.2A).

Since $\mathbf{G}$ is locally compact, we can assume $\overline{\mathbf{E}}_1$ is compact. Thus, any covering of $\overline{\mathbf{E}}_1$ by a collection $\{g_k \cdot \mathbf{E}_n\}_{k=1}^\infty$ of translates of $\mathbf{E}_n$ must have a finite subcover (Figure 2.2B). Say that a finite cover $\{g_k \cdot \mathbf{E}_n\}_{k=1}^K$ of $\overline{\mathbf{E}}_1$ is **minimal** if no subcollection of $\{g_k \cdot \mathbf{E}_n\}_{k=1}^K$ covers $\overline{\mathbf{E}}_1$. Let

$$C_n \;=\; \min\left\{K \in \mathbb{N}\;;\; \text{There is a minimal covering } \{g_k \cdot \mathbf{E}_n\}_{k=1}^K \text{ of } \overline{\mathbf{E}}_1 \text{ of cardinality } K\right\}$$

Thus, C_n measures the 'relative sizes' of $\mathbf{E}_1$ and $\mathbf{E}_n$. Loosely speaking, $\mathbf{E}_1$ is 'C_n times as big' as $\mathbf{E}_n$.

Let $\mathcal{B} \;=\; \{g \cdot \mathbf{E}_n\;;\; g \in \mathbf{G},\; n \in \mathbb{N}\}$, and, for any $(g \cdot \mathbf{E}_n) \in \mathcal{B}$, define $\rho(g \cdot \mathbf{E}_n) = \dfrac{1}{C_n}$.



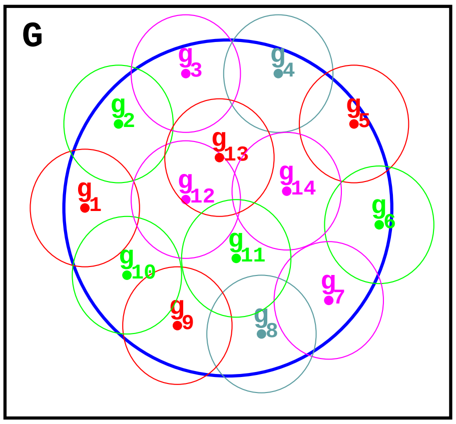


Figure 2.2: **(A)** $\mathbf{E}_1 \supset \mathbf{E}_2 \supset \mathbf{E}_3 \supset \ldots$ are neighbourhoods of the identity element $e \in \mathbf{G}$. **(B)** $\{g_k \cdot \mathbf{E}_3\}_{k=1}^{14}$ is a covering of $\mathbf{E}_1$.

For example, suppose that $\mathbf{G} = \mathbb{R}^D$, and for all n, let $\mathbf{E}_n = \left(\frac{-1}{2^n}, \frac{1}{2^n}\right)$ be the open cube of sidelength $\frac{-1}{2^{n-1}}$ around the origin. Then for all $n \in \mathbb{N}$,

$$2^{nD} \;\leq\; C_n \;\leq\; (1+c)2^{nD} \qquad (c \geq 0 \text{ some constant}) \tag{2.7}$$

so that $\rho(\mathbf{E}_n) \approx \frac{C}{2^{nD}}$ for some constant $C \geq 1$.

Exercise 52 Verify (2.7). Hint: 2^{nD} copies of $\mathbf{E}_n$ can't cover all of $\overline{\mathbf{E}}_1$, because they are *open* cubes. Push them slightly closer together and shift them slightly to one side to cover all of $\overline{\mathbf{E}}_1$ except for D faces, which are each $(D-1)$-cubes. Now argue inductively.

Let $\mathbf{U} \subset \mathbf{X}$. A (countable) **$\mathcal{B}$-covering** of $\mathbf{U}$ is a collection $\{\mathbf{B}_n\}_{n=1}^{\infty}$ of elements in $\mathcal{B}$ such that $\mathbf{U} \subset \bigcup_{n=1}^{\infty} \mathbf{B}_n$. The 'total size' of $\mathbf{U}$ should therefore be less than $\sum_{n=1}^{\infty} \rho(\mathbf{B}_n)$. Thus we can measure $\mathbf{U}$ by taking the infimum over all such coverings:

$$\widetilde{\mu}(\mathbf{U}) \;=\; \inf_{\substack{\{\mathbf{B}_n\}_{n=1}^{\infty} \subset \mathcal{B} \\ \text{covers } \mathbf{U}}} \sum_{n=1}^{\infty} \rho(\mathbf{B}_n) \tag{2.8}$$

This is the **covering outer measure** induced by $\mathcal{B}$ and ρ.

Proposition 55 *The covering outer measure defined by (2.8) is an outer measure.*

Proof: Property **(OM1)** follows immediately: $\emptyset \in \mathcal{B}$, and $\emptyset$ is a 'covering' of itself, so we have $\widetilde{\mu}(\emptyset) \leq \rho(\emptyset) = 0$.

To see property **(OM2)**, suppose that $\mathbf{U} \subset \mathbf{V}$. Then any $\mathcal{B}$-covering of $\mathbf{V}$ is also a $\mathcal{B}$-covering of $\mathbf{U}$. Thus,

$$\Big\{\{\mathbf{B}_n\}_{n=1}^{\infty} \subset \mathcal{B}\ ;\ \text{a covering of }\ \mathbf{U}\Big\} \quad \subset \quad \Big\{\{\mathbf{B}_n\}_{n=1}^{\infty} \subset \mathcal{B}\ ;\ \text{a covering of }\ \mathbf{V}\Big\}$$

so, taking the infimum over both sets in (2.8), we get $\widetilde{\mu}(\mathbf{U}) \leq \widetilde{\mu}(\mathbf{V})$.

To see property **(OM3)**, let $\mathbf{A}_m \subset \mathbf{X}$ for all $m \in \mathbb{N}$. Fix $\epsilon > 0$, and for all $m \in \mathbb{N}$, find a $\mathcal{B}$-covering $\{\mathbf{B}_n^{(m)}\}_{n=1}^{\infty}$ of $\mathbf{A}_m$ such that

$$\sum_{n=1}^{\infty} \rho\left(\mathbf{B}_n^{(m)}\right) \quad \leq \quad \widetilde{\mu}(\mathbf{A}_m) \ + \ \frac{\epsilon}{2^m}$$

Then the collection $\{\mathbf{B}_n^{(m)}\}_{n,m=1}^{\infty}$ is a $\mathcal{B}$-covering of $\bigcup_{n=1}^{\infty} \mathbf{A}_n$, so we conclude:

$$\widetilde{\mu}\left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right) \quad \leq \quad \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \rho\left(\mathbf{B}_n^{(m)}\right) \quad \leq \quad \sum_{m=1}^{\infty}\left(\widetilde{\mu}(\mathbf{A}_m) + \frac{\epsilon}{2^m}\right) \quad = \quad \left(\sum_{m=1}^{\infty} \widetilde{\mu}(\mathbf{A}_m)\right) \ + \ \epsilon$$

Since $\epsilon > 0$ is arbitrarily small, conclude that $\widetilde{\mu}\left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right) \ \leq \ \sum_{m=1}^{\infty} \widetilde{\mu}(\mathbf{A}_m)$, as desired. $\Box$

At this point, we can apply Carathéodory's theorem (page 36) to obtain a measure from $\widetilde{\mu}$.

Example 56:

(a) **The Lebesgue Outer Measure:** If $\mathbf{X} = \mathbb{R}$ and $\rho(a,b] \ = \ b - a$ as in Example (54a), then the covering outer measure

$$\widetilde{\mu}(\mathbf{U}) \quad = \quad \inf_{\substack{\{(a_n,b_n]\}_{n=1}^{\infty} \\ \text{covers } \mathbf{U}}} \ \sum_{n=1}^{\infty}(b_n - a_n)$$

is called the **Lebesgue outer measure**, and was already introduced in Example 51 on page 34. Application of Carathéodory's theorem (page 36) yields the **Lebesgue Measure** (Example 53 on page 38). The **Lebesgue sigma algebra** contains all Borel sets (see §2.1(b)).

(b) **The Stieltjes Outer Measure:** Let $\mathbf{X} = \mathbb{R}$ and let $\mathcal{B}$ be the set of all finite left-open intervals; let $f : \mathbb{R} \longrightarrow \mathbb{R}$ right-continuous and nondecreasing, and define $\rho(a,b] \ = \ f(b) - f(a)$ as in Example (54b). The covering outer measure

$$\widetilde{\mu}(\mathbf{U}) \quad = \quad \inf_{\substack{\{(a_n,b_n]\}_{n=1}^{\infty} \\ \text{covers } \mathbf{U}}} \ \sum_{n=1}^{\infty}\Big(f(b_n) - f(a_n)\Big)$$

is called the **Stieltjes outer measure** defined by f. Application of Carathéodory's theorem (page 36) yields a **Stieltjes Measure**, on the Lebesgue sigma algebra.

(c) **Product Outer Measure:** Suppose $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are measure spaces, with $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$, and $\rho(\mathbf{V}^1 \times \mathbf{V}^2) = \mu_1[\mathbf{V}^1] \cdot \mu_2[\mathbf{V}^2]$ as in Example (54e). Then the covering outer measure

$$\widetilde{\mu}(\mathbf{U}) \quad = \quad \inf_{\substack{\{\mathbf{V}_n^1 \times \mathbf{V}_n^2\}_{n=1}^\infty \\ \text{covers } \mathbf{U}}} \sum_{n=1}^\infty \Big(\mu_1\left[\mathbf{V}_n^1\right] \cdot \mu_2\left[\mathbf{V}_n^2\right]\Big)$$

is called the **product outer measure**. Application of Carathéodory's theorem (page 36) yields the **product measure**. We will show in §2.1(b) that the Carathéodory sigma-algebra contains all elements of $\mathcal{X}_1 \otimes \mathcal{X}_2$.

Exercise 53 Construct the product measure on a finite product space $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_N$.

Exercise 54 Construct the product measure on an infinite product space $\prod_{\lambda \in \Lambda} \mathbf{X}_\lambda$. Find necessary/sufficient conditions for a subset to have a finite but nonzero measure.

Sometimes we construct an outer measure as the supremum of a family of outer measures.

Lemma 57 *Suppose $\{\widetilde{\mu}_i\}_{i \in \mathcal{I}}$ is a family of outer measures, and define $\widetilde{\mu}(\mathbf{U}) = \sup_{i \in \mathcal{I}} \widetilde{\mu}_i(\mathbf{U})$ (for all $\mathbf{U} \subset \mathbf{X}$). Then $\widetilde{\mu}$ is also an outer measure.*

Proof: **Exercise 55** □

Example 58:

(a) **The Hausdorff Outer Measures:** Let $\mathbf{X}$ be a metric space. For any $\delta > 0$, let $\mathcal{B}_\delta$ the set of open balls in $\mathbf{X}$ of radius less than δ. Fix $\alpha > 0$, and define $\rho^\alpha(\mathbf{B}) = \mathsf{diam}[\mathbf{B}]^\alpha$ as in Example (54d). By Proposition 55, we obtain a covering outer measure:

$$\widetilde{\mu}_\delta^\alpha(\mathbf{U}) \quad = \quad \inf_{\substack{\{\mathbf{B}_n\}_{n=1}^\infty \subset \mathcal{B}_\delta \\ \text{covers } \mathbf{U}}} \sum_{n=1}^\infty \mathsf{diam}[\mathbf{B}_n]^\alpha. \tag{2.9}$$

We call this the **Hausdorff δ-outer measure**.

Notice that, as $\delta \to 0$, the set $\mathcal{B}_\delta$ becomes smaller and smaller. Thus, the infimum in (2.9) gets *larger*. For example, if $\mathbf{U}$ was an extremely 'tangled' curve embedded in $\mathbb{R}^2$, then we could achive a (bad) estimate of its length by simply covering it with a single large ball (see Figure 1.8(A) on page 9); a *better* estimate might be achieved by covering it with many smaller balls. We thus take the *limit* as δ goes to zero:

$$\widetilde{\mu}^\alpha(\mathbf{U}) \quad = \quad \lim_{\delta \to 0} \widetilde{\mu}_\delta^\alpha(\mathbf{U}) \quad = \quad \sup_{\delta > 0} \widetilde{\mu}_\delta(\mathbf{U})$$

By Lemma 57, $\widetilde{\mu}_\alpha$ is also an outer measure, and is called the **Hausdorff outer measure** (of dimension α). Application of Carathéodory's theorem (page 36) yields the **Hausdorff measure of dimension** α. We will show in §2.1(c) that the Carathéodory sigma-algebra contains the Borel sigma-algebra of $\mathbf{X}$.

(b) **The Haar Outer Measures:** Let $\mathbf{G}$ be a locally compact topological group, and let $\{\mathbf{E}_n\}_{n=1}^\infty$ and $\{C_n\}_{n=1}^\infty$ be as in Example (54f). For each $M \in \mathbb{N}$, let

$$\mathcal{B}_M \quad = \quad \{g \cdot \mathbf{E}_m \ ; \ m \geq M \text{ and } g \in \mathbf{G}\}.$$

By Proposition 55, we obtain a covering outer measure:

$$\widetilde{\mu}_M(\mathbf{U}) \quad = \quad \inf_{\substack{\{\mathbf{B}_n\}_{n=1}^\infty \subset \mathcal{B}_M \\ \text{covers } \mathbf{U}}} \sum_{n=1}^\infty \rho(\mathbf{B}_n) \tag{2.10}$$

As with the Hausdorff measure, as $M \to \infty$, the set $\mathcal{B}_M$ becomes smaller and smaller. Thus, the infimum in (2.10) gets *larger*. We thus define the (left) **Haar outer measure** by taking the limit as M goes to infinity:

$$\widetilde{\mu}(\mathbf{U}) \quad = \quad \lim_{M \to \infty} \widetilde{\mu}_M(\mathbf{U}) \quad = \quad \sup_{M \in \mathbb{N}} \widetilde{\mu}_M(\mathbf{U})$$

By Lemma 57, $\widetilde{\mu}$ is an outer measure. Application of Carathéodory's theorem (page 36) yields the **Haar measure** η on $\mathbf{G}$.

Exercise 56 Show that η is left-translation invariant, by construction.

Exercise 57 Modify the construction to obtain the *right* Haar measure.

Exercise 58 If $\mathbf{G} = \mathbb{R}^D$ and $\mathbf{E}_n = \left(\frac{-1}{2^n}, \frac{1}{2^n}\right)$ as in Example (54f), show that η agrees with the D-dimensional Lebesgue measure on $\mathbb{R}^D$.

We will show in §2.1(c) that the Carathéodory sigma-algebra contains the Borel sigma-algebra of $\mathbf{G}$.

2.1(b) Premeasures

Prerequisites: §2.1(a)

Let $\mathcal{B} \subset \mathcal{P}(\mathbf{X})$ and let ρ be a 'gauge' as in §2.1 above. It would be nice if the elements of $\mathcal{B}$ were themselves Carathéodory-measurable, and if the Carathéodory measure $\widetilde{\mu}$ agreed with the 'gauge' ρ on $\mathcal{B}$.

A collection $\mathcal{B} \subset \mathcal{P}(\mathbf{X})$ is called a **prealgebra**[1] if

- $\emptyset \in \mathcal{B}$.
- $\mathcal{B}$ is closed under finite intersections: if $\mathbf{A}, \mathbf{B} \in \mathcal{B}$, then $\mathbf{A} \cap \mathbf{B} \in \mathcal{B}$;

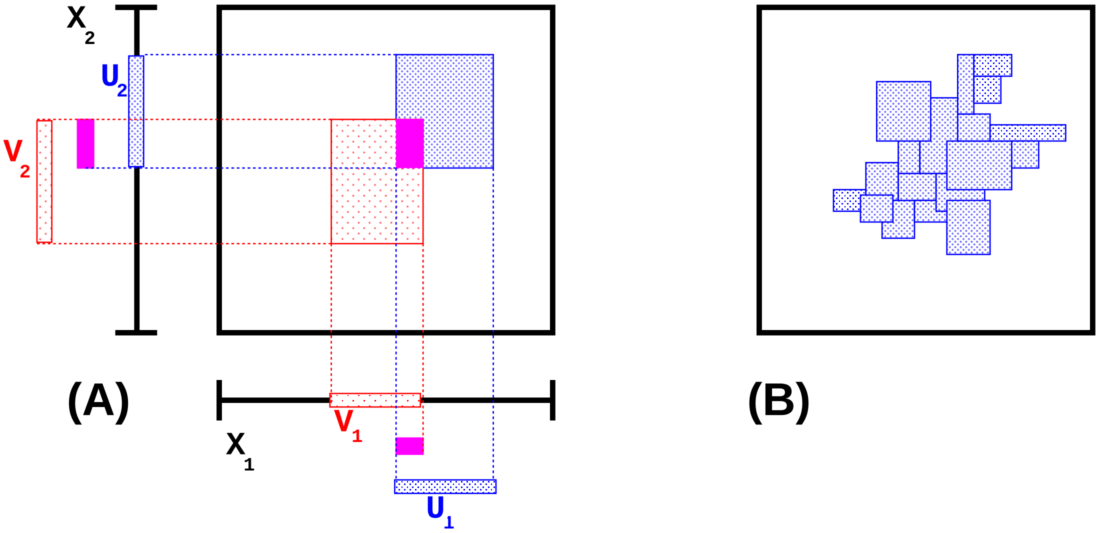


Figure 2.3: The product sigma-algebra is a prealgebra

- If $\mathbf{B} \in \mathcal{B}$, then $\mathbf{B}^{\complement}$ is a *finite disjoint union* of elements in $\mathcal{B}$.

Example 59:

(a) Let $\mathbf{X} = \mathbb{R}$ and let $\mathcal{B}$ be the set of left-open intervals, as in Example (54a). Then $\mathcal{B}$ is a pre-algebra. (**Exercise 59**)

(b) Let $(\mathbf{X}_1, \mathcal{X}_1)$ and $(\mathbf{X}_2, \mathcal{X}_2)$ be measurable spaces. Let $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$, and let $\mathcal{R}$ be the set of all **rectangles** as in Example (54e). Then $\mathcal{R}$ is a prealgebra.

Exercise 60 Verify this. Hint: If $\mathbf{U}_1 \times \mathbf{U}_2$ and $\mathbf{V}_1 \times \mathbf{V}_2$ are two such rectangles, then $(\mathbf{U}_1 \times \mathbf{U}_2) \cap (\mathbf{V}_1 \times \mathbf{V}_2) = (\mathbf{U}_1 \cap \mathbf{V}_1) \times (\mathbf{U}_2 \cap \mathbf{V}_2)$ (see Figure 2.3A) ________

We call $\mathcal{A} \subset \mathcal{P}(\mathbf{X})$ an **algebra** if $\mathcal{A}$ is closed under complementation and under *finite* unions and intersections. If $\mathcal{B} \subset \mathcal{P}(\mathbf{X})$ is any collection of sets, then let $\widetilde{\mathcal{B}}$ be the smallest algebra in $\mathcal{P}(\mathbf{X})$ containing all elements of $\mathcal{B}$.

Lemma 60 *If $\mathcal{B}$ is a prealgebra, then every element of $\widetilde{\mathcal{B}}$ can be written in a unique way as a* **disjoint** *union of elements in $\mathcal{B}$.*

Proof: **Exercise 61** Hint:

First observe that $\mathbf{A} \cup \mathbf{B} = \left(\mathbf{A}^{\complement} \cap \mathbf{B}\right) \sqcup (\mathbf{A} \cap \mathbf{B}) \sqcup \left(\mathbf{A} \cap \mathbf{B}^{\complement}\right)$. Conclude that all finite unions of $\mathcal{B}$-elements can be rewritten as finite *disjoint* unions of $\mathcal{B}$-elements. Next, if $\mathbf{A} = \bigsqcup_{n=1}^{N} \mathbf{A}_n$ and

[1] Sometimes this is called an **elementary family**.

$\mathbf{B} = \bigsqcup_{m=1}^{M} \mathbf{B}_m$ are two finite disjoint unions of $\mathcal{B}$-elements, show that $\mathbf{A} \cap \mathbf{B} = \bigsqcup_{n=1}^{N} \bigsqcup_{m=1}^{M} \left(\mathbf{A}_n \cap \mathbf{B}_n\right)$. Thus, the intersection of two finite disjoint unions of $\mathcal{B}$-elements is also a finite disjoint union of $\mathcal{B}$-elements. $\square$

Example 61:

(a) Let $\mathbf{X} = \mathbb{R}$ and let $\mathcal{B}$ be the set of left-open intervals, as in Example (59a). Then $\widetilde{\mathcal{B}}$ is the set of all finite disjoint unions of left-open intervals.

(b) Let $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$, and let $\mathcal{R}$ be the set of all **rectangles** as in Example (59b). Then $\widetilde{\mathcal{R}}$ is the set of all finite disjoint unions of rectangles (Figure 2.3B)

We call ρ a **premeasure** if, for any finite or countable disjoint collection $\{\mathbf{B}_n\}_{n=1}^{\infty} \subset \mathcal{B}$,

$$\left(\text{The disjoint union } \bigsqcup_{n=1}^{\infty} \mathbf{B}_n \text{ is also in } \mathcal{B}\right) \Longrightarrow \left(\rho\left[\bigsqcup_{n=1}^{\infty} \mathbf{B}_n\right] \quad = \quad \sum_{n=1}^{\infty} \rho[\mathbf{B}_n]\right)$$

Example 62:

(a) If $\mathbf{X} = \mathbb{R}$ and $\mathcal{B}$ is the set of all finite left-open intervals as in Example (61a), and $\rho(a, b] = b - a$ as in Example (56a) on page 41, then ρ is a premeasure. (**Exercise 62**)

(b) Let $\mathbf{X} = \mathbb{R}$ and let $\mathcal{B}$ be the set of all finite left-open intervals. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be some right-continuous, nondecreasing function, and $\rho(a, b] = f(b) - f(a)$ as in Example (56b) on page 41. Then ρ is a premeasure. (**Exercise 63**)

(c) If $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ an $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are measure spaces, and $\mathcal{R}$ is as in Example (61b), and $\rho[\mathbf{U}_1 \times \mathbf{U}_2] = \mu_1[\mathbf{U}_1] \cdot \mu_K[\mathbf{U}_2]$ as in Example (56c) on page 42, then ρ is a premeasure (**Exercise 64**).

If ρ is a premeasure, we can extend ρ to a function $\widetilde{\rho} : \widetilde{\mathcal{B}} \longrightarrow [0, \infty]$ as follows: if $\widetilde{\mathbf{B}} \in \widetilde{\mathcal{B}}$ and $\widetilde{\mathbf{B}} = \bigsqcup_{n=1}^{N} \mathbf{B}_n$ for some $\mathbf{B}_1, \ldots, \mathbf{B}_N \in \mathcal{B}$, then define

$$\widetilde{\rho}\left[\widetilde{\mathbf{B}}\right] \quad = \quad \sum_{n=1}^{N} \rho[\mathbf{B}_n] \tag{2.11}$$

Lemma 63 *Suppose $\mathcal{B}$ is a prealgebra and ρ is a premeasure.*

1. *$\widetilde{\rho}$ is well-defined by formula (2.11). That is: if $\widetilde{\mathbf{B}} \in \widetilde{\mathcal{B}}$, and $\bigsqcup_{n=1}^{N} \mathbf{B}_n = \widetilde{\mathbf{B}} = \bigsqcup_{m=1}^{M} \mathbf{B}'_m$, then $\sum_{n=1}^{N} \rho[\mathbf{B}_n] = \sum_{m=1}^{m} \rho[\mathbf{B}'_m]$.*

2. *$\widetilde{\rho}$ is also a premeasure.*

3. *The covering outer measure defined by $(\widetilde{\mathcal{B}}, \widetilde{\rho})$ is identical with that defined by $(\mathcal{B}, \rho)$.*

Proof: **(1)** Since $\mathcal{B}$ is a prealgebra, $\mathbf{B}_n \cap \mathbf{B}'_m$ is in $\mathcal{B}$ for all m and n. Observe that $\mathbf{B}_n = \bigsqcup_{m=1}^{M} (\mathbf{B}_n \cap \mathbf{B}'_m)$ and $\mathbf{B}'_m = \bigsqcup_{n=1}^{N} (\mathbf{B}_n \cap \mathbf{B}'_m)$. Since ρ is a premeasure, it follows that

$$\rho(\mathbf{B}_n) = \sum_{m=1}^{M} \rho\left(\mathbf{B}_n \cap \mathbf{B}'_m\right), \qquad \text{and} \qquad \rho(\mathbf{B}'_m) = \sum_{n=1}^{N} \rho\left(\mathbf{B}_n \cap \mathbf{B}'_m\right)$$

Thus, $$\sum_{n=1}^{N} \rho(\mathbf{B}_n) = \sum_{n=1}^{N}\sum_{m=1}^{M} \rho\left(\mathbf{B}_n \cap \mathbf{B}'_m\right), = \sum_{m=1}^{M} \rho(\mathbf{B}'_m).$$

(2 & 3) **Exercise 65** ________ □

Hence, for the remainder of this section, we can assume without loss of generality that $\mathcal{B}$ *is an algebra.*

Proposition 64 *If $\mathcal{B}$ is an algebra and ρ is a premeasure on $\mathcal{B}$, then any $\mathbf{A} \in \mathcal{B}$ is $\widetilde{\mu}$-measurable, and $\widetilde{\mu}(\mathbf{A}) = \rho(\mathbf{A})$.*

Proof: Let $\mathbf{A} \in \mathcal{B}$; first we will show that $\widetilde{\mu}(\mathbf{A}) = \rho(\mathbf{A})$.

Proof that $\widetilde{\mu}(\mathbf{A}) \leq \rho(\mathbf{A})$: This follows immediately from the definition (2.8) on page 40, because $\{\mathbf{A}\}$ is itself a $\mathcal{B}$-covering of $\mathbf{A}$.

Proof that $\widetilde{\mu}(\mathbf{A}) \geq \rho(\mathbf{A})$: Suppose $\{\mathbf{B}_n\}_{n=1}^{\infty}$ is a $\mathcal{B}$-covering of $\mathbf{A}$. For all $N \in \mathbb{N}$, define

$$\mathbf{A}_N = \mathbf{A} \cap \left(\mathbf{B}_N \setminus \bigcup_{n=1}^{N-1} \mathbf{B}_n\right)$$

Observe that:

(1) $\mathbf{A}_n \in \mathcal{B}$ (because $\mathcal{B}$ is an algebra), and $\mathbf{A}_n \subset \mathbf{B}_n$.

(2) $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots$ are disjoint, and $\mathbf{A} = \bigsqcup_{n=1}^{\infty} \mathbf{A}_n$.

Since ρ is a premeasure, it follows from **(2)** that $\rho(\mathbf{A}) = \sum_{n=1}^{\infty} \rho(\mathbf{A}_n)$. However, by **(1)**, $\rho(\mathbf{A}_n) \leq \rho(\mathbf{B}_n)$; hence $\rho(\mathbf{A}) \leq \sum_{n=1}^{\infty} \rho(\mathbf{B}_n)$. Since this is true for any $\mathcal{B}$-covering, we conclude that $\rho(\mathbf{A}) \leq \inf_{\substack{\{\mathbf{B}_n\}_{n=1}^{\infty} \\ \text{a covering of } \mathbf{A}}} \sum_{n=1}^{\infty} \rho(\mathbf{B}_n) = \widetilde{\mu}(\mathbf{A})$.

Proof that A is measurable: By Claim 1 from the proof of Carathéodory's theorem on page 36, it is sufficient to show, for any $\mathbf{Y} \subset \mathbf{X}$, that $\widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}] + \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}^{\complement}] \leq \widetilde{\mu}[\mathbf{Y}]$.
For any $\epsilon > 0$, find a $\mathcal{B}$-covering $\{\mathbf{B}_n\}_{n=1}^{\infty}$ for $\mathbf{Y}$ such that

$$\left(\sum_{n=1}^{\infty} \rho(\mathbf{B}_n) \right) \leq \widetilde{\mu}[\mathbf{Y}] + \epsilon \tag{2.12}$$

Since $\mathcal{B}$ is an algebra, the elements $\mathbf{B}_n \cap \mathbf{A}$ and $\mathbf{B}_n \cap \mathbf{A}^{\complement}$ are in $\mathcal{B}$, and, $\{\mathbf{B}_n \cap \mathbf{A}\}_{n=1}^{\infty}$ is a $\mathcal{B}$-covering for $\mathbf{Y} \cap \mathbf{A}$, while $\{\mathbf{B}_n \cap \mathbf{A}^{\complement}\}_{n=1}^{\infty}$ is a $\mathcal{B}$-covering for $\mathbf{Y} \cap \mathbf{A}^{\complement}$. Hence, we have:

$$\widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}] \leq \sum_{n=1}^{\infty} \rho\,[\mathbf{B}_n \cap \mathbf{A}] \quad \text{and} \quad \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}^{\complement}] \leq \sum_{n=1}^{\infty} \rho \left[\mathbf{B}_n \cap \mathbf{A}^{\complement}\right]. \tag{2.13}$$

Also,

$$\sum_{n=1}^{\infty} \rho[\mathbf{B}_n] = \sum_{n=1}^{\infty} \rho \left[(\mathbf{B}_n \cap \mathbf{A}) \sqcup \left(\mathbf{B}_n \cap \mathbf{A}^{\complement}\right)\right] =_{(*)} \sum_{n=1}^{\infty} \rho\,[\mathbf{B}_n \cap \mathbf{A}] + \sum_{n=1}^{\infty} \rho \left[\mathbf{B}_n \cap \mathbf{A}^{\complement}\right]. \tag{2.14}$$

where $(*)$ is because ρ is a premeasure. Thus,

$$\begin{aligned} \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}] + \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}^{\complement}] &\leq_{\text{by (2.13)}} \sum_{n=1}^{\infty} \rho\,[\mathbf{B}_n \cap \mathbf{A}] + \sum_{n=1}^{\infty} \rho \left[\mathbf{B}_n \cap \mathbf{A}^{\complement}\right] \\ &=_{\text{by (2.14)}} \sum_{n=1}^{\infty} \rho[\mathbf{B}_n] \leq_{\text{by (2.12)}} \widetilde{\mu}[\mathbf{Y}] + \epsilon. \end{aligned}$$

Let $\epsilon \to 0$ to conclude $\widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}] + \widetilde{\mu}[\mathbf{Y} \cap \mathbf{A}^{\complement}] \leq \widetilde{\mu}[\mathbf{Y}]$, as desired. ☐

Example 65:

(a) In the Lebesgue measure (Example 53 on page 38), all half-open intervals —and therefore all Borel sets in $\mathbb{R}$ —are Lebesgue-measurable. Also, the Lebesgue measure of $[a, b)$ is $b - a$, as desired.

(b) In the Stieltjes measure (Example (56b) on page 41), all half-open intervals —and therefore all Borel sets in $\mathbb{R}$ —are Lebesgue-measurable. Also, the measure of $[a, b)$ is $f(b) - f(a)$, as desired.

(c) In the product measure (Example (56c) on page 42), all rectangles —and therefore all elements of $\mathcal{X}_1 \otimes \mathcal{X}_2$ —are measurable with respect to the product measure. Also, the measure of $\mathbf{U}_1 \times \mathbf{U}_2$ is $\mu_1(\mathbf{U}_1) \times \mu_2(\mathbf{U}_2)$, as desired.

2.1(c) Metric Outer Measures

Prerequisites: §2.1(a), particularly Examples (54d), (54f), (58a) and (58b)

Let $(\mathbf{X}, d)$ be a metric space. For any $\mathbf{U}, \mathbf{V} \subset \mathbf{X}$, define

$$d(\mathbf{U}, \mathbf{V}) \quad = \quad \inf_{u \in \mathbf{U}} \inf_{v \in \mathbf{V}} d(u, v).$$

Thus, if $\mathbf{U} \cap \mathbf{V} \neq \emptyset$, then $d(\mathbf{U}, \mathbf{V}) = 0$. Conversely, if $d(\mathbf{U}, \mathbf{V}) > 0$, then $\mathbf{U}$ and $\mathbf{V}$ are disjoint.

Let $\widetilde{\mu}$ be an outer measure[2]; we call $\widetilde{\mu}$ a **metric outer measure** if:

$$\Big(d(\mathbf{U}, \mathbf{V}) > 0 \Big) \quad \Longrightarrow \quad \Big(\widetilde{\mu}(\mathbf{U} \sqcup \mathbf{V}) \;=\; \widetilde{\mu}(\mathbf{U}) + \widetilde{\mu}(\mathbf{V}) \Big)$$

Proposition 66 (Examples of Metric Outer Measures)

1. *Let $(\mathbf{X}, d)$ be a metric space. For any $\alpha > 0$, the α-dimensional Hausdorff outer measure (Example (58a) on page 42) is a metric outer measure.*

2. *Let $\mathbf{G}$ be a locally compact Hausdorff topological group. Then $\mathbf{G}$ is metrizable[3], and the Haar outer measure (Example (58b) on page 43) is a metric outer measure.*

Proof: **Proof of (1)** Suppose $\mathbf{U}^1, \mathbf{U}^2 \subset \mathbf{X}$ and $d(\mathbf{U}^1, \mathbf{U}^2) = \epsilon > 0$. Fix $\delta < \epsilon$; let $\mathcal{B}_\delta$ the set of open balls of radius less than δ, and let $\{\mathbf{B}_n\}_{n=1}^\infty \subset \mathcal{B}_\delta$.

Claim 1: *$\{\mathbf{B}_n\}_{n=1}^\infty$ covers $\mathbf{U}^1 \sqcup \mathbf{U}^2$ if and only if we can split $\{\mathbf{B}_n\}_{n=1}^\infty$ into two subfamilies: $\{\mathbf{B}_n\}_{n=1}^\infty = \{\mathbf{B}_n^1\}_{n=1}^\infty \sqcup \{\mathbf{B}_n^2\}_{n=1}^\infty$, where $\{\mathbf{B}_n^1\}_{n=1}^\infty$ covers $\mathbf{U}^1$ and $\{\mathbf{B}_n^2\}_{n=1}^\infty$ covers $\mathbf{U}^2$.*

Proof: If $\{\mathbf{B}_n^j\}_{n=1}^\infty$ covers $\mathbf{U}^j$ for $j = 1, 2$, then clearly $\{\mathbf{B}_n^1\}_{n=1}^\infty \sqcup \{\mathbf{B}_n^2\}_{n=1}^\infty$ covers $\mathbf{U}^1 \sqcup \mathbf{U}^2$. Conversely, suppose that $\{\mathbf{B}_n\}_{n=1}^\infty$ covers $\mathbf{U}^1 \sqcup \mathbf{U}^2$. Since $d(\mathbf{U}^1, \mathbf{U}^2) = \epsilon > \delta$, no element of $\{\mathbf{B}_n\}_{n=1}^\infty$ can simultaneously intersect $\mathbf{U}^1$ and $\mathbf{U}^2$. Thus, we can split $\{\mathbf{B}_n\}_{n=1}^\infty$ into two disjoint subcollections $\{\mathbf{B}_n^1\}_{n=1}^\infty$ and $\{\mathbf{B}_n^2\}_{n=1}^\infty$ which cover $\mathbf{U}^1$ and $\mathbf{U}^2$ respectively. □ [Claim 1]

[2]See page 34.

[3]That is, there is a metric on $\mathbf{G}$ which is compatible with the topology. Note that we do *not* assume this metric is $\mathbf{G}$-invariant.

If $\{\mathbf{B}_n\}_{n=1}^\infty = \{\mathbf{B}_n^1\}_{n=1}^\infty \sqcup \{\mathbf{B}_n^2\}_{n=1}^\infty$ as in Claim 1, then clearly,

$$\sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n\right]^\alpha = \sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n^1\right]^\alpha + \sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n^2\right]^\alpha$$

Then, taking infimums, we get:

$$\begin{aligned}\widetilde{\mu}_\delta^\alpha(\mathbf{U}) &= \inf_{\substack{\{\mathbf{B}_n\}_{n=1}^\infty\subset\mathcal{B}_\delta \\ \text{covers } \mathbf{U}}} \sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n\right]^\alpha \\ &= \inf_{\substack{\{\mathbf{B}_n^1\}_{n=1}^\infty\subset\mathcal{B}_\delta \\ \text{covers } \mathbf{U}^1}} \sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n^1\right]^\alpha + \inf_{\substack{\{\mathbf{B}_n^2\}_{n=1}^\infty\subset\mathcal{B}_\delta \\ \text{covers } \mathbf{U}^2}} \sum_{n=1}^\infty \mathsf{diam}\left[\mathbf{B}_n^2\right]^\alpha \\ &= \widetilde{\mu}_\delta^\alpha(\mathbf{U}^1) + \widetilde{\mu}_\delta^\alpha(\mathbf{U}^2)\end{aligned}$$

This is true for any $\delta < \epsilon$. Thus, taking the limit as $\delta \to 0$, we conclude:

$$\widetilde{\mu}^\alpha(\mathbf{U}) = \widetilde{\mu}^\alpha(\mathbf{U}^1) + \widetilde{\mu}^\alpha(\mathbf{U}^2).$$

Proof of (2) The metrizability of $\mathbf{G}$ is a standard result[4]. Let $\mathbf{E}_1 \supset \mathbf{E}_2 \supset \mathbf{E}_3 \supset \ldots$ be a descending sequence of open neighbourhoods of the identity element $e \in \mathbf{G}$, as in Example (54f) on page 39. Assume without loss of generality that $\mathsf{diam}\left[\mathbf{E}_n\right] < \frac{1}{n}$. Now argue exactly as in **(1)**. □

Proposition 67 *If $\widetilde{\mu}$ is a metric outer measure, then all Borel subsets of $\mathbf{X}$ are $\widetilde{\mu}$-measurable.*

Proof: Since the Borel sigma-algebra is generated by the closed subsets of $\mathbf{X}$, it suffices to show that all closed subsets of $\mathbf{X}$ are $\widetilde{\mu}$-measurable. So, let $\mathbf{C} \subset \mathbf{X}$ be closed. By Claim 1 from the proof of Caratheodory's theorem on page 36, $\mathbf{C} \in \mathcal{X}$ if and only if:

$$\text{For any } \mathbf{Y} \subset \mathbf{X}, \text{ with } \widetilde{\mu}\left[\mathbf{Y}\right] < \infty, \qquad \widetilde{\mu}\left[\mathbf{Y}\right] \geq \widetilde{\mu}\left[\mathbf{C} \cap \mathbf{Y}\right] + \widetilde{\mu}\left[\mathbf{C}^{\complement} \cap \mathbf{Y}\right]. \tag{2.15}$$

So, let $\mathbf{Y} \subset \mathbf{X}$ with $\widetilde{\mu}\left[\mathbf{Y}\right] < \infty$, and for all $n \in \mathbb{N}$, define

$$\mathbf{Y}_n = \left\{ y \in \mathbf{Y} \cap \mathbf{C}^{\complement} \; ; \; d(y, \mathbf{C}) > \frac{1}{2^n} \right\}.$$

Claim 1: *For any $n \in \mathbb{N}$,* $\widetilde{\mu}\left(\mathbf{Y}\right) \geq \widetilde{\mu}\left(\mathbf{Y} \cap \mathbf{C}\right) + \widetilde{\mu}\left(\mathbf{Y}_n\right)$.

[4] See for example [?] Exercise 38C, p. 260.

Proof: $\widetilde{\mu}(\mathbf{Y}) \;=\; \widetilde{\mu}\left((\mathbf{Y}\cap\mathbf{C}) \;\sqcup\; \left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right)\right) \;\geq_{(a)}\; \widetilde{\mu}\left((\mathbf{Y}\cap\mathbf{C}) \;\sqcup\; \mathbf{Y}_n\right)$
$=_{(b)} \;\; \widetilde{\mu}(\mathbf{Y}\cap\mathbf{C}) \;+\; \widetilde{\mu}(\mathbf{Y}_n)$.

(a) Because $\mathbf{Y}_n \subset \mathbf{Y}\cap\mathbf{C}^{\complement}$, and $\widetilde{\mu}$ is an outer measure (property **(OM2)** on page 34).

(b) Because $d(\mathbf{Y}\cap\mathbf{C},\ \mathbf{Y}_n) \;>\; \frac{1}{2^n}$, and $\widetilde{\mu}$ is a *metric* outer measure. . . . . ☐ `[Claim 1]`

Claim 2: $\widetilde{\mu}\left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right) \;=\; \lim_{n\to\infty} \widetilde{\mu}(\mathbf{Y}_n)$.

Proof:

Claim 2.1: $\lim_{n\to\infty} \widetilde{\mu}(\mathbf{Y}_n) \;=\; \sup_{n\in\mathbb{N}} \widetilde{\mu}(\mathbf{Y}_n) \;\leq\; \widetilde{\mu}\left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right)$.

Proof: $\mathbf{Y}_1 \subset \mathbf{Y}_2 \subset \ldots \subset \mathbf{Y}\cap\mathbf{C}^{\complement}$, so $\widetilde{\mu}(\mathbf{Y}_1) \leq \widetilde{\mu}(\mathbf{Y}_2) \leq \ldots \leq \widetilde{\mu}\left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right)$. The claim follows. . . . . ☐ `[Claim 2.1]`

It remains to show the reverse inequality, namely, that $\widetilde{\mu}\left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right) \;\leq\; \sup_{n\in\mathbb{N}} \widetilde{\mu}(\mathbf{Y}_n)$. For all $n\in\mathbb{N}$, define $\mathbf{U}_n \;=\; \mathbf{Y}_n \setminus \mathbf{Y}_{n-1}$.

Claim 2.2: *For any $n\in\mathbb{N}$,* $\;d(\mathbf{U}_{n+2},\mathbf{U}_n) \geq \frac{1}{2^{n+1}}$.

Proof: By contradiction, suppose $d(\mathbf{U}_{n+2},\mathbf{U}_n) < \frac{1}{2^{n+1}}$. Thus, there is some $u_0\in\mathbf{U}_n$ and $u_2\in\mathbf{U}_{n+2}$ so that $d(u_0,u_2) < \frac{1}{2^{n+1}}$. But $u_0\in\mathbf{U}_n\subset\mathbf{Y}_n$, so $d(u_0,\mathbf{C}) > \frac{1}{2^n}$. Meanwhile, $u_2\in\mathbf{U}_{n+2} \;=\; \mathbf{Y}_{n+2}\setminus\mathbf{Y}_{n+1}$, so that $d(u_2,\mathbf{C}) \leq \frac{1}{2^{n+1}}$. Thus, $d(u_0,\mathbf{C}) \;\leq\; d(u_0,u_2) + d(u_2,\mathbf{C}) \;<\; \frac{1}{2^{n+1}} + \frac{1}{2^{n+1}} = \frac{1}{2^n}$, contradicting that $d(u_0,\mathbf{C}) > \frac{1}{2^n}$. . . . ☐ `[Claim 2.2]`

Claim 2.3: $\widetilde{\mu}\left(\bigsqcup_{n=1}^{N} \mathbf{U}_{2n}\right) \;=\; \sum_{n=1}^{N} \widetilde{\mu}(\mathbf{U}_{2n})$ *and* $\widetilde{\mu}\left(\bigsqcup_{n=0}^{N} \mathbf{U}_{2n+1}\right) \;=\; \sum_{n=0}^{N} \widetilde{\mu}(\mathbf{U}_{2n+1})$.

Proof: $\widetilde{\mu}$ is a *metric* outer measure, so by Claim 2.2, $\widetilde{\mu}(\mathbf{U}_2\sqcup\mathbf{U}_4\sqcup\ldots\sqcup\mathbf{U}_{2N}) \;=\; \widetilde{\mu}(\mathbf{U}_2) + \widetilde{\mu}(\mathbf{U}_4\sqcup\ldots\sqcup\mathbf{U}_{2N})$. Apply induction. Do the same for the $\mathbf{U}_{2n+1}$ equation. ☐ `[Claim 2.3]`

Claim 2.4: $\sum_{n=1}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n})$ *and* $\sum_{n=0}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n+1})$ *are finite.*

Proof: By Claim 2.3, $\sum_{n=1}^{N} \widetilde{\mu}(\mathbf{U}_{2n}) \;=\; \widetilde{\mu}\left(\bigsqcup_{n=1}^{N} \mathbf{U}_{2n}\right) \;\leq\; \widetilde{\mu}\left(\bigsqcup_{n=1}^{\infty} \mathbf{U}_n\right) \;\leq\; \widetilde{\mu}(\mathbf{Y}\cap\mathbf{C}^{\complement})$.

Take the limit as $N\to\infty$ to conclude that $\sum_{n=1}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n}) \;\leq\; \widetilde{\mu}(\mathbf{Y}\cap\mathbf{C}^{\complement}) \;\leq\; \widetilde{\mu}(\mathbf{Y})$. But $\widetilde{\mu}(\mathbf{Y})$ is finite by hypothesis. . . . . ☐ `[Claim 2.4]`

Claim 2.5: $\widetilde{\mu}\left(\mathbf{Y}\cap\mathbf{C}^{\complement}\right) \;\leq\; \sup_{n\in\mathbb{N}} \widetilde{\mu}(\mathbf{Y}_n)$.

Proof: Let $\epsilon > 0$. It follows from Claim 2.4 that, for large enough N, we have

$$\sum_{n=N}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n}) < \frac{\epsilon}{2}, \qquad \text{and} \qquad \sum_{n=N}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n+1}) < \frac{\epsilon}{2}.$$

But we also know that $\mathbf{Y} \cap \mathbf{C}^{\complement} = \mathbf{Y}_{2N} \sqcup \left(\bigsqcup_{n=N}^{\infty} \mathbf{U}_{2n}\right) \sqcup \left(\bigsqcup_{n=N}^{\infty} \mathbf{U}_{2n+1}\right)$, so that, by property **(OM3)** of any outer measure (page 34),

$$\widetilde{\mu}\left(\mathbf{Y} \cap \mathbf{C}^{\complement}\right) \leq \widetilde{\mu}(\mathbf{Y}_{2N}) + \sum_{n=N}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n}) + \sum_{n=N}^{\infty} \widetilde{\mu}(\mathbf{U}_{2n+1}) \leq \sup_{n \in \mathbb{N}} \widetilde{\mu}(\mathbf{Y}_n) + \epsilon$$

Since ϵ is arbitrary, we conclude that $\widetilde{\mu}\left(\mathbf{Y} \cap \mathbf{C}^{\complement}\right) \leq \sup_{n \in \mathbb{N}} \widetilde{\mu}(\mathbf{Y}_n)$, as desired. □ `[Claims 2.5 & 2`

Combining Claim 1 and Claim 2 yields assertion (2.15), as desired. □

Corollary 68 *If* $\mathbf{X}$ *is a metric space, then the Hausdorff measure is a Borel measure.*

If $\mathbf{G}$ *is a locally compact Hausdorff group, then the Haar measure is a Borel measure.* □

2.2 More About Stieltjes Measures

Prerequisites: §2.1(a), particularly Examples (54b) and (56b); §2.1(b), particularly Examples (62b) and (65b)

Recommended: Example 1.1(c)⟨vii⟩ on page 10

Stieltjes measures are the simplest class of measures on $\mathbb{R}$ besides the Lebesgue measure, and often arise in probability theory. If μ_f is the Stieltjes measure determined by f, then f is called the **accumulation function** or **cumulative distribution function** for μ_f. It will help to keep the following examples in mind:

Example 69:

(a) **The Lebesgue Measure:** Let $f(x) = x$. Then μ_f is the Lebesgue measure.

(b) **Antiderivatives:** Let $f(x) = \arctan(x)$ (Figure 2.4A) Then, for any interval $[a, b]$,

$$\mu_f[a,b] = \arctan(b) - \arctan(a) = \int_a^b \frac{1}{1+x^2}\, dx, \qquad \text{(Figure 2.4B)}$$

because $f(x) = \arctan(x)$ is the antiderivative of $f'(x) = \frac{1}{1+x^2}$.

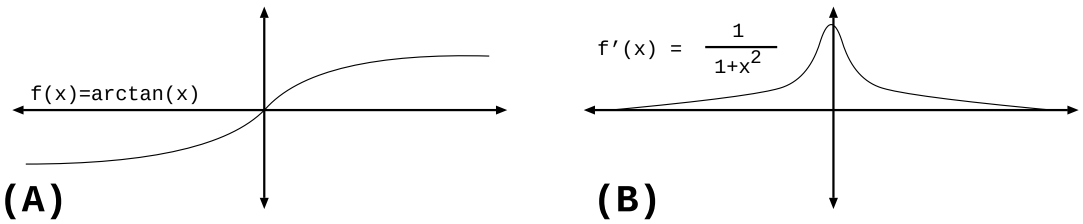


Figure 2.4: **Antiderivatives:** If $f(x) = \arctan(x)$, then $\mu_f[a,b] \;=\; \int_a^b \frac{1}{1+x^2}\,dx$.

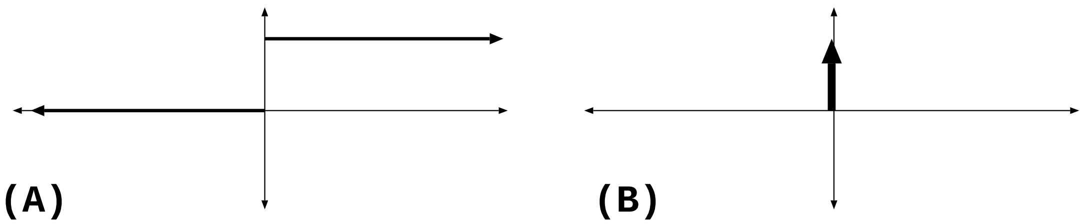


Figure 2.5: The Heaviside step function.

(c) **The Heaviside Step Function:** Let $f(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases}$ (Figure 2.5A). Then for any measurable $\mathbf{U} \subset \mathbb{R}$,

$$\mu_f[\mathbf{U}] \quad = \quad \begin{cases} 1 & \text{if } 0 \in \mathbf{U} \\ 0 & \text{if } 0 \not\in \mathbf{U} \end{cases} \qquad (\textbf{\underline{Exercise 66}}\,; \text{ see Figure 2.5B})$$

Thus, μ_f possess a single 'atom' of mass 1 at zero. μ_f is sometimes called the **point mass** or the **Dirac delta function** (even though it is not a function), and written as δ_0. In general, if $y \in \mathbb{R}$ and we want an atom at y, we define $f(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases}$ Thus, for any measurable $\mathbf{U} \subset \mathbb{R}$, $\mu_f[\mathbf{U}] = \begin{cases} 1 & \text{if } y \in \mathbf{U}; \\ 0 & \text{if } y \not\in \mathbf{U}. \end{cases}$ We call $\delta_y = \mu_f$ the **point mass at** y.

(d) **The Floor Function:** Let $f(x) = \lfloor x \rfloor$. That is, $f(x) = \max\{n \in \mathbb{Z}\;;\; n \leq x\}$ (Figure 2.6A). Then for any measurable $\mathbf{U} \subset \mathbb{R}$,

$$\mu_f[\mathbf{U}] \quad = \quad \mathsf{card}\,[\mathbf{Z} \cap \mathbf{U}] \qquad (\textbf{\underline{Exercise 67}}\,; \text{see Figure 2.6B})$$

Thus, μ_f has an atom of mass 1 at every integer. In other words, $\mu_f \;=\; \sum_{n=-\infty}^{\infty} \delta_n$.

(e) **The Devil's Staircase:** Let $\mathbf{K} \subset \mathbb{R}$ be the Cantor set (Example 16b on page 18), and define $f(x) = \sup\{k \in \mathbf{K}\;;\; k \leq x\}$ (Figure 2.7A). Then f is nondecreasing and right-continuous (**Exercise 68**). If μ_f is the corresponding Stieltjes measure (Figure 2.7B), then $\mu_f(\mathbf{K}) = 1$, and $\mu_f(\mathbf{K}^{\complement}) = 0$ (**Exercise 69**).

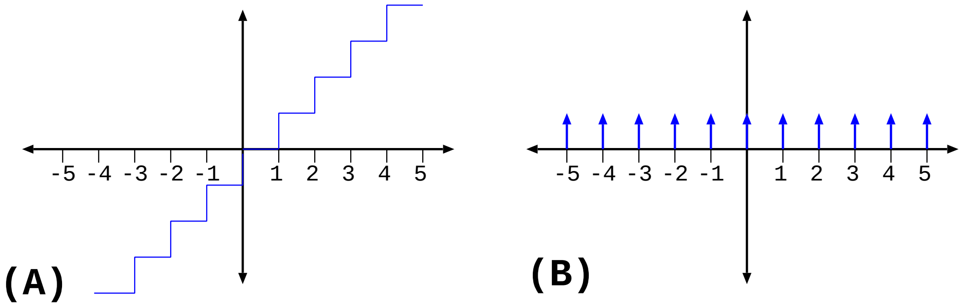


Figure 2.6: The floor function $f(x) = \lfloor x \rfloor$.

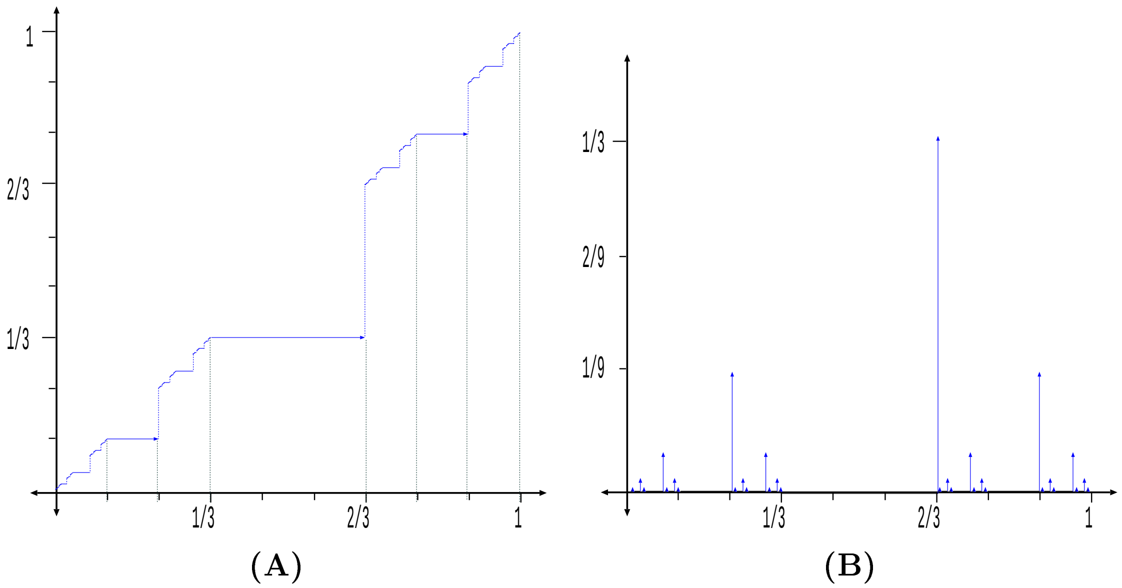


Figure 2.7: The Devil's staircase.

(f) Suppose we enumerate the rational numbers: $\mathbb{Q} = \{q_1, q_2, q_3, \ldots\}$. Define $f(x) = \sum_{q_n \leq x} \frac{1}{2^n}$. Then f is nondecreasing and right-continuous (**Exercise 70**). If μ_f is the corresponding Stieltjes measure, then $\mu_f(\mathbb{Q}) = 1$, and $\mu_f(\not{\mathbb{Q}}) = 0$ (**Exercise 71**).

Exercise 72 Suppose we apply the 'Devil's staircase' construction to the rational numbers, and define $f(x) = \sup\{q \in \mathbb{Q}\, ;\ q \leq x\}$. Show that μ_f is just the Lebesgue measure.

Theorem 70 (Properties of the Stieltjes Measure)

Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a nondecreasing, right-continuous function, and let μ_f be the corresponding Stieltjes measure.

1. *If $-\infty \leq a \leq b \leq \infty$, then*

$$\begin{array}{rclcrcl} \mu_f(a,b] & = & f(b) - f(a); & & \mu_f(a,b) & = & \lim_{x \nearrow b} f(x) - f(a); \\ \mu_f[a,b] & = & f(b) - \lim_{x \nearrow a} f(x); & \text{and} & \mu_f[a,b) & = & \lim_{x \nearrow b} f(x) - \lim_{x \nearrow a} f(a); \end{array}$$

2. *Suppose $a \in \mathbb{R}$ is a point of discontinuity of f, so that $\lim_{x \nearrow a} f(x) < f(a)$. Then a is an atom of μ_f, and $\mu_f\{a\} = f(a) - \lim_{x \nearrow a} f(x)$. Furthermore, every atom arises in this manner.*

3. *μ_f has at most a countable number of atoms.*

Proof: **Exercise 73** $\square$

A measure μ on the Borel sigma algebra of $\mathbb{R}$ is called **locally finite** if $\mu[\mathbf{K}] < \infty$ for any compact $\mathbf{K} \subset \mathbb{R}$.

Theorem 71 (Local Finiteness)

Let μ be a measure on the Borel sigma algebra of $\mathbb{R}$. Then:
$\Big($ *μ is locally finite* $\Big) \iff \Big($ *μ is the Stieltjes measure for some function f.* $\Big)$

Proof: '$\Longleftarrow$': Suppose μ_f is a Stieltjes measure, and $\mathbf{K} \subset \mathbb{R}$ is compact. Then $\mathbf{K} \subset [a,b)$ for some $-\infty < a < b < \infty$, and thus, $\mu_f[\mathbf{K}] \leq \mu_f(a,b] = f(b) - f(a) < \infty$.

'$\Longrightarrow$': Suppose μ is locally finite, and define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by: $f(x) = \begin{cases} \mu(0,x] & \text{if } x > 0 \\ -\mu(x,0] & \text{if } x \leq 0 \end{cases}$.

It is **Exercise 74** to verify:

1. f is nondecreasing and right-continuous.
2. μ is the Stieltjes measure defined by f. $\square$

A measure μ on the Borel sigma algebra of $\mathbb{R}$ is called **regular** if, for any measurable $\mathbf{U} \subset \mathbb{R}$, we have:

$$\mu[\mathbf{U}] \quad = \quad \inf_{\substack{\mathbf{O} \text{ open} \\ \mathbf{U}\subset\mathbf{O}}} \mu[\mathbf{O}] \qquad \text{and} \qquad \mu[\mathbf{U}] \quad = \quad \sup_{\substack{\mathbf{K} \text{ compact} \\ \mathbf{K}\subset\mathbf{U}}} \mu[\mathbf{K}]$$

Theorem 72 (Regularity)

Let μ_f be a Stieltjes measure on $\mathbb{R}$. Then μ is regular.

Proof:

Claim 1: *If $-\infty < a \leq b < \infty$, then for any $\epsilon > 0$, there is some $B > b$ so that $\mu_f(a,B) < \epsilon + \mu_f(a,b]$.*

Proof: f is right-continuous, so for any ϵ, there is some $B > b$ so that $f(B) < \epsilon + f(b)$. Thus, $\lim_{x\nearrow B} f(x) \leq f(B) < f(b)+\epsilon$. Thus, $\mu_f(a,b) = \lim_{x\nearrow B} f(x)-f(a) < f(b)-f(a)+\epsilon = \mu_f(a,b]+\epsilon$. .... □ [Claim 1]

Claim 2: *Let $\mathbf{U} \subset \mathbb{R}$ be measurable. Then $\mu_f[\mathbf{U}] = \inf_{\substack{\mathbf{O} \text{ open} \\ \mathbf{U}\subset\mathbf{O}}} \mu_f[\mathbf{O}]$.*

Proof: Fix $\epsilon > 0$, and let $\{(a_n,b_n]\}_{n=1}^{\infty}$ be a covering of $\mathbf{U}$ so that $\sum_{n=1}^{\infty} \mu_f(a_n,b_n] < \epsilon+\mu_f[\mathbf{U}]$. By Claim 1, for all $n \in \mathbb{N}$, find B_n so that $\mu_f(a_n,B_n) < \frac{\epsilon}{2^n} + \mu_f(a_n,b_n]$. Now let $\mathbf{V} = \bigcup_{n=1}^{\infty} (a_n,B_n)$. Then $\mathbf{V}$ is open, $\mathbf{U} \subset \mathbf{V}$, and

$$\inf_{\substack{\mathbf{O} \text{ open} \\ \mathbf{U}\subset\mathbf{O}}} \mu_f[\mathbf{O}] \;\leq\; \mu_f[\mathbf{V}] \;\leq\; \sum_{n=1}^{\infty} \mu_f(a_n,B_n) \;\leq\; \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} + \mu_f(a_n,b_n]$$

$$= \;\epsilon + \sum_{n=1}^{\infty} \mu_f(a_n,b_n] \;<\; 2\epsilon + \mu_f[\mathbf{U}].$$

Since ϵ is arbitrary, we conclude that $\inf_{\substack{\mathbf{U}\subset\mathbf{O} \\ \mathbf{O} \text{ open}}} \mu_f[\mathbf{O}] \leq \mu_f[\mathbf{U}]$. The reverse inequality is immediate, so we're done. .... □ [Claim 2]

Claim 3: *If $n \in \mathbb{N}$ and $\mathbf{U} \subset [-n,n]$ is measurable, then $\mu[\mathbf{U}] = \sup_{\substack{\mathbf{K}\subset\mathbf{U} \\ \mathbf{K} \text{ compact}}} \mu[\mathbf{K}]$.*

Proof: Let $\mathbb{I} = [-n,n]$, and let $\mathbf{U}^{\complement} = \mathbb{I} \setminus \mathbf{U}$. By Claim 1, $\mu[\mathbf{U}^{\complement}] = \inf_{\substack{\mathbf{U}^{\complement}\subset\mathbf{O}\subset\mathbb{I} \\ \mathbf{O} \text{ open}}} \mu[\mathbf{O}]$. Thus,

$$\mu[\mathbf{U}] \;=\; \mu[\mathbb{I}] - \mu[\mathbf{U}^{\complement}] \;=\; 2n - \inf_{\substack{\mathbf{U}^{\complement}\subset\mathbf{O}\subset\mathbb{I} \\ \mathbf{O} \text{ open}}} \mu[\mathbf{O}] \;=\; \sup_{\substack{\mathbf{U}^{\complement}\subset\mathbf{O}\subset\mathbb{I} \\ \mathbf{O} \text{ open}}} (2n - \mu[\mathbf{O}])$$

$$= \sup_{\substack{\mathbf{U}^\complement \subset \mathbf{O} \subset \mathbb{I} \\ \mathbf{O} \text{ open}}} \mu\left[\mathbf{O}^\complement\right], \qquad \text{where } \mathbf{O}^\complement = \mathbb{I} \setminus \mathbf{O}.$$

However, the complements of any open $\mathbf{O} \subset \mathbb{I}$ is compact, and vice versa. Indeed,

$$\left\{\mathbf{O}^\complement \ ; \ \mathbf{U}^\complement \subset \mathbf{O} \subset \mathbb{I} \text{ and } \mathbf{O} \text{ open}\right\} \quad = \quad \{\mathbf{K} \ ; \ \mathbf{U} \subset \mathbf{K} \subset \mathbb{I} \text{ and } \mathbf{K} \text{ compact}\}$$

Thus, $\displaystyle\sup_{\substack{\mathbf{U}^\complement \subset \mathbf{O} \subset \mathbb{I} \\ \mathbf{O} \text{ open}}} \mu\left[\mathbf{O}^\complement\right] \quad = \quad \sup_{\substack{\mathbf{K} \subset \mathbf{U} \\ \mathbf{K} \text{ compact}}} \mu[\mathbf{K}]$. .... □ `[Claim 3]`

Claim 4: *Suppose $\mathbf{U} \subset \mathbb{R}$ is measurable. Then* $\displaystyle\mu[\mathbf{U}] = \sup_{\substack{\mathbf{K} \subset \mathbf{U} \\ \mathbf{K} \text{ compact}}} \mu[\mathbf{K}]$.

Proof: Let $\mathbf{U}_n = \mathbf{U} \cap [-n, n]$ for all $n \in \mathbb{N}$. By Claim 3, we can find a compact $\mathbf{K}_n \subset \mathbf{U}_n$ so that $\mu[\mathbf{U}_n] < \frac{1}{n} + \mu[\mathbf{K}_n]$. Thus,

$$\mu[\mathbf{U}] \quad =_{(a)} \quad \lim_{n\to\infty} \mu[\mathbf{U}_n] \quad = \quad \lim_{n\to\infty} \mu[\mathbf{U}_n] - \frac{1}{n} \quad \leq \quad \lim_{n\to\infty} \mu[\mathbf{K}_n] \quad \leq_{(b)} \quad \mu[\mathbf{U}].$$

(a) Because $\mathbf{U}_1 \subset \mathbf{U}_2 \subset \ldots$ and $\mathbf{U} = \bigcup_{n=1}^{\infty} \mathbf{U}_n$. (b) Because $\mathbf{K}_n \subset \mathbf{U}_n \subset \mathbf{U}$ for all n.

We conclude that $\displaystyle\lim_{n\to\infty} \mu[\mathbf{K}_n] = \mu[\mathbf{U}]$. .... □ `[Claim 4 & Theorem]`

This yields the following result, which says that, modulo sets of measure zero, all measurable subsets of $\mathbb{R}$ are quite 'nice'.

Corollary 73 *Let μ_f be a Stieltjes measure, and let $\mathbf{U} \subset \mathbb{R}$. The following are equivalent:*

1. *$\mathbf{U}$ is measurable.*
2. *$\mathbf{U} = \mathbf{G} \setminus \mathbf{Z}$ where $\mathbf{G}$ is G_δ and $\mathbf{Z}$ has μ_f-measure zero.*
3. *$\mathbf{U} = \mathbf{F} \sqcup \mathbf{Z}$ where $\mathbf{F}$ is F_σ and $\mathbf{Z}$ has μ_f-measure zero.*

Proof: Clearly (2$\Longrightarrow$1) and (3$\Longrightarrow$1)

(1$\Longrightarrow$2): Let $\mathbf{O}_1 \supset \mathbf{O}_2 \supset \ldots \supset \mathbf{U}$ be a sequence of open sets such that $\mu_f[\mathbf{U}] = \lim_{n\to\infty} \mu_f[\mathbf{O}_n]$. Let $\mathbf{G} = \bigcap_{n=1}^{\infty} \mathbf{O}_n$. Then $\mathbf{G}$ is G_δ, and $\mathbf{U} \subset \mathbf{G}$, and $\mu_f[\mathbf{G}] = \lim_{n\to\infty} \mu_f[\mathbf{O}_n] = \mu_f[\mathbf{U}]$. Thus, $\mathbf{Z} = \mathbf{G} \setminus \mathbf{U}$ has measure zero.

(1$\Longrightarrow$3) is proved similarly, using compact sets. □

Example (2.4) is just a special case of the following:

Theorem 74 (Stieltjes Measures and the Fundamental Theorem of Calculus)

1. *Let $\phi : \mathbb{R} \longrightarrow [0, \infty)$ be continuous. Define μ by $\mu[\mathbf{U}] = \int_{\mathbf{U}} \phi(x)\ d\lambda[x]$, where λ is the Lebesgue measure. Then μ is the Stieltjes measure determined by f, where f is any* **antiderivative** *of ϕ —that is, any function satisfying $f' = \phi$.*

2. *Conversely, suppose $f : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable, and let μ_f be the corresponding Stieltjes measure. Then $\mu_f[\mathbf{U}] = \int_{\mathbf{U}} f'(x)\ d\lambda[x]$.*

Proof: <u>**Exercise 75**</u> □

Even when f is nondifferentiable —or even discontinuous —we can think of the measure μ_f as a kind of generalized 'derivative' of f. For example, the Dirac delta function δ_0 from Example (69c) is the 'derivative' of the Heaviside step function. These loose ideas are made precise in the theory of **distributions** developed by Laurent Schwartz [**?**].

2.3 Signed and Complex-valued Measures

Measures were invented to endow subsets with a notion of *magnitude*, but it is useful to allow measures to take on negative, or even complex, values. This makes set of all measures on a measurable space $(\mathbf{X}, \mathcal{X})$ into a *vector space*.

Definition 75 Signed Measure

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

1. $\mu[\emptyset] = 0$.

2. *μ assumes at most* **one** *of the values ∞ or $-\infty$; it cannot assume both.*

3. *μ* **countably additive***: for disjoint collection $\mathbf{U}_1, \mathbf{U}_2, \ldots \in \mathcal{X}$, we have:*

$$\mu\left[\bigsqcup_{n=1}^{\infty} \mathbf{U}_n\right] \quad = \quad \sum_{n=1}^{\infty} \mu\left[\mathbf{U}_n\right], \tag{2.16}$$

where the sum on the right hand side of (2.16) converges **absolutely***, either to a real number, or to $+\infty$ or $-\infty$.*

Remark:

- It is important that μ cannot contain both $+\infty$ and $-\infty$ in its range. After all, if $\mu[\mathbf{A}] = \infty$, and $\mu[\mathbf{B}] = -\infty$, then what possible value could $\mu[\mathbf{A} \sqcup \mathbf{B}]$ have?
- The set $\bigsqcup_{n=1}^{\infty} \mathbf{U}_n$ is the same no matter what 'order' in which we perform the disjoint union. Thus, the value of the summation in (2.16) must also be independent of the ordering; this is why the sum must converge absolutely.
- One physical interpretation of a (nonnegative) measure is as the *density* of some distribution of matter. Hence, the measure of a subset $\mathbf{U}$ represents the 'mass' contained in $\mathbf{U}$. Similarly, we can interpret a (signed) measure as a distribution of *electric charge*. Hence, the measure of a subset $\mathbf{U}$ represents the total charge contained in $\mathbf{U}$. For this reason, signed measures are sometimes called **charges**.

Example 76:

(a) Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f \in \mathbf{L}^1(\mathbf{X}, \mu)$. Define measure ν by:

$$\nu[\mathbf{U}] \;=\; \int_{\mathbf{U}} f(x)\; d\mu[x] \qquad \text{for any } \mathbf{U} \in \mathcal{X}.$$

If f is nonnegative, then ν is a 'normal' measure, but if f assumes negative values, then ν is a signed measure.

(b) Let $(\mathbf{X}, \mathcal{X})$ be a measurable space, and let μ and ν be two measures on $\mathbf{X}$. Define $\lambda = \mu - \nu$; that is, for any $\mathbf{U} \in \mathcal{X}$, $\lambda[\mathbf{U}] = \mu[\mathbf{U}] - \nu[\mathbf{U}]$. Then λ is a signed measure. .

We will see that Example(76b) is in fact prototypical.

Definition 77 Positive, Negative & Null sets; Mutually singular

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a (signed) measure space, and let $\mathbf{U} \subset \mathbf{X}$. Then:

- $\mathbf{U}$ *is* **positive** *if* $\mu[\mathbf{U}] \geq 0$, *and* $\mu[\mathbf{V}] \geq 0$ *for all* $\mathbf{V} \subset \mathbf{U}$.
- $\mathbf{U}$ *is* **negative** *if* $\mu[\mathbf{U}] \leq 0$, *and* $\mu[\mathbf{V}] \leq 0$ *for all* $\mathbf{V} \subset \mathbf{U}$.
- $\mathbf{U}$ *is* **null** *if if* $\mu[\mathbf{U}] = 0$, *and* $\mu[\mathbf{V}] = 0$ *for all* $\mathbf{V} \subset \mathbf{U}$.

If μ and ν are two measures on $\mathcal{X}$, we say μ and ν are **mutually singular**, *and write "$\mu \perp \nu$", if there exist disjoint subsets $\mathbf{U}, \mathbf{V} \subset \mathbf{X}$ so that:*

1. $\mathbf{U} \cap \mathbf{V} = \emptyset$ *and* $\mathbf{U} \sqcup \mathbf{V} = \mathbf{X}$.
2. $\mathbf{U}$ *is* ν*-null, and* $\mathbf{V}$ *is* μ*-null.*

Heuristically, $\mathbf{U}$ *contains the 'support' of* μ, *and* $\mathbf{V}$ *contains the 'support' of* ν, *and the two measures have 'disjoint support'.*

Example 78:

Let $(\mathbf{X}, \mathcal{X}, \mu)$, f, and ν be as in Example (76a). Then:

$$\begin{aligned} \mathbf{P} &= \{x \in \mathbf{X} \; ; \; f(x) \geq 0\} \quad \text{is a } \nu\text{-positive set;} \\ \mathbf{N} &= \{x \in \mathbf{X} \; ; \; f(x) < 0\} \quad \text{is a } \nu\text{-negative set;} \\ \text{and } \mathbf{Z} &= \{x \in \mathbf{X} \; ; \; f(x) = 0\} \quad \text{is a } \nu\text{-null set.} \end{aligned}$$

Theorem 79 (Hahn-Jordan Decomposition Theorem)

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space, with signed measure* μ.

1. *There exist disjoint subsets* $\mathbf{P}, \mathbf{N} \subset \mathbf{X}$ *so that*
   - $\mathbf{P} \sqcup \mathbf{N} = \mathbf{X}$.
   - $\mathbf{P}$ *is* μ*-positive and* $\mathbf{N}$ *is* μ*-negative.*
2. *These sets are 'almost' unique: If* $\mathbf{P}', \mathbf{N}'$ *are another pair satisfying these conditions, then* $\mathbf{P} \triangle \mathbf{P}'$ *and* $\mathbf{N} \triangle \mathbf{N}'$ *are* μ*-null.*
3. *There exist unique nonnegative measures* μ^+ *and* μ^- *on* $(\mathbf{X}, \mathcal{X})$ *so that* $\mu^+ \perp \mu^-$ *and* $\mu = \mu^+ - \mu^-$.

Proof: Assume without loss of generality that $\mu[\mathbf{U}] < \infty$ for all $\mathbf{U}$ (otherwise, consider $-\mu$).

Proof of Part 1:

Claim 1: *Let* $\epsilon > 0$. *If* $\mathbf{Y} \subset \mathbf{X}$, *and* $\mu[\mathbf{Y}] > -\infty$, *then there is some* $\mathbf{U} \subset \mathbf{Y}$ *such that* $\mu[\mathbf{U}] \geq \mu[\mathbf{Y}]$, *and so that, for all* $\mathbf{V} \subset \mathbf{U}$, $\mu[\mathbf{V}] \geq -\epsilon$.

Proof: Suppose not. Then there is some $\mathbf{N}_0 \subset \mathbf{Y}$ with $\mu[\mathbf{N}_0] < -\epsilon$ (otherwise $\mathbf{Y}$ itself could be the $\mathbf{U}$ of the claim). Now, let $\mathbf{Y}_1 = \mathbf{Y} \setminus \mathbf{N}_0$. Then

$$\mu[\mathbf{Y}_1] \;=\; \mu[\mathbf{Y}] - \mu[\mathbf{N}_0] \;>\; \mu[\mathbf{Y}] + \epsilon \;>\mu[\mathbf{Y}].$$

Next, there is $\mathbf{N}_1 \subset \mathbf{Y}_1$ so that $\mu[\mathbf{N}_1] < -\epsilon$ (otherwise, $\mathbf{Y}_1$ could be $\mathbf{U}$ since $\mu[\mathbf{Y}_1] > \mu[\mathbf{Y}]$). Let $\mathbf{Y}_2 = \mathbf{Y}_1 \setminus \mathbf{N}_1$. Continuing this way, we get an infinite sequence of disjoint sets $\mathbf{Y}_0, \mathbf{Y}_1, \ldots$ such that $\mu[\mathbf{Y}_n] < -\epsilon$ for all n. Now let $\mathbf{Y}_\infty = \displaystyle\bigsqcup_{n=1}^{\infty} \mathbf{Y}_n$. Then

$$\mu[\mathbf{Y}_\infty] \;=\; \sum_{n=1}^{\infty} \mu[\mathbf{Y}_n] \;\leq\; \sum_{n=1}^{\infty} (-\epsilon) \;=\; -\infty.$$

But then $\mu[\mathbf{Y} \setminus \mathbf{Y}_\infty] = \mu[\mathbf{Y}] + \infty = \infty$, contradicting our starting assumption. □ `[Claim 1]`

Claim 2:

(a) If $\mathbf{U}_1 \supset \mathbf{U}_2 \supset \mathbf{U}_3 \supset \ldots$ *and* $\mathbf{U} = \bigcap_{n=1}^{\infty} \mathbf{U}_n$*, then* $\mu[\mathbf{U}] = \lim_{n\to\infty} \mu[\mathbf{U}_n]$.

(b) If $\mathbf{U}_1 \subset \mathbf{U}_2 \subset \mathbf{U}_3 \subset \ldots$ *and* $\mathbf{U} = \bigcup_{n=1}^{\infty} \mathbf{U}_n$*, then* $\mu[\mathbf{U}] = \lim_{n\to\infty} \mu[\mathbf{U}_n]$.

Proof: <u>**Exercise 76**</u> Hint: The proof is similar to the one for nonnegative measures. □ `[Claim 2]`

Claim 3: *If* $\mathbf{Y} \subset \mathbf{X}$*, and* $\mu[\mathbf{Y}] > -\infty$*, then there is a* μ*-positive subset* $\mathbf{P} \subset \mathbf{Y}$ *with* $\mu[\mathbf{P}] \geq \mu[\mathbf{Y}]$.

Proof: By Claim 1, there is $\mathbf{U}_1 \subset \mathbf{Y}$ so that $\mu[\mathbf{U}_1] \geq \mu[\mathbf{Y}]$ and $\mu[\mathbf{V}] \geq -1$ for all $\mathbf{V} \subset \mathbf{U}_1$. Next, by Claim 1, there is $\mathbf{U}_2 \subset \mathbf{U}_1$ so that $\mu[\mathbf{U}_2] \geq \mu[\mathbf{U}_1] \geq \mu[\mathbf{Y}]$, and $\mu[\mathbf{V}] \geq \frac{-1}{2}$ for all $\mathbf{V} \subset \mathbf{U}_2$.

Inductively, we build a sequence $\mathbf{U}_1 \supset \mathbf{U}_2 \supset \mathbf{U}_3 \supset \ldots$ such that, for all $n \in \mathbb{N}$, $\mu[\mathbf{U}_n] \geq \mu[\mathbf{Y}]$, and $\mu[\mathbf{V}] \geq \frac{-1}{n}$ for all $\mathbf{V} \subset \mathbf{U}_n$. Now, let $\mathbf{P} = \bigcap_{n=1}^{\infty} \mathbf{U}_n$.

Claim 3.1: $\mathbf{P}$ *is* μ*-positive.*

Proof: Let $\mathbf{V} \subset \mathbf{P}$. For every $n \in \mathbb{N}$, $\mathbf{V} \subset \mathbf{U}_n$, so that $\mu[\mathbf{V}] \geq \frac{-1}{n}$. Thus, $\mu[\mathbf{V}] \geq 0$. □ `[Claim 3.1]`

Also, by Claim 2(a), $\mu[\mathbf{P}] = \lim_{n\to\infty} \mu[\mathbf{U}_n] \geq \mu[\mathbf{Y}]$. .... □ `[Claim 3]`

Claim 4: *If* $\mathbf{P}_1, \mathbf{P}_2, \ldots$ *are all* μ*-positive sets, then* $\mathbf{P} = \bigcup_{n=1}^{\infty} \mathbf{P}_n$ *is also* μ*-positive.*

Proof: <u>**Exercise 77**</u> .... □ `[Claim 4]`

Let $S = \sup_{\mathbf{Y}\in\mathcal{X}} \mu[\mathbf{Y}]$, and let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of subsets such that $\mu[\mathbf{Y}_n] \xrightarrow[n\to\infty]{} S$. By Claim 3, we have μ-positive subsets $\mathbf{P}_n \subset \mathbf{Y}_n$ so that $\mu[\mathbf{P}_n] \geq \mu[\mathbf{Y}_n]$; hence $\mu[\mathbf{P}_n] \xrightarrow[n\to\infty]{} S$ also. By Claim 4, $\mathbf{P} = \bigcup_{n=1}^{\infty} \mathbf{P}_n$ is also μ-positive.

Claim 5: $\mu[\mathbf{P}] = S$.

Proof: First, note that $\mu[\mathbf{P}] \leq S$, by definition of S. But by Claim 2(b), $\mu[\mathbf{P}] = \lim_{N\to\infty} \mu\left[\bigcup_{n=1}^{N} \mathbf{P}_n\right] \geq \lim_{n\to\infty} \mu\left[\mathbf{P}_n\right] = S$. Thus, $\mu[\mathbf{P}] \geq S$ also. .... □ `[Claim 5]`

Now, let $\mathbf{N} = \mathbf{X} \setminus \mathbf{P}$.

Claim 6: $\mathbf{N}$ *is* μ*-negative.*

Proof: Suppose not. Then there is $\mathbf{U} \subset \mathbf{N}$ with $\mu[\mathbf{U}] > 0$. But $\mathbf{U}$ is disjoint from $\mathbf{P}$, and thus $\mu[\mathbf{P} \sqcup \mathbf{U}] > \mu[\mathbf{U}] = S$, contradicting the supremality of S. .... □ `[Claim 6]`

Thus, $\mathbf{X} = \mathbf{P} \sqcup \mathbf{N}$ is the decomposition we seek.

Proof of Part 2: Suppose $\mathbf{X} = \mathbf{P}' \sqcup \mathbf{N}'$ was another such decomposition.

Claim 7: $\mathbf{P}' \setminus \mathbf{P}$ *is* μ*-null.*

Proof: Suppose not. Then there is some $\mathbf{U} \subset (\mathbf{P}' \setminus \mathbf{P})$ with $\mu[\mathbf{U}] > 0$. But $\mathbf{U}$ is disjoint from $\mathbf{P}$, so that $\mu[\mathbf{U} \sqcup \mathbf{P}] > \mu[\mathbf{P}] = S$, contradicting the supremality of S. □ `[Claim 7]`

Likewise, $\mathbf{P} \setminus \mathbf{P}'$ is μ-null, so that $\mathbf{P} \triangle \mathbf{P}' = (\mathbf{P} \setminus \mathbf{P}') \sqcup (\mathbf{P}' \setminus \mathbf{P})$ is μ-null. A similar proof works for $\mathbf{N}$ and $\mathbf{N}'$.

Proof of Part 2: <u>**Exercise 78**</u> . □

We refer to $\mathbf{X} = \mathbf{P} \sqcup \mathbf{N}$ as a **Hahn-Jordan decomposition** of the (signed) measure space $(\mathbf{X}, \mathcal{X}, \mu)$, and $\mu = \mu^+ - \mu^-$ as a **Hahn-Jordan decomposition** of the signed measure μ. The **total variation** of μ is the (positive) measure $|\mu|$ defined $|\mu| = \mu^+ + \mu^-$.

Example 80:

(a) Suppose μ, f, and ν are as Example (76a), and define $\mathbf{P}$ and $\mathbf{N}$ as in Example 78. Then $\mathbf{X} = \mathbf{P} \sqcup \mathbf{N}$, and this is a Hahn-Jordan decomposition of $(\mathbf{X}, \mathcal{X}, \nu)$. Now, define

$$f^+(x) \;=\; \mathbb{1}_{\mathbf{P}}(x) \cdot f(x) \qquad \text{and} \qquad f^-(x) \;=\; \mathbb{1}_{\mathbf{N}}(x) \cdot f(x)$$

and, for any $\mathbf{U} \in \mathcal{X}$, let

$$\nu^+[\mathbf{U}] \;=\; \int_{\mathbf{U}} f^+(x)\ d\mu[x] \qquad \text{and} \qquad \nu^-[\mathbf{U}] \;=\; \int_{\mathbf{U}} f^-(x)\ d\mu[x]$$

Then $\nu = \nu^+ - \nu^-$ is a Hahn-Jordan decomposition for ν. The total variation of ν is the measure $|\nu|$ defined

$$|\nu|[\mathbf{U}] \;=\; \int_{\mathbf{U}} |f(x)|\ d\mu[x] \qquad \text{for all } \mathbf{U} \in \mathcal{X}.$$

(b) $\lambda = \mu - \nu$ as Example (76b). If $\mu \perp \nu$, then the Hahn-Jordan decomposition for λ is:

$$\lambda^+ \;=\; \mu \qquad \text{and} \qquad \lambda^- \;=\; \nu.$$

However, if μ and ν have overlapping support, then they do *not* yield the Hahn-Jordan decomposition

Definition 81 Complex-valued Measure

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space. A* **complex-valued measure** *on* $(\mathbf{X}, \mathcal{X})$ *is a function* $\mu : \mathcal{X} \longrightarrow \mathbb{C}$ *so that* $\mathsf{re}[\mu]$ *and* $\mathsf{im}[\mu]$ *are both strictly finite signed measures.*

Remark: We need $\mathsf{re}[\mu]$ and $\mathsf{im}[\mu]$ to be *finite*, because otherwise we might end up with nonsensical results like "$\mu[\mathbf{A}] = \infty + 3 \cdot i$".

Example 82:

(a) Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f \in \mathbf{L}^1(\mathbf{X}, \mu;\ \mathbb{C})$. Define measure ν by:

$$\nu[\mathbf{U}] \;=\; \int_{\mathbf{U}} f(x)\, d\mu[x] \qquad \text{for any } \mathbf{U} \in \mathcal{X}.$$

Then ν is a complex-valued measure.

(b) Let $(\mathbf{X}, \mathcal{X})$ be a measurable space, and let λ, μ, ν, ρ be four finite measures on $\mathbf{X}$. Define $\eta = (\lambda - \mu) + \mathbf{i}(\nu - \rho)$; that is,

$$\eta[\mathbf{U}] \;=\; \lambda[\mathbf{U}] \;-\; \mu[\mathbf{U}] \;+\; \mathbf{i}\nu[\mathbf{U}] \;-\; \mathbf{i}\rho[\mathbf{U}] \qquad \text{for any } \mathbf{U} \in \mathcal{X}.$$

Then ν is a complex-valued measure.

In a similar vein, we can define "$\mathbb{V}$-valued measures" where $\mathbb{V}$ is any topological vector space. For example:

- **Bochner Integrals** are measures taking their values on on a Banach space.
- **Spectral Measures** are measures whose values range over the set of bounded linear operators on a Hilbert space.

2.4 The Space of Measures

2.4(a) Introduction

Prerequisites: §2.3

Definition 83 The Space of Measures

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space. The* **space of measures** *over* $(\mathbf{X}, \mathcal{X})$ *is the complex vector space*

$$\mathcal{M}(\mathbf{X}, \mathcal{X}) \;:=\; \{\mu : \mathcal{X} \longrightarrow \mathbb{C}\ ;\ \mu \ \text{ is a (complex-valued) measure on } \mathcal{X}\}$$

Exercise 79 Show that this is a vector space.

Example 84: Finite State Space

If $\mathbf{X} = \{x_1, \ldots, x_N\}$ is a finite set, and $\mathcal{X} := \mathcal{P}(\mathbf{X})$, then $\mathcal{M}(\mathbf{X}, \mathcal{X})$ is canonically isomorphic to $\mathbb{R}^{\mathbf{X}} = \mathbb{R}^N$. (**Exercise 80**)

Definition 85 Probability Simplex

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space. The* **probability simplex** *on* $(\mathbf{X}, \mathcal{X})$ *is defined:*

$$\Delta(\mathbf{X}, \mathcal{X}) \quad := \quad \{\mu \in \mathcal{M}(\mathbf{X}, \mathcal{X}) \ ; \ \mu \ \text{ is non negative and } \ \mu[\mathbf{X}] = 1. \ \}$$

Example 86: Finite State Space

If $\mathbf{X} = \{x_1, \ldots, x_N\}$ is a finite set, and $\mathcal{X} := \mathcal{P}(\mathbf{X})$, then $\Delta(\mathbf{X}, \mathcal{X})$ is canonically isomorphic to the positive simplex $\Delta^N \ = \ \left\{ \mathbf{x} \in [0,1]^N \ ; \ \sum_{n=1}^{N} x_n \ = \ 1 \right\}$. (**Exercise 81**)

Definition 87 The Space of Radon Measures

If $\mathbf{X}$ *is a topological space, and* $\mathcal{X}$ *is the Borel sigma algebra. The* **space of Radon Measures** *on* $\mathbf{X}$ *is:*

$$\mathcal{M}_R(\mathbf{X}, \mathcal{X}) \quad := \quad \{\mu : \mathcal{X} \longrightarrow \mathbb{C} \ ; \ \mu \ \text{ is a Radon measure on } \ \mathcal{X}\}$$

Exercise 82 Show that $\mathcal{M}_R$ is a linear subspace of $\mathcal{M}$.

Definition 88 Radon Probability Simplex

Let $\mathbf{X}$ *be a topological space, and* $\mathcal{X}$ *be the* **Borel** *sigma algebra. Then the* **Radon probability simplex** *on* $(\mathbf{X}, \mathcal{X})$ *is defined:*

$$\Delta_R(\mathbf{X}, \mathcal{X}) \quad := \quad \{\mu \in \mathcal{M}_R(\mathbf{X}, \mathcal{X}) \ ; \ \mu \ \text{ is nonnegative and } \ \mu[\mathbf{X}] = 1.\}$$

2.4(b) The Norm Topology

Prerequisites: §2.4(a),[Banach Space Theory]

Definition 89 Total Variation Norm

Let $(\mathbf{X}, \mathcal{X})$ *be a measurable space, and define*

$$\mathbb{B} \ = \ \{f \in \mathbf{L}(\mathbf{X}, \mathcal{X}) \ ; \ \ |f| \text{ has constant value } 1\}$$

If $\mu \in \mathcal{M}(\mathbf{X}, \mathcal{X})$, *then the* **total variation** *of* μ *is defined:*

$$\|\mu\| \quad := \quad \sup_{f \in \mathbb{B}} \int_{\mathbf{X}} f \ d\mu.$$

Lemma 90 Equivalent Definitions

1. *Other, equivalent definitions of the total variation norm are:*

$$\begin{aligned} \|\mu\| &= \sup\left\{\sum_{P\in\mathcal{P}} |\mu[P]| \;;\; \mathcal{P} \text{ is a finite measurable partition of } \mathbf{X}\right\} \\ \text{and } \|\mu\| &= \sup\left\{\sum_{P\in\mathcal{P}} |\mu[P]| \;;\; \mathcal{P} \text{ is a countable measurable partition of } \mathbf{X}\right\} \end{aligned}$$

2. *If μ is real-valued, then:* $\|\mu\| \;=\; \left|\sup_{\mathbf{U}\in\mathcal{X}} \mu[\mathbf{U}]\right| \;+\; \left|\inf_{\mathbf{U}\in\mathcal{X}} \mu[\mathbf{U}]\right|$.

3. *If μ is nonnegative, then:* $\|\mu\| \;=\; \mu[\mathbf{X}]$.

4. *If μ is nonegative, and ν is absolutely continuous with respect to μ, then* $\|\nu\| \;=\; \left\|\frac{d\nu}{d\mu}\right\|_1 \cdot \|\mu\|$, *where* $\left\|\frac{d\nu}{d\mu}\right\|_1$ *is the $\mathbf{L}^1$-norm of the function* $\frac{d\nu}{d\mu}$ *in the space* $\mathbf{L}^1(\mathbf{X},\mathcal{X},\mu)$.

Proof: <u>**Exercise 83**</u> □

Example 91:

(a) If $\mathbf{X} = \{x_1,\ldots,x_n\}$ is a finite set, then, for any $\mu \in \mathcal{M}(\mathbf{X})$, $\|\mu\| \;=\; \mu\{x_1\}+\ldots+\mu\{x_n\}$.

(b) If μ is a probability measure, then $\|\mu\| = 1$.

Theorem 92 *Under the total variation norm, $\mathcal{M}(\mathbf{X},\mathcal{X})$ is a Banach Space.*

Proof: <u>**Exercise 84**</u> Hint: First show that $\|\bullet\|$ acts as a norm on $\mathcal{M}(\mathbf{X},\mathcal{X})$.

Next, suppose $\{\mu_n\}_{n=1}^{\infty}$ is a Cauchy sequence; we need to show that it converges to an element of $\mathcal{M}(\mathbf{X},\mathcal{X})$.

1. Show that, for any $\mathbf{U} \in \mathcal{X}$, the sequence of complex numbers $\{\mu_n(\mathbf{U})\}_{n=1}^{\infty}$ is Cauchy.
2. Define $\mu(\mathbf{U}) \;=\; \lim_{n\to\infty} \mu_n(\mathbf{U})$. Then $\mu : \mathcal{X} \longrightarrow \mathbb{C}$ is a complex valued measure.
3. $\|\mu\| \;=\; \lim_{n\to\infty} \|\mu_n\|$. □

Theorem 93 *If $(\mathbf{X}_1,\mathcal{X}_1)$ and $(\mathbf{X}_2,\mathcal{X}_2)$ are measurable spaces, and $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ is a measurable map, then the push-forward map*

$$f^* : \mathcal{M}(\mathbf{X}_1,\mathcal{X}_1) \longrightarrow \mathcal{M}(\mathbf{X}_2,\mathcal{X}_2)$$

is a linear isometric embedding of the Banach space $\mathcal{M}(\mathbf{X}_1,\mathcal{X}_1)$ into Banach space $\mathcal{M}(\mathbf{X}_2,\mathcal{X}_2)$.

Proof: <u>**Exercise 85**</u> □

2.4(c) The Weak Topology

Prerequisites: §2.4(a),[Weak Topologies and Locally Convex Spaces]

The elements of $\mathcal{M}(\mathbf{X}, \mathcal{X})$ act as *linear functionals* on other spaces; this induces corresponding *weak topologies* on $\mathcal{M}(\mathbf{X}, \mathcal{X})$.

Definition 94 Weak Topology

Let $(\mathbf{X}, \mathcal{X})$ be a measurable space. The **weak topology** *on $\mathcal{M}(\mathbf{X}, \mathcal{X})$ is the topology defined by the following convergence condition. If $\{\mu_\lambda\}_{\lambda\in\Lambda}$ is some net of measures, then*

$$\left(\mu_\lambda \xrightarrow[\lambda\in\Lambda]{} \mu \right) \iff \left(\text{For every } \mathbf{U} \in \mathcal{X},\ \ \mu_\lambda[\mathbf{U}] \xrightarrow[\lambda\in\Lambda]{} \mu[\mathbf{U}] \right)$$

Theorem 95 *With respect to the weak topology, $\mathcal{M}(\mathbf{X}, \mathcal{X})$ is a locally convex space.*

Proof: <u>**Exercise 86**</u> □

There is another (slightly stronger) weak topology, defined on the set of Radon measures. Let $\mathbf{X}$ be a topological space; recall from § 4.1(a) on page 95 that $\mathcal{C}_0(\mathbf{X})$ is the vector space of continuous functions $f : \mathbf{X} \longrightarrow \mathbb{R}$ which *vanish at infinity*, which is a Banach space when equipped with the uniform norm $\|f\|_{\mathrm{u}} = \sup_{x\in\mathbf{X}} |f(x)|$. Let $\mathcal{C}_0^*(\mathbf{X})$ be the space of **continuous linear functionals** on $\mathcal{C}_0(\mathbf{X})$; that is, linear functions $\mathcal{C}_0(\mathbf{X}) \longrightarrow \mathbb{R}$ which are continuous relative to $\|\bullet\|_{\mathrm{u}}$.

Definition 96 Vague Topology

Let $\mathbf{X}$ be a topological space, and $\mathcal{X}$ the Borel sigma-algebra. The **vague topology** *on $\mathcal{M}_R(\mathbf{X}, \mathcal{X})$ is the topology defined by the following convergence condition. If $\{\mu_\lambda\}_{\lambda\in\Lambda}$ is some net of Radon measures, then*

$$\left(\mu_\lambda \xrightarrow[\lambda\in\Lambda]{} \mu \right) \iff \left(\text{For every } f \in \mathcal{C}_0(\mathbf{X}),\quad \int_{\mathbf{X}} f\ d\mu_\lambda \xrightarrow[\lambda\in\Lambda]{} \int_{\mathbf{X}} f\ d\mu \right)$$

Theorem 97 *With respect to the vague topology, $\mathcal{M}_R(\mathbf{X}, \mathcal{X})$ is a locally convex space.*

Proof: <u>**Exercise 87**</u> □

Theorem 98 (Second Riesz Representation Theorem)

Let $\mathbf{X}$ is a locally compact Hausdorff space with Borel sigma-algebra $\mathcal{X}$. Define a map $\mathcal{M}(\mathbf{X}, \mathcal{X}) \longrightarrow \mathcal{C}_0^*$ so that, for any $\mu \in \mathcal{M}(\mathbf{X}, \mathcal{X})$ and $f \in \mathcal{C}_0$, $\mu[f] = \int_{\mathbf{X}} f\ d\mu$.

This map is an isomorphism of $\mathcal{M}(\mathbf{X}, \mathcal{X})$ and $\mathcal{C}_0^*$ as locally convex vector spaces.

□

2.5 Disintegrations of Measures

Prerequisites: §5.2(a),§2.4(a)

Integration is process whereby a multitude of separate and distinct quantities are combined into a unified whole. *Disintegration* is the opposite: a process whereby a whole is shattered into many separate components. In measure theory, *disintegration* is a process whereby we can decompose a measure into many constituent fibres.

Definition 99 Disintegration of Measures

Let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be measure spaces. A **disintegration** *of ξ with respect to Υ is a map $\mathbf{Y} \ni y \mapsto \xi_y \in \mathcal{M}(\mathbf{X}, \mathcal{X})$ such that:*

(D1) *The map $(y \mapsto \xi_y)$ is* **measurable:**, *for any fixed measurable subset $\mathbf{U} \subset \mathbf{X}$, the function $\mathbf{Y} \ni y \mapsto \xi_y[\mathbf{U}] \in \mathbb{C}$ is measurable.*

(D2) *The measure ξ is the* **integral** *of the measures $\{\xi_y\}_{y \in \mathbf{Y}}$ with respect to Υ, in the sense that, for any measurable subset $\mathbf{U} \subset \mathbf{X}$, we have:*

$$\xi[\mathbf{U}] \quad = \quad \int_Y \xi_y[\mathbf{U}] \;\; d\Upsilon[y]$$

Formally, we write this: "$\xi \quad = \quad \int_Y \xi_y \;\; d\Upsilon[y]$".

Example 100: The Cube

Let $\mathbf{X} = \mathbb{I}^3$ be the three-dimensional unit cube with Lebesgue measure $\xi = \lambda^3$; let $\mathbf{Y} = \mathbb{I}^2$ be the unit square with Lebesgue measure $\Upsilon = \lambda^2$. Let $P : \mathbf{X} \longrightarrow \mathbf{Y}$ be the projection map: $P(x_1, x_2, x_3) = (x_1, x_2)$. Fix $y \in \mathbb{I}^2$, and let $\mathbf{F}_y = P^{-1}\{y\} = \{y\} \times \mathbb{I}$ be the **fibre** over y (see Figure 2.8A.)

Fibre $\mathbf{F}_y$ is just a copy of the unit interval, so let ξ_x be a 'copy' of the Lebesgue measure λ on $\mathbf{F}_y$. In other words: for any $\mathbf{U} \subset \mathbb{I}^3$, let $\mathbf{U}_y \;=\; \{x \in \mathbb{I} \;;\; (y, x) \in \mathbf{U}\}$, and then define:

$$\xi_y[\mathbf{U}] \;=\; \lambda[\mathbf{U}_y] \qquad \text{(see Figure 2.8B.)}$$

It then follows that $\lambda^3 \;=\; \int_{\mathbb{I}^2} \xi_y \; d\lambda^2[y]$.

<u>**Exercise 88**</u> Verify this, by checking that properties **(D1)** and **(D2)** hold.

Example 101: The Fubini-Tonelli Theorem

Let $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ and $(\mathbf{Z}, \mathcal{Z}, \zeta)$ be measure spaces, with product measure space $(\mathbf{X}, \mathcal{X}, \xi)$. Let $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \xi)$, and define measure Υ on $\mathcal{X}$ by:

$$\forall \; \mathbf{U} \in \mathcal{X}, \quad \Upsilon\,[\mathbf{U}] \;\; := \;\; \int_{\mathbf{U}} f \; d\xi.$$

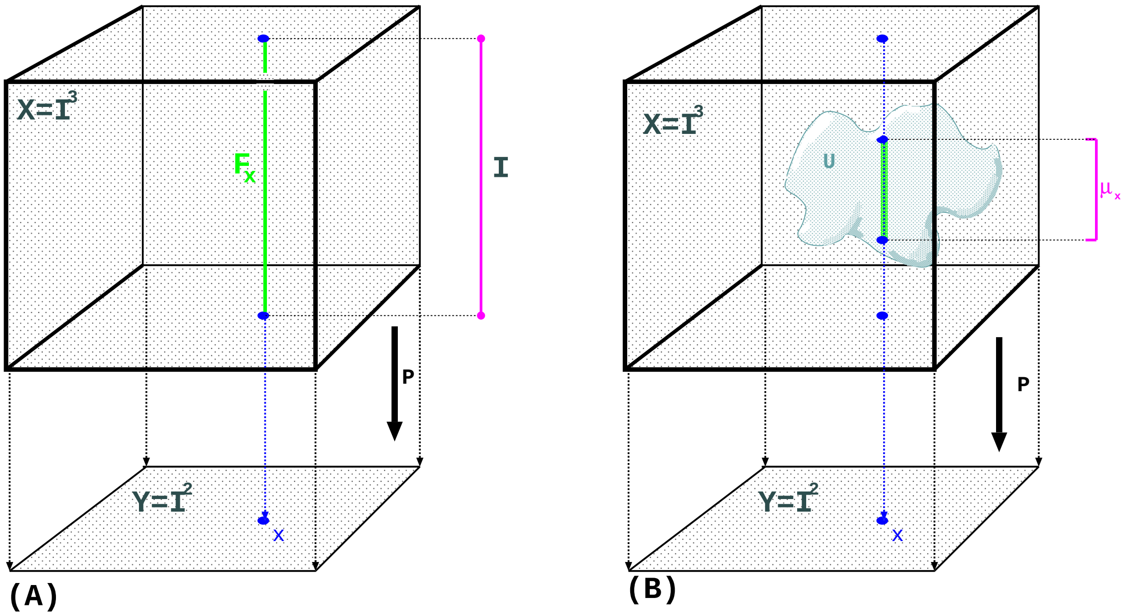


Figure 2.8: The fibre measure over a point.

For all $y \in \mathbf{Y}$, let $\mathbf{F}_y = P^{-1}\{y\} = \{y\} \times \mathbf{Z}$ be the **fibre** over y. For any $\mathbf{U} \subset \mathbf{X}$, define $\mathbf{U}_y = \{z \in \mathbf{Z}\,;\ (y,z) \in \mathbf{U}\}$. Define the **fibre measure** ξ_y by:

$$\xi_y[\mathbf{U}] \quad = \quad \int_{\mathbf{U}_z} f(y,z) \ \ d\zeta[z].$$

Then ξ has disintigration: $\xi \quad = \quad \int_{\mathbf{Y}} \xi_y \ d\Upsilon[y].$

This is really just a restatement of the **Fubini-Tonelli Theorem**.

Exercise 89 Verify this disintegration, by checking that properties **(D1)** and **(D2)** hold.

Exercise 90 Verify that this is equivalent to the Fubini-Tonelli theorem.

These examples illustrate the most common disintegrations of measures: decompositions into 'fibres' over some projection map.

Theorem 102 *Let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be measure spaces, and $p : (\mathbf{X}, \mathcal{X}, \xi) \longrightarrow (\mathbf{Y}, \mathcal{Y}, \Upsilon)$ a measure-preserving map. Then p induces a disintegration of ξ over Υ:*

$$\xi \quad = \quad \int_{\mathbf{Y}} \xi_y \ \ d\Upsilon[y],$$

where, for all $y \in \mathbf{Y}$, the fibre measure ξ_y has its support confined to the fibre $\mathbf{F}_y = p^{-1}\{y\}$.

Proof: Let $\widetilde{\mathcal{Y}} := p^{-1}(\mathcal{Y})$, a sigma-subalgebra of $\mathcal{X}$.

For a fixed subset $\mathbf{U} \in \mathcal{X}$, consider the conditional measure

$$\xi\left[\mathbf{U} \;\langle\!\langle\; \widetilde{\mathcal{Y}}\right] \;:=\; \mathbb{E}_{\widetilde{\mathcal{Y}}}\left[\mathbb{1}_{\mathbf{U}}\right]$$

of $\mathbf{U}$ with respect to $\widetilde{\mathcal{Y}}$. This is a $\widetilde{\mathcal{Y}}$-measurable function, therefore constant on each fibre of the map p. Thus, we can treat it as a measurable function on $\mathbf{Y}$. Specifically, for any fixed $y \in \mathbf{Y}$, define

$$\xi_y[\mathbf{U}] \;:=\; \xi\left[\mathbf{U} \;\langle\!\langle\; \widetilde{\mathcal{Y}}\right](x), \qquad \text{where } x \in p^{-1}\{y\} \text{ is arbitrary.}$$

Claim 1: *For fixed $y \in \mathbf{Y}$, the map $\xi_y : \mathcal{X} \ni \mathbf{U} \mapsto \xi_y[\mathbf{U}] \in \mathbb{R}$ is a measure on $\mathbf{X}$, and is supported entirely on the fibre $\mathbf{F}_y$. In other words, for any $\mathbf{U} \in \mathcal{X}$, if $\mathbf{U} \cap \mathbf{F}_y = \emptyset$, then $\xi_y[\mathbf{U}] = 0$.*

Proof: **Exercise 91** .... □ [Claim 1]

Claim 2: $\xi \;=\; \int_{\mathbf{Y}} \xi_y \; d\Upsilon[y]$.

Proof: **Exercise 92** .... □ [Claim 2]

□

Definition 103 Fibre Spaces, Fibre Sets

Let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be measure spaces, and $p : (\mathbf{X}, \mathcal{X}, \xi) \longrightarrow (\mathbf{Y}, \mathcal{Y}, \Upsilon)$ a measure-preserving map.

Fix $y \in \mathbf{Y}$, and let ξ_y be the **fibre measure** *of Υ over y, so that we have the disintegration*

$$\xi \;=\; \int_{\mathbf{Y}} \xi_y \; d\Upsilon[y],$$

Let $\mathbf{F}_y := p^{-1}\{y\}$ be the p-fibre over y, and define sigma-algebra

$$\mathcal{F}_y \;=\; \{\mathbf{U} \cap \mathbf{F}_y \;;\; \mathbf{U} \in \mathcal{X}\}$$

*Then $(\mathbf{F}_y, \mathcal{F}_y, \xi_y)$ is a measure space (**Exercise 93**), and is called the* **fibre space** *over the point y.*

For any measurable subset $\mathbf{U} \subset \mathbf{X}$, let $\mathbf{U}_y := \mathbf{U} \cap \mathbf{F}_y$ be the corresponding element of $\mathcal{F}_y$; this is called the **fibre** *of $\mathbf{U}$ over y.*

*If $f \in \mathbf{L}^p(\mathbf{X}, \mathcal{X}, \xi)$, then let $f_y = f\big|_{\mathbf{F}_y}$. Then f_y is well-defined for $\forall_\Upsilon\; y \in \mathbf{Y}$, and $f_y \in \mathbf{L}^p(\mathbf{F}_y, \mathcal{F}_y, \xi_y)$ (**Exercise 94**). We write, formally,*

$$f \;=\; \int_{\mathbf{Y}} f_y \; d\Upsilon[y]. \qquad \textit{(see Figure 2.9 on the next page)}$$

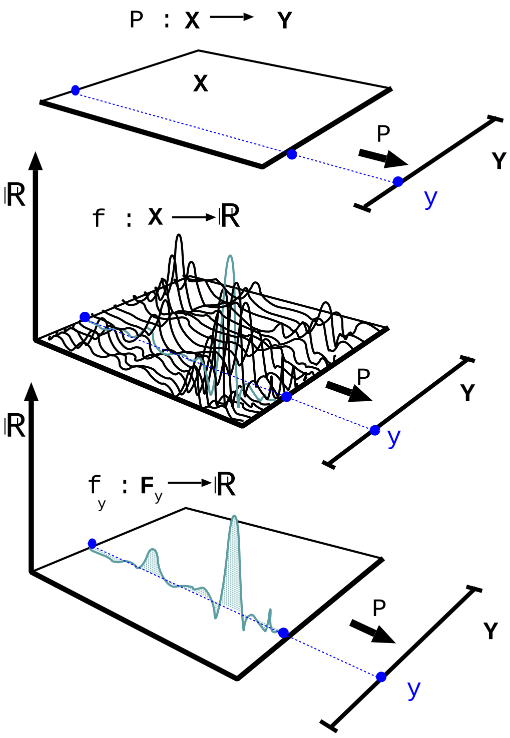


Figure 2.9: The **fibre** of the function f over the point y.

For example:

- If $\mathbf{U} \subset \mathbf{X}$, then $\mathbb{1}_{(\mathbf{U}_y)} \quad = \quad (\mathbb{1}_{\mathbf{U}})_y$. (**Exercise 95**)
- If $f \in \mathbf{L}^p(\mathbf{X}, \mathcal{X}, \xi)$, then $\|f\|_p \quad = \quad \left(\int_{\mathbf{Y}} \|f_y\|_p^{\ p} \ d\Upsilon[y] \right)^{1/p}$. (**Exercise 96**)

3 Integration Theory

3.1 Construction of the Lebesgue Integral

Prerequisites: §1.1, §1.3(a)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space. If $f : \mathbf{X} \longrightarrow \mathbb{C}$ is a measurable function, we might want to define the *integral* of f, over $\mathbf{X}$, relative to μ. Our goal is to generalize the classic Riemann

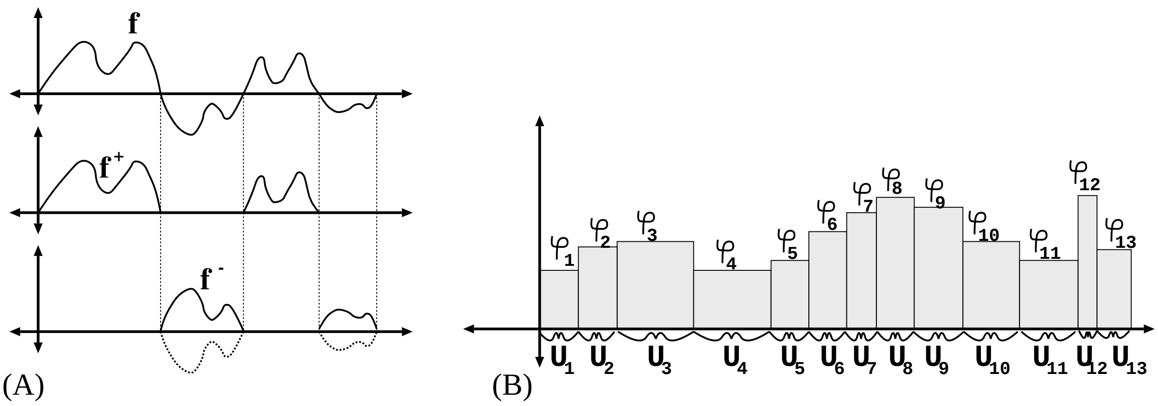


Figure 3.1: **(A)** $f \;=\; f^+ - f^-$; **(B)** A simple function.

integral to a broader class of domains and functions. Let us denote this (as yet hypothetical) integral by $\int_{\mathbf{X}} f \; d\mu$. It should have at least three properties:

Compatibility with μ: If $f = \mathbb{1}_{\mathbf{U}}$ is an indicator function, then $\int_{\mathbf{X}} \mathbb{1}_{\mathbf{U}} \; d\mu \;=\; \mu[\mathbf{U}]$.

Linearity: $\int_{\mathbf{X}} (c \cdot f + g) \; d\mu \;=\; c \cdot \int_{\mathbf{X}} f \; d\mu + \int_{\mathbf{X}} g \; d\mu$, for any functions $f, g : \mathbf{X} \longrightarrow \mathbb{C}$ and $c \in \mathbb{C}$.

Continuity: If $f_1, f_2, \ldots$ is a sequence of functions converging to f (in some sense), then $\lim_{n\to\infty} \int_{\mathbf{X}} f_n \; d\mu \;=\; \int_{\mathbf{X}} f \; d\mu$.

These properties suggest a strategy for defining the integral:

(1) Use '**Compatibility**' and '**Linearity**' to define the integral of any sum of characteristic functions —that is, any *simple* function.

(2) Use some kind of '**Continuity**' to define the integral of an arbitrary function by approximating it with simple functions.

There are two ways to realize this strategy, which we describe in §3.1(b) and §3.1(d). The reader need only be familiar with §3.1(b); the approach in §3.1(d) is developed only for interest. For either approach, we need some facts about simple functions, developed in §3.1(a). After constructing the integral, we will establish some of it's basic properties in §3.1(c). We will then develop important limit theorems in §3.2.

Preliminaries: If $f : \mathbf{X} \longrightarrow [-\infty, \infty]$, then we can write f as a difference of two nonnegative functions:

$f \;=\; f^+ - f^-$, where $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = -\min\{f(x), 0\}$ (see Figure 3.1A)

If $f : \mathbf{X} \longrightarrow \mathbb{C}$ then we use the decomposition: $f \;=\; f_{\mathrm{re}} + \mathbf{i} \cdot f_{\mathrm{im}}$.

Lemma 104 *If $f : \mathbf{X} \longrightarrow [-\infty, \infty]$ is measurable, then f^+ and f^- are measurable and nonnegative. If $f : \mathbf{X} \longrightarrow \mathbb{C}$ is measurable, then f_{re}^+, f_{re}^-, f_{im}^+ and f_{im}^- are measurable and nonnegative.*

Proof: Exercise 97 □

3.1(a) Simple Functions

Definition 105 Simple Function

A **simple function** *is a finite or countable linear combination of characteristic functions. That is, $\Phi : \mathbf{X} \longrightarrow \mathbb{C}$ is* **simple** *if it has the form:*

$$\Phi(x) \;=\; \sum_{n=1}^{\infty} \varphi_n \cdot \mathbb{1}_{\mathbf{U}_n}(x) \qquad \text{(see Figure 3.1B)}$$

where $\varphi_n \in \mathbb{C}$ are constants and $\mathbf{U}_n \subset \mathbf{X}$ are measurable subsets, for $n \in \mathbb{N}$. (A finite *linear combination is the special case when $\mathbf{U}_n = \emptyset$ for almost all $n \in \mathbb{N}$).*

A **repartition** of $\Phi(x)$ is a different collection of subsets $\{\widetilde{\mathbf{U}}_k\}_{k=1}^{\infty}$ and constants $\{\widetilde{\varphi}_k\}_{k=1}^{\infty}$ so that $\Phi(x) \;=\; \sum_{k=1}^{\infty} \widetilde{\varphi}_k \cdot \mathbb{1}_{\widetilde{\mathbf{U}}_k}(x)$ for all $x \in \mathbf{X}$ (see Figure 3.2A).

If $\Psi \;=\; \sum_{n=1}^{\infty} \psi_n \cdot \mathbb{1}_{\mathbf{V}_n}$ is another simple function, we say Φ and Ψ are **compatible** if $\mathbf{U}_n = \mathbf{V}_n$ for all n.

Lemma 106 *Φ and Ψ can always be repartitioned to be compatible; ie. there exists a collection of subsets $\{\mathbf{W}_k\}_{k=1}^{\infty}$ and constants $\{\widetilde{\varphi}_k\}_{k=1}^{\infty}$ and $\{\widetilde{\psi}_k\}_{k=1}^{\infty}$ such that $\Phi = \sum_{k=1}^{\infty} \widetilde{\varphi}_k \cdot \mathbb{1}_{\mathbf{W}_k}$ and $\Psi = \sum_{k=1}^{\infty} \widetilde{\psi}_k \cdot \mathbb{1}_{\mathbf{W}_k}$.*

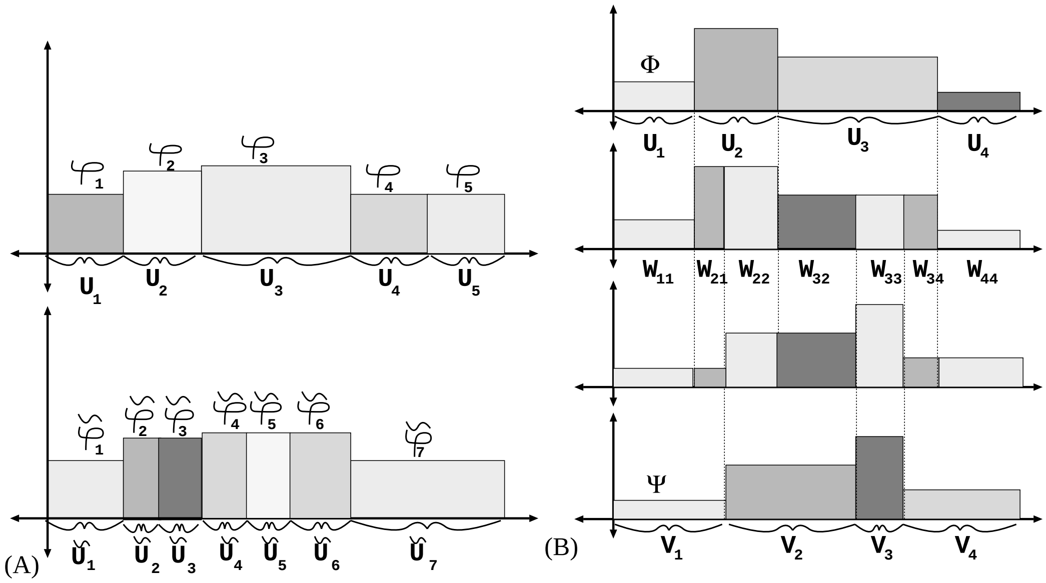


Figure 3.2: **(A)** Repartitioning a simple function. **(B)** Repartitioning two simple functions to be compatible.

Proof: (see Figure 3.2B) For all $n, m \in \mathbb{N}$, define $\mathbf{W}_{n,m} = \mathbf{U}_n \cap \mathbf{V}_m$ (possibly empty), and define $\widetilde{\varphi}_{n,m} = \varphi_n$ and $\widetilde{\psi}_{n,m} = \psi_m$. Now identify $\mathbb{N} \cong \mathbb{N} \times \mathbb{N}$ in some way, thereby reindexing $\mathbf{W}_{n,m}$, $\widetilde{\varphi}_{n,m}$ and $\widetilde{\psi}_{n,m}$ with respect to $k \in \mathbb{N}$. □

Lemma 107 *If Φ and Ψ are simple functions, then so is $\Phi + \Psi$.*

Proof: This follows from Lemma 106. □

Lemma 108 (Approximation by Simple Functions)

Let $f : \mathbf{X} \longrightarrow \mathbb{C}$ be measurable. There is a sequence of simple functions $\{\Phi_n\}_{n=1}^{\infty}$ converging uniformly to f. That is:

- *For all $x \in \mathbf{X}$, $\lim_{n\to\infty} \Phi_n(x) = f(x)$, and furthermore, $\lim_{n\to\infty} \sup_{x \in \mathbf{X}} \left|\Phi_n(x) - f(x)\right| = 0$.*
- *Furthermore, if f is real-valued and nonnegative, we can assume that for all $x \in \mathbf{X}$, $\Phi_1(x) \leq \Phi_2(x) \leq \ldots \leq f(x)$.*

Proof: **Case 1:** (*f is is real-valued*) The construction is illustrated in Figure 3.3. For all $n \in \mathbb{N}$, and all $z \in \mathbb{Z}$, define $\mathbf{U}_n^z = \left\{x \in \mathbf{X} \,;\, f(x) \in \left[\frac{z}{2^n}, \frac{z+1}{2^n}\right)\right\}$, and then define $\Phi_n(x) = \sum_{z=-\infty}^{\infty} \frac{z}{2^n} \cdot \mathbb{1}_{\mathbf{U}_n^z}$. Thus, for all $x \in \mathbf{X}$, $\left|\Phi_n(x) - f(x)\right| < \frac{1}{2^n}$.

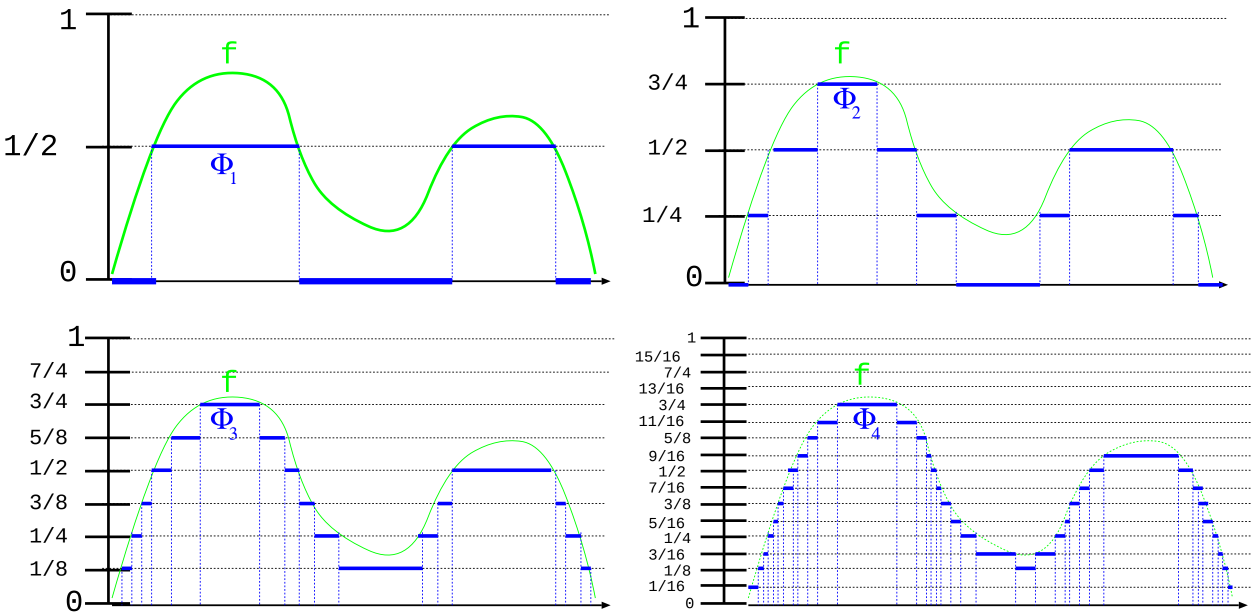


Figure 3.3: Approximating f with simple functions.

Also, if f is nonnegative, observe that $\Phi_1 \leq \Phi_2 \leq \ldots \leq f$.

Case 2: (*f is complex-valued*) Apply Case 2 to find $\{\Phi_n^{\mathrm{re}}\}_{n=1}^{\infty}$ and $\{\Phi_n^{\mathrm{im}}\}_{n=1}^{\infty}$ converging uniformly to f_{re} and f_{im}. Now define $\Phi_n = \Phi_n^{\mathrm{re}} + \mathbf{i} \cdot \Phi_n^{\mathrm{im}}$; then Φ_n is simple by Lemma 107, and $\{\Phi_n\}_{N=1}^{\infty}$ converges uniformly to f. □

3.1(b) Definition of Lebesgue Integral (First Approach)

Prerequisites: §3.1(a)

Definition 109 Lebesgue Integral of a Simple Function

Let $\Phi = \sum_{n=1}^{\infty} \varphi_n \cdot \mathbb{1}_{\mathbf{U}_n}$ *be a simple function. We define the* **(Lebesgue) integral** *of* Φ*:*

$$\int_{\mathbf{X}} \Phi \, d\mu \quad = \quad \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n]$$

if this sum is absolutely convergent. In this case, we say that Φ *is* **integrable**.

First, we check that any repartition of Φ yields the same value for $\int_{\mathbf{X}} \Phi \, d\mu$ in Definition 109.

Lemma 110 *If* $\sum_{n=1}^{N} \varphi_n \cdot \mathbb{1}_{\mathbf{U}_n} \;=\; \Phi \;=\; \sum_{k=1}^{K} \widetilde{\varphi}_n \cdot \mathbb{1}_{\widetilde{\mathbf{U}}_k}$, *then* $\sum_{n=1}^{N} \varphi_n \cdot \mu[\mathbf{U}_n] \;=\; \sum_{k=1}^{K} \widetilde{\varphi}_k \cdot \mu[\widetilde{\mathbf{U}}_k]$.

Proof: <u>**Exercise 98**</u> □

Now we extend this definition to arbitrary measurable functions.

Definition 111 Lebesgue Integral

Let $f : \mathbf{X} \longrightarrow \mathbb{C}$ *be an arbitrary measurable function.*

1. *If* $f : \mathbf{X} \longrightarrow [0, \infty]$, *then let* $\mathcal{S} \;=\; \{\Phi : \mathbf{X} \longrightarrow [0, \infty]\ ;\ \Phi \text{ simple, and } \Phi \leq f\}$, *and then define* $\int_{\mathbf{X}} f \; d\mu \;=\; \sup_{\Phi \in \mathcal{S}} \int_{\mathbf{X}} \Phi \; d\mu$. *If this supremum is finite, we say* f *is* **integrable**.
2. *If* $f : \mathbf{X} \longrightarrow [-\infty, \infty]$, *then we say* f *is* **integrable** *if* f^+ *and* f^- *are integrable, and then we define* $\int_{\mathbf{X}} f \; d\mu \;=\; \int_{\mathbf{X}} f^+ \; d\mu \;-\; \int_{\mathbf{X}} f^- \; d\mu$
3. *If* $f : \mathbf{X} \longrightarrow \mathbb{C}$, *then we say* f *is* **integrable** *if* f_{re} *and* f_{im} *are integrable, and then we define* $\int_{\mathbf{X}} f \; d\mu \;=\; \int_{\mathbf{X}} f_{\mathrm{re}} \; d\mu \;+\; \mathbf{i} \cdot \int_{\mathbf{X}} f_{\mathrm{im}} \; d\mu$.

Remark: Sometimes the notation "$\int_{\mathbf{X}} f(x) \; d\mu[x]$" is used, to emphasise the fact that we are integrating f as a function of x. Other times, the notation "$\int_{\mathbf{X}} f$" or even just "$\int f$" is used, when $\mathbf{X}$ and μ are understood.

3.1(c) Basic Properties of the Integral

Prerequisites: §3.1(b) **or** §3.1(d)

Lemma 112 $\Big(f \text{ is integrable} \Big) \iff \Big(|f| \text{ is integrable} \Big) \iff \left(\int_{\mathbf{X}} |f| \; d\mu \;<\; \infty \right)$.

Proof: <u>**Exercise 99**</u> □

The set of all integrable functions is denoted $\mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$.

Proposition 113 (Properties of the Integral)

The Lebesgue integral has the following properties:

Linearity: *If f, g are integrable, then* $\int_{\mathbf{X}} (f+g)\ d\mu \ = \ \int_{\mathbf{X}} f\ d\mu + \int_{\mathbf{X}} g\ d\mu$. *If $c \in \mathbb{C}$, then* $\int_{\mathbf{X}} c \cdot f(x)\ d\mu[x] \ = \ c \cdot \int_{\mathbf{X}} f(x)\ d\mu[x]$.

Monotonicity: *If f and g are real-valued, and $f(x) \leq g(x)$ for μ-almost every $x \in \mathbf{X}$, then* $\int_{\mathbf{X}} f\ d\mu \ \leq \ \int_{\mathbf{X}} g\ d\mu$.

Determines a Density: *If $f : \mathbf{X} \longrightarrow [0, \infty]$ is nonnegative, then we can define a new measure:*

$$\mu_f : \mathcal{X} \ni \mathbf{U} \mapsto \left(\int_{\mathbf{U}} f\ d\mu \right) \in [0, \infty].$$

Identity: *The following are equivalent:*

1. $f = g$ *almost everywhere.*
2. $\int_{\mathbf{U}} f\ d\mu \ = \ \int_{\mathbf{U}} g\ d\mu$, *for every* $\mathbf{U} \in \mathcal{X}$.
3. $\int_{\mathbf{X}} |f - g|\ d\mu \ = \ 0$.

We will establish these properties in two stages: first for simple functions, then for arbitrary functions. To pass from the first stage to the second, we must pause to develop an important convergence theorem.

Proof of Theorem 113 'Monotonicity':

Case 1: (*Simple functions*) By Lemma 106, we can repartition Φ and Ψ so that they are compatible; by Lemma 110, this does not change the values of their integrals. So, let $\Phi = \sum_{n=1}^{\infty} \varphi_n \cdot \mathbb{1}_{\mathbf{U}_n}$ and $\Psi = \sum_{n=1}^{\infty} \psi_n \cdot \mathbb{1}_{\mathbf{U}_n}$ If $\Phi \leq \Psi$, this means that $\varphi_n \leq \psi_n$ for all n. Thus,

$$\int_{\mathbf{X}} \Phi\ d\mu \ = \ \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n] \ \leq \ \sum_{n=1}^{\infty} \psi_n \cdot \mu[\mathbf{U}_n] \ = \ \int_{\mathbf{X}} \Psi\ d\mu.$$

Case 2: (*Nonnegative functions*) Let

$$\begin{aligned} \mathcal{S}(f) &= \{\Phi : \mathbf{X} \longrightarrow [0, \infty]\ ;\ \Phi \text{ simple, and } \Phi \leq f\}, \\ \text{and } \mathcal{S}(g) &= \{\Gamma : \mathbf{X} \longrightarrow [0, \infty]\ ;\ \Gamma \text{ simple, and } \Gamma \leq g\}. \end{aligned}$$

If $f \leq g$, then clearly $\mathcal{S}(f) \subset \mathcal{S}(g)$; hence,

$$\int_{\mathbf{X}} f\ d\mu \ = \ \sup_{\Phi \in \mathcal{S}(f)} \int_{\mathbf{X}} \Phi\ d\mu \ \leq \ \sup_{\Gamma \in \mathcal{S}(g)} \int_{\mathbf{X}} \Gamma\ d\mu \ = \ \int_{\mathbf{X}} g\ d\mu.$$

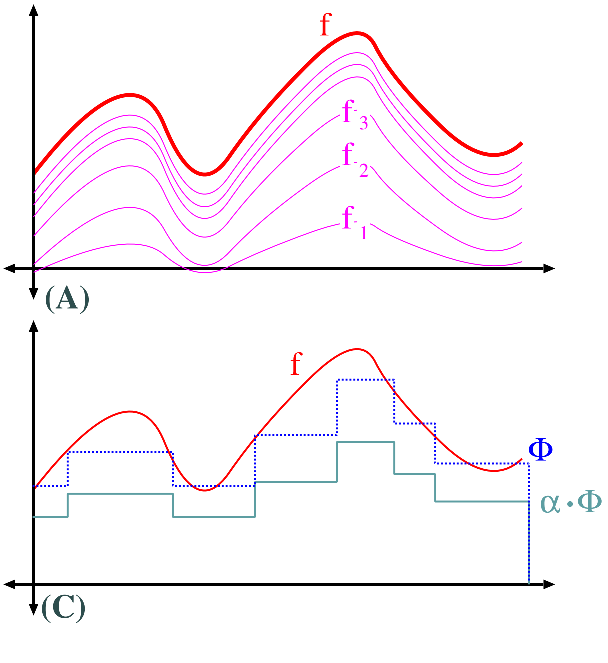


Figure 3.4: The Monotone Convergence Theorem

Case 3: (*Real functions*) If $f \leq g$, then $f^+ \leq g^+$ and $f^- \geq g^-$. Thus, by Case 2,

$$\int_{\mathbf{X}} f^+ \, d\mu \;\leq\; \int_{\mathbf{X}} g^+ \, d\mu \qquad \text{and} \qquad \int_{\mathbf{X}} f^- \, d\mu \;\geq\; \int_{\mathbf{X}} g^- \, d\mu.$$

Thus $\displaystyle\int_{\mathbf{X}} f \, d\mu \;=\; \int_{\mathbf{X}} f^+ \, d\mu - \int_{\mathbf{X}} f^- \, d\mu \;\leq\; \int_{\mathbf{X}} g^+ \, d\mu - \int_{\mathbf{X}} g^- \, d\mu \;=\; \int_{\mathbf{X}} g \, d\mu.$ □

Proof of Theorem 113 'Density' for simple functions: First suppose that $\Phi = \varphi \cdot \mathbb{1}_{\mathbf{U}}$ for some $\varphi \in \mathbb{C}$ and $\mathbf{U} \in \mathcal{X}$. Then for any $\mathbf{V} \in \mathcal{X}$,

$$\mu_\Phi(\mathbf{V}) \;=\; \varphi \cdot \mu\left[\mathbf{U} \cap \mathbf{V}\right] \;=\; \varphi \cdot \mu_{|_{\mathbf{U}}}[\mathbf{V}],$$

where $\mu_{|_{\mathbf{U}}}$ is the measure μ restricted to $\mathbf{U}$ (see § 1.2(c) on page 20). Thus, $\mu_\Phi = \varphi \cdot \mu_{|_{\mathbf{U}}}$ is also a measure.

Next, if $\Phi = \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n]$, then $\mu_\Phi \;=\; \sum_{n=1}^{\infty} \varphi_n \cdot \mu_{|_{\mathbf{U}_n}}$ is a linear combination of measures, so it is also a measure. □

Proposition 114 (Monotone Convergence Theorem)

Let $f_n : \mathbf{X}\longrightarrow[0,\infty]$ be measurable for all $n \in \mathbb{N}$, and suppose that, for μ-almost every $x \in \mathbf{X}$, $f_1(x) \leq f_2(x) \leq \ldots \leq f(x)$ and $\lim_{n\to\infty} f_n(x) = f(x)$, as in Figure 3.4(A).

Then $\displaystyle\int_{\mathbf{X}} f \, d\mu \;=\; \lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu \;=\; \sup_{n\in\mathbb{N}}\int_{\mathbf{X}} f_n \, d\mu.$

Proof: Case 1: ($\lim_{n\to\infty} f_n(x) = f(x)$ *everywhere on* $\mathbf{X}$)

Claim 1: $\displaystyle\lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu$ *exists and is not greater than* $\displaystyle\int_{\mathbf{X}} f \, d\mu$.

Proof: By the **'Monotonicity'** property from Theorem 113, we know that $\int_{\mathbf{X}} f_1 \, d\mu \leq \int_{\mathbf{X}} f_2 \, d\mu \leq \int_{\mathbf{X}} f_3 \, d\mu \leq \ldots$, forms an increasing sequence, and also that $\int_{\mathbf{X}} f_n \, d\mu \leq \int_{\mathbf{X}} f \, d\mu$ for all n. Thus, $\displaystyle\lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu = \sup_{n\in\mathbb{N}}\int_{\mathbf{X}} f_n \, d\mu \leq \int_{\mathbf{X}} f \, d\mu.$ ..... □ `[Claim 1]`

Now, let $\Phi \leq f$ be some nonnegative simple function, as in Figure 3.4(B).

Claim 2: $\displaystyle\int_{\mathbf{X}} \Phi \, d\mu \;\leq\; \lim_{n\to\infty}\int_{\mathbf{X}} f_n.$

Proof: Fix $0 < \alpha < 1$, so that $\alpha\cdot\Phi < \Phi$ (Figure 3.4C). For all $n \in \mathbb{N}$, define $\mathbf{V}_n = \{x \in \mathbf{X}\,;\, f_n(x) \geq \alpha\cdot\Phi(x)\}$, as in Figure 3.4(D). Thus,

$$\alpha\cdot\int_{\mathbf{V}_n} \Phi \, d\mu \;=\; \int_{\mathbf{V}_n} \alpha\cdot\Phi \, d\mu \;\leq\; \int_{\mathbf{V}_n} f_n \, d\mu \;\leq\; \int_{\mathbf{X}} f_n \, d\mu. \tag{3.1}$$

Note that $\mathbf{V}_1 \subset \mathbf{V}_2 \subset \ldots$, and $\mathbf{X} = \bigcup_{n=1}^{\infty} \mathbf{V}_n$. The **Density** property of Theorem 113 says μ_Φ is a measure. Thus,

$$\alpha\int_{\mathbf{X}} \Phi \, d\mu \;=\; \alpha\cdot\mu_\Phi[\mathbf{X}] \;=_{(a)}\; \alpha\cdot\lim_{n\to\infty}\mu_\Phi[\mathbf{V}_n] \;=\; \alpha\cdot\lim_{n\to\infty}\int_{\mathbf{V}_n} \Phi \, d\mu \;\leq_{(b)}\; \lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu \tag{3.2}$$

(a) by the **Lower Continuity** property from Proposition 15 on page 17 (b) By formula (3.1).

Let $\alpha\to 1$ in (3.2) to conclude that $\displaystyle\int_{\mathbf{X}} \Phi \, d\mu \leq \lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu$ .... □ `[Claim 2]`

Thus, taking the supremum over all simple functions $\Phi \leq f$, we conclude from Claim 2 that

$$\int_{\mathbf{X}} f \, d\mu \;=\; \sup_{\Phi\leq f}\int_{\mathbf{X}} \Phi \, d\mu \;\leq\; \lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu.$$

By combining this with Claim 1, we conclude that $\displaystyle\int_{\mathbf{X}} f \, d\mu = \lim_{n\to\infty}\int_{\mathbf{X}} f_n \, d\mu.$ ______□

Corollary 115 (Levi's Theorem)

Let $0 \leq f_1(x) \leq f_2(x) \leq \ldots \leq f(x)$ be as in the Monotone Convergence Theorem, and let $M < \infty$ be such that $\int_{\mathbf{X}} f_n \, d\mu \leq M$ for all n. Then $f(x)$ is finite almost everywhere, and $\int_{\mathbf{X}} f \, d\mu \; \leq \; M$.

Proof: It follows immediately from the Monotone Convergence Theorem that $\int_{\mathbf{X}} f \, d\mu \; \leq \; M$. To see that f must be almost-everywhere finite, let $\mathbf{Y} = \{x \in \mathbf{X} \; ; \; f(x) = \infty\}$; if $\mu[\mathbf{Y}] > 0$, then $\int_{\mathbf{X}} f \, d\mu = \infty \cdot \mu[\mathbf{Y}] = \infty$. □

Corollary 116 *If $f : \mathbf{X} \longrightarrow [0, \infty]$ is measurable, then there is sequence of nonnegative simple functions $\{\Phi_n\}_{n=1}^{\infty}$ so that,*

- *For all $x \in \mathbf{X}$, $\quad 0 \leq \Phi_1(x) \leq \Phi_2(x) \leq \ldots \leq f(x)$ and $\lim_{n\to\infty} \Phi_n(x) = f(x)$.*
- $\int_{\mathbf{X}} f \, d\mu \; = \; \lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n \, d\mu.$

Proof: Combine Lemma 108 with the Monotone Convergence Theorem. □

Proof of the rest of Theorem 113:

'Linearity' (*for simple functions*) Suppose Φ and Ψ are simple. By Lemma 106, we can assume that Φ and Ψ are compatible —that is, $\Phi = \sum_{n=1}^{\infty} \varphi_n \cdot \mathbb{1}_{\mathbf{U}_n}$ and $\Psi = \sum_{n=1}^{\infty} \psi_n \cdot \mathbb{1}_{\mathbf{U}_n}$. Thus, $\Phi + \Psi = \sum_{n=1}^{\infty} (\varphi_n + \psi_n) \cdot \mathbb{1}_{\mathbf{U}_n}$, so that

$$\int_{\mathbf{X}} \Phi + \Psi \, d\mu \; = \; \sum_{n=1}^{\infty} (\varphi_n + \psi_n) \cdot \mu[\mathbf{U}_n] \; = \; \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n] + \sum_{n=1}^{\infty} \psi_n \cdot \mu[\mathbf{U}_n] \; = \; \int_{\mathbf{X}} \Phi \, d\mu + \int_{\mathbf{X}} \Psi \, d\mu.$$

'Linearity' (*for nonnegative functions*) By Corollary 116, find sequences of simple functions $\{\Phi_n\}_{n=1}^{\infty}$ and $\{\Gamma_n\}_{n=1}^{\infty}$ converging pointwise to f and g from below, with $\int_{\mathbf{X}} f \, d\mu \; = \; \lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n \, d\mu$ and $\int_{\mathbf{X}} g \, d\mu \; = \; \lim_{n\to\infty} \int_{\mathbf{X}} \Gamma_n \, d\mu$.

Let $\Theta_n = \Phi_n + \Gamma_n$. Then $\{\Theta_n\}_{n=1}^{\infty}$ is a sequence of functions converging pointwise to $h = f+g$ from below, so by the Monotone Convergence Theorem,

$$\int_{\mathbf{X}} h \, d\mu \quad = \quad \lim_{n\to\infty} \int_{\mathbf{X}} \Theta_n \, d\mu \quad = \quad \lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n \, d\mu + \int_{\mathbf{X}} \Gamma_n \, d\mu \quad = \quad \int_{\mathbf{X}} f \, d\mu + \int_{\mathbf{X}} g \, d\mu.$$

'Linearity' (*for real- or complex-valued functions*) **Exercise 100** .

'Density' (*for arbitrary functions*) Let $\{\Phi_n\}_{n=1}^{\infty}$ be a sequence of simple functions converging to f from below. Then, for any $\mathbf{U} \in \mathcal{X}$, the Monotone Convergence Theorem says $\mu_f[\mathbf{U}] = \lim_{n\to\infty} \mu_{\Phi_n}[\mathbf{U}]$. Now apply Part 2 of Proposition 21 on page 22.

'Identity': First suppose f is a nonnegative function.

Claim 1: $\left(f =_\mu 0 \right) \iff \left(\int_{\mathbf{U}} f \, d\mu \ = 0 \textit{ for every } \mathbf{U} \in \mathcal{X} \right)$

Proof: ($\Longrightarrow$): Fix $\mathbf{U} \in \mathcal{X}$. Suppose $\Phi = \sum_{n=1}^{\infty} \varphi_n \cdot \mathbb{1}_{\mathbf{Z}_n}$ is a simple function, with $\Phi \leq f$, and with $\varphi_n > 0$ for all n. Then since $f =_\mu 0$, we must have $\mu[\mathbf{Z}_n] = 0$ for all n, and thus, $\mu[\mathbf{U} \cap \mathbf{Z}_n] = 0$. Thus, $\int_{\mathbf{U}} \Phi \quad d\mu \ = \ \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U} \cap \mathbf{Z}_n] = 0$. Since this is true for any $\Phi \leq f$, take the supremum over all Φ to conclude that $\int_{\mathbf{U}} f \quad d\mu \ = \ 0$.

($\Longleftarrow$): For all $n \in \mathbb{N}$, let $\mathbf{U}_n = \left\{ x \in \mathbf{X} \ ; \ f(x) \geq \frac{1}{n} \right\}$, and let $f_n = \frac{1}{n} \cdot \mathbb{1}_{\mathbf{U}_n}$. Then clearly $f_n \leq f$, so:

$$0 \quad \leq \quad \frac{1}{n} \cdot \mu\left[\mathbf{U}_n\right] \quad = \quad \int_{\mathbf{U}_n} f_n \, d\mu \quad \leq_{(*)} \quad \int_{\mathbf{U}_n} f \, d\mu \quad =_{(\dagger)} \quad 0,$$

where $(*)$ is by **'Monotonicity'** and $(\dagger)$ is by hypothesis. Thus, $\mu\left[\mathbf{U}_n\right] = 0$. Thus is true for all n; and $\mathsf{supp}\,[f] = \bigcup_{n=1}^{\infty} \mathbf{U}_n$, so conclude that $\mu\left(\mathsf{supp}\,[f]\right) = 0$. ...... $\square$ `[Claim 1]`

The remaining proof of the **'Identity'** property is **Exercise 101** . $\square$

Proof of Monotone Convergence Theorem *for* **a.e.***convergence*: **Exercise 102** $\square$

3.1(d) Definition of Lebesgue Integral (Second Approach)

Prerequisites: §3.1(a) **Recommended:** §3.1(b)

Case 1: (*Finite Measure Space*) Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a *finite* measure space. If Φ is a simple function, we define the integral of Φ exactly as in Definition 109 on page 73, and we say that Φ is **integrable** if this integral is finite.

Now, let $f : \mathbf{X} \longrightarrow \mathbb{C}$ be an arbitrary measurable function. By Lemma 108, there is a sequence of simple functions converging uniformly to f. We say that f is **integrable** if there is a sequence of *integrable* simple functions converging uniformly to f.

Definition 117 Lebesgue Integral

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a finite measure space. Let $f : \mathbf{X} \longrightarrow \mathbb{C}$ be integrable, and suppose $\{\Phi_n\}_{n=1}^{\infty}$ is a sequence of integrable simple functions converging uniformly to f. Then we define the **(Lebesgue) integral** *of f:*

$$\int_{\mathbf{X}}^{\dagger} f \; d\mu \quad = \quad \lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n \; d\mu \tag{3.3}$$

Also, if $\mathbf{U} \subset \mathbf{X}$ is measurable, then define $\displaystyle\int_{\mathbf{U}}^{\dagger} f \; d\mu \quad = \quad \int_{\mathbf{X}} \mathbb{1}_{\mathbf{U}} \cdot f \; d\mu.$

We use the notation "$\int_{\mathbf{X}}^{\dagger} f \; d\mu$" to distinguish this integral from the integral of Definition 111 on page 74. However, we will soon show that the two are equivalent. We must first ensure that the expression (3.3) is well-defined:

Lemma 118

1. *If $\{\Phi_n\}_{n=1}^{\infty}$ converges uniformly to f, then the limit in (3.3) exists.*
2. *The limit in (3.3) is independent of the sequence $\{\Phi_n\}_{n=1}^{\infty}$. That is: if $\{\Psi_n\}_{n=1}^{\infty}$ is another sequence of simple functions converging uniformly to f, then $\displaystyle\lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n \; d\mu = \lim_{n\to\infty} \int_{\mathbf{X}} \Psi_n \; d\mu$.*

Proof: **(1)** By multiplying μ by a scalar if necessary, we can assume $\mu[\mathbf{X}] = 1$. Let $I_n = \displaystyle\int_{\mathbf{X}} \Phi_n \; d\mu$ for all n. Thus, $\{I_n\}_{n=1}^{\infty}$ is a sequence of real numbers; we will show it is Cauchy, and thus, has a limit. Fix $\epsilon > 0$. Since $\{\Phi_n\}_{n=1}^{\infty}$ converges uniformly to f, find N so that $\displaystyle\sup_{x\in\mathbf{X}} |f(x) - \Phi_n(x)| < \frac{\epsilon}{2}$ for any $n > N$. Thus,

$$\sup_{x\in\mathbf{X}} |\Phi_m(x) - \Phi_n(x)| \quad \leq \quad \sup_{x\in\mathbf{X}} |\Phi_m(x) - \Phi(x)| \; + \; \sup_{x\in\mathbf{X}} |\Phi(x) - \Phi_n(x)| \quad < \quad \epsilon \tag{3.4}$$

for any $n, m > N$. By Lemma 106 on page 71, we are free to assume that Φ_n and Φ_m are compatible —in other words, $\Phi_n = \displaystyle\sum_{k=1}^{\infty} \varphi_n^{(k)} \cdot \mathbb{1}_{\mathbf{U}_k}$ and $\Phi_m = \displaystyle\sum_{k=1}^{\infty} \varphi_m^{(k)} \cdot \mathbb{1}_{\mathbf{U}_k}$, where $\mathbf{U}_1, \mathbf{U}_2, \ldots$

are disjoint subsets of $\mathbf{X}$. Thus, (3.4) implies that $\left|\varphi_n^{(k)} - \varphi_m^{(k)}\right| < \epsilon$ for all $k \in \mathbb{N}$. From this, we conclude

$$\begin{aligned} |I_n - I_m| &= \left| \int_{\mathbf{X}} \Phi_n \, d\mu - \int_{\mathbf{X}} \Phi_m \, d\mu \right| = \left| \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n] - \sum_{n=1}^{\infty} \varphi_n \cdot \mu[\mathbf{U}_n] \right| \\ &\leq \sum_{n=1}^{\infty} |\varphi_n - \varphi_n| \cdot \mu[\mathbf{U}_n] \leq \sum_{n=1}^{\infty} \epsilon \cdot \mu[\mathbf{U}_n] \leq \epsilon \cdot \mu[\mathbf{X}] = \epsilon. \end{aligned}$$

Since ϵ was arbitrary, it follows that $\{I_n\}_{n=1}^{\infty}$ is Cauchy, thus, convergent.

(2): Suppose $\{\Phi_n\}_{n=1}^{\infty}$ and $\{\Psi_n\}_{n=1}^{\infty}$ are two sequence of integrable simple functions converging uniformly to f. Let $I_n = \int_{\mathbf{X}} \Phi_n \, d\mu$ and $J_n = \int_{\mathbf{X}} \Phi_n \, d\mu$ for all n. We want to show that $\lim_{n\to\infty} I_n = \lim_{n\to\infty} J_n$. Consider the sequence $\{\Phi_1, \Psi_1, \Phi_2, \Psi_2, \Phi_3, \ldots\}$. This is *also* a sequence of simple functions converging uniformly to f, so by **Part (1)**, we know that the sequence $\{I_1, J_1, I_2, J_2, I_3, \ldots\}$ converges. In particular, this means that the subsequences $\{I_n\}_{n=1}^{\infty}$ and $\{J_n\}_{n=1}^{\infty}$ must converge to the same value. ______ □

Lemma 119 (Monotonicity)

If $f, g : \mathbf{X} \longrightarrow \mathbb{R}$ are integrable, and $f \leq g$, then $\int_{\mathbf{X}}^{\dagger} f \, d\mu \leq \int_{\mathbf{X}}^{\dagger} g \, d\mu$.

Proof: The case when f and g are simple is exactly as in the proof of Theorem 113 on page 74. So, suppose f and g are real-valued. Let $\{\Phi_n\}_{n=1}^{\infty}$ and $\{\Gamma_n\}_{n=1}^{\infty}$ be sequences of simple functions converging uniformly to f and g, respectively. Define $\underline{\Phi}_n(x) = \min\{\Phi_n(x), \Gamma_n(x)\}$ and $\overline{\Gamma}_n(x) = \max\{\Phi_n(x), \Gamma_n(x)\}$ for all $n \in \mathbb{N}$ and $x \in \mathbf{X}$.

Claim 1: *$\{\underline{\Phi}_n\}_{n=1}^{\infty}$ and $\{\overline{\Gamma}_n\}_{n=1}^{\infty}$ are also simple functions, and also converge uniformly to f and g, respectively.*

Proof: **Exercise 103** .... □ [Claim 1]

By construction, $\underline{\Phi}_n \leq \overline{\Gamma}_n$ for all $n \in \mathbb{N}$. Thus, by Case 1, $\int_{\mathbf{X}}^{\dagger} \underline{\Phi}_n \, d\mu \leq \int_{\mathbf{X}}^{\dagger} \overline{\Gamma}_n \, d\mu$. Thus, $\int_{\mathbf{X}}^{\dagger} f \, d\mu = \lim_{n\to\infty} \int_{\mathbf{X}}^{\dagger} \underline{\Phi}_n \, d\mu \leq \lim_{n\to\infty} \int_{\mathbf{X}}^{\dagger} \overline{\Gamma}_n \, d\mu = \int_{\mathbf{X}}^{\dagger} g \, d\mu$. ______ □

Lemma 120 *Definition 117 is equivalent to Definition 111.*

Proof: Let $\int_{\mathbf{X}} f\ d\mu$ be the integral of Definition 111, and let $\int_{\mathbf{X}}^{\dagger} f\ d\mu$ be the integral of Definition 117.

Case 1: (*f nonnegative*) Let $\mathcal{S} = \{\Phi : \mathbf{X}\longrightarrow[0,\infty]\ ;\ \Phi \text{ simple, and } \Phi \leq f\}$. Recall that $\displaystyle\int_{\mathbf{X}} f\ d\mu \ = \ \sup_{\Phi\in\mathcal{S}} \int_{\mathbf{X}} \Phi\ d\mu$.

Claim 1: $\displaystyle\int_{\mathbf{X}} f\ d\mu \leq \int_{\mathbf{X}}^{\dagger} f\ d\mu.$

Proof: If $\Phi \in \mathcal{S}$, then $\displaystyle\int_{\mathbf{X}} \Phi\ d\mu \ = \ \int_{\mathbf{X}}^{\dagger} \Phi\ d\mu \ \leq_{(*)} \ \int_{\mathbf{X}}^{\dagger} f\ d\mu$, where $(*)$ is by Lemma 119.

We conclude that $\displaystyle\sup_{\Phi\in\mathcal{S}} \int_{\mathbf{X}} \Phi\ d\mu \ \leq \ \int_{\mathbf{X}}^{\dagger} f\ d\mu$. .... □ [Claim 1]

Claim 2: $\displaystyle\int_{\mathbf{X}}^{\dagger} f\ d\mu \ \leq \ \int_{\mathbf{X}} f\ d\mu.$

Proof: By Lemma 108 on page 72, there exists a sequence of simple functions $\Phi_1 \leq \Phi_2 \leq \ldots \leq f$ converging uniformly to f. For all $n \in \mathbb{N}$, Φ_n is integrable, because $\displaystyle\int_{\mathbf{X}} \Phi_n\ d\mu \leq \int_{\mathbf{X}}^{\dagger} f\ d\mu < \infty$, by Lemma 119. Thus,

$$\int_{\mathbf{X}}^{\dagger} f\ d\mu \ =_{(b)} \ \lim_{n\to\infty} \int_{\mathbf{X}} \Phi_n\ d\mu \ \leq \ \sup_{n\in\mathbb{N}} \int_{\mathbf{X}} \Phi_n\ d\mu \ \leq \ \sup_{\Phi\in\mathcal{S}} \int_{\mathbf{X}} \Phi\ d\mu \ =_{(b)} \ \int_{\mathbf{X}} f\ d\mu.$$

(a) by Definition 117 (b) by Definition 111. .... □ [Claim 2]

Case 2 (*f real or complex-valued*) **<u>Exercise 104</u>** Hint: Combine Definition 111 with the various cases of Lemma 108. □

<u>Exercise 105</u> Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ is an *infinite* measure space. Construct a counterexample to show that the integral from Definition 117 is *not* well-defined in this case.

Case 2: (*Sigma-Finite Measure Space*) Let $(\mathbf{X}, \mathcal{X}, \mu)$ be sigma-finite. Thus, there is a countable partition $\mathcal{P}$ of $\mathbf{X}$ so that $\mathbf{X} = \bigsqcup_{\mathbf{P}\in\mathcal{P}}^{\infty} \mathbf{P}$, where every atom $\mathbf{P}$ of $\mathcal{P}$ is measurable, and $\mu[\mathbf{P}] < \infty$. If $f : \mathbf{X}\longrightarrow\mathbb{R}$ is measurable, we define

$$\int_{\mathbf{X}}^{\dagger} f\ d\mu \ = \ \sum_{\mathbf{P}\in\mathcal{P}} \int_{\mathbf{P}} f\ d\mu, \tag{3.5}$$

if this sum converges absolutely; in this case, f is called **integrable**.

First we must check that the definition does not depend upon the choice of partition.

Lemma 121 *Suppose $\mathcal{P}$ and $\mathcal{Q}$ are countable partitions. Then* $\displaystyle\sum_{\mathbf{P}\in\mathcal{P}} \int_{\mathbf{P}} f\, d\mu \;=\; \sum_{\mathbf{Q}\in\mathcal{Q}} \int_{\mathbf{Q}} f\, d\mu.$

Proof: **Case 1:** ($\mathcal{Q}$ *refines* $\mathcal{P}$) Suppose we index $\mathcal{P}$ in some arbitrary fashion: $\mathcal{P} = \{\mathbf{P}_n\}_{n=1}^{\infty}$. Since $\mathcal{Q}$ refines $\mathcal{P}$, we can write every atom of $\mathcal{P}$ as a (countable) disjoint union of $\mathcal{Q}$-atoms —that is, $\mathbf{P}_n = \displaystyle\bigsqcup_{m=1}^{\infty} \mathbf{Q}_m^n$ for some $\mathbf{Q}_1^n, \mathbf{Q}_2^n, \ldots$ in $\mathcal{Q}$. Then we can index $\mathcal{Q}$ like: $\mathcal{Q} = \{\mathbf{Q}_m^n\}_{n,m=1}^{\infty}$. Then clearly,

$$\sum_{\mathbf{P}\in\mathcal{P}} \int_{\mathbf{P}} f\, d\mu \;=\; \sum_{n=1}^{\infty} \int_{\mathbf{P}_n} f\, d\mu \;=\; \sum_{n,m=1}^{\infty} \int_{\mathbf{Q}_m^n} f\, d\mu \;=\; \sum_{\mathbf{Q}\in\mathcal{Q}} \int_{\mathbf{Q}} f\, d\mu.$$

Case 2: ($\mathcal{Q}$ *and* $\mathcal{P}$ *arbitrary*) Let $\mathcal{R} = \mathcal{Q} \vee \mathcal{P}$. Then $\mathcal{R}$ refines both $\mathcal{P}$ and $\mathcal{Q}$. Apply Case 1 to conclude that: $\displaystyle\sum_{\mathbf{P}\in\mathcal{P}} \int_{\mathbf{P}} f\, d\mu = \sum_{\mathbf{R}\in\mathcal{R}} \int_{\mathbf{R}} f\, d\mu = \sum_{\mathbf{Q}\in\mathcal{Q}} \int_{\mathbf{Q}} f\, d\mu.$ ☐

Lemma 122

1. *The value of* $\displaystyle\int_{\mathbf{X}}^{\dagger} f\, d\mu$ *in formula (3.5) agrees with the value of* $\displaystyle\int_{\mathbf{X}} f\, d\mu$ *in Definition 111.*
2. *If* $(\mathbf{X}, \mathcal{X}, \mu)$ *is finite, then the value of* $\displaystyle\int_{\mathbf{X}}^{\dagger} f\, d\mu$ *in formula (3.5) agrees with the value from formula (3.3).*

Proof: <u>**Exercise 106**</u> ☐

Case 3: (*Arbitrary Measure Space*) Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ is an arbitrary measure space. Let $\mathcal{F} = \{\mathbf{U} \in \mathcal{X}\ ;\ \mu[\mathbf{U}] < \infty\}$ be the collection of all subsets with finite measure. If $f : \mathbf{X} \longrightarrow \mathbb{R}$ is measurable, then we define

$$\int_{\mathbf{X}}^{\dagger} f\, d\mu \;=\; \sup_{\mathbf{F}\in\mathcal{F}} \int_{\mathbf{F}}^{\dagger} f\, d\mu \;-\; \inf_{\mathbf{F}\in\mathcal{F}} \int_{\mathbf{F}}^{\dagger} f\, d\mu \tag{3.6}$$

Lemma 123

1. *The value of* $\displaystyle\int_{\mathbf{X}}^{\dagger} f\, d\mu$ *in formula (3.6) agrees with the value of* $\displaystyle\int_{\mathbf{X}} f\, d\mu$ *in Definition 111.*
2. *If* $(\mathbf{X}, \mathcal{X}, \mu)$ *is sigma-finite, then the value of* $\displaystyle\int_{\mathbf{X}}^{\dagger} f\, d\mu$ *in formula (3.6) agrees with the value from formula (3.5)*

Proof: <u>**Exercise 107**</u> ☐

3.2 Limit Theorems

Prerequisites: §3.1

Proposition 124 (Fatou's Lemma)

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of nonnegative integrable functions.

1. *Let $\underline{f}(x) = \liminf_{n\to\infty} f_n(x)$ for all $x \in \mathbf{X}$. Then $\underline{f}$ is also integrable, and*

$$\int_{\mathbf{X}} \underline{f}\ d\mu \ \leq\ \liminf_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu.$$

2. *If $f(x) = \lim_{n\to\infty} f_n(x)$ exists for μ-almost all $x \in \mathbf{X}$, then f is also integrable, and*

$$\int_{\mathbf{X}} f\ d\mu \ \leq\ \liminf_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu.$$

Proof: **(1):** For all $N \in \mathbb{N}$, define $\underline{F}_N(x) = \inf_{n\in[N...\infty]} f_n(x)$ for all $x \in \mathbf{X}$. Then $\underline{F}_N$ is measurable, and $\underline{F}_N \leq f_n$ for any $n \geq N$. Thus, by **'Monotonicity'** from Theorem 113 on page 74,

$$\int_{\mathbf{X}} \underline{F}_N \quad d\mu \ \leq\ \int_{\mathbf{X}} f_n \quad d\mu.$$

Since this holds for all $n \geq N$, it follows:

$$\int_{\mathbf{X}} \underline{F}_N \quad d\mu \ \leq\ \inf_{n\in[N...\infty]} \int_{\mathbf{X}} f_n \quad d\mu. \tag{3.7}$$

However, $\underline{F}_1 \leq \underline{F}_2 \leq \ldots$ forms an increasing sequence, and $\underline{f}(x) = \lim_{N\to\infty} \underline{F}_N(x)$; Hence,

$$\begin{aligned}\int_{\mathbf{X}} \underline{f} \quad d\mu \ &=_{(a)}\ \lim_{N\to\infty} \int_{\mathbf{X}} \underline{F}_N \quad d\mu \quad \leq_{(b)} \quad \lim_{N\to\infty}\ \inf_{n\in[N...\infty]} \int_{\mathbf{X}} f_n \quad d\mu \\ &=\ \liminf_{n\to\infty} \int_{\mathbf{X}} f_n \quad d\mu, \qquad \text{as desired.}\end{aligned}$$

(a) by Monotone Convergence (Theorem 114 on page 77); (b) by (3.7).

(2): By modifying the functions f_n on a set of measure zero, we can assume that $f(x) = \lim_{n\to\infty} f_n(x)$ for *all* $x \in \mathbf{X}$; this does not change the value of the integrals because of the **'Identity'** property Theorem 113. At this point, **(2)** follows from **(1)**. □

Theorem 125 (Lebesgue's Dominated Convergence Theorem)

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of integrable functions. Suppose:

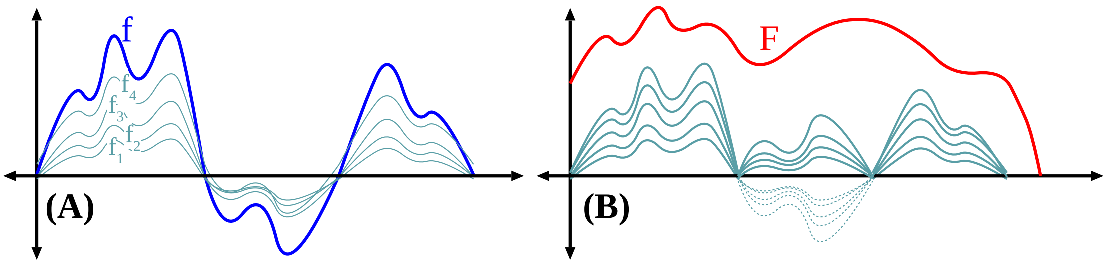


Figure 3.5: Lebesgue's Dominated Convergence Theorem

- *$f(x) = \lim_{n\to\infty} f_n(x)$ exists for μ-almost all $x \in \mathbf{X}$ (Figure 3.5A).*
- *There is some integrable $F \in \mathbf{L}^1(\mathbf{X}, \mu)$ so that $|f_n(x)| \leq F(x)$ for almost all $x \in \mathbf{X}$ and all $n \in \mathbb{N}$ (Figure 3.5B).*

Then $\displaystyle\int_{\mathbf{X}} f\ d\mu \ = \ \lim_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu.$

Proof: We will prove the theorem when f_n are real-valued. If f_n are complex-valued, then simply apply the proof to $\mathsf{re}[f_n]$ and $\mathsf{im}[f_n]$ separately.

By modifying the functions f_n on a set of measure zero, we can assume that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in \mathbf{X}$, and that $|f_n(x)| \leq F(x)$ for all $x \in \mathbf{X}$. By the **'Identity'** property, this does not modify any of the relevant integrals.

Case 1: $(f \equiv 0)$ Since $|f_n| < F$, it follows that $F + f_n > 0$ and $F - f_n > 0$. Since $\lim_{n\to\infty} f_n \ = \ f \ \equiv \ 0$, we have:

$$\liminf_{n\to\infty} f_n \quad = \quad 0 \quad = \quad \liminf_{n\to\infty} (-f_n).$$

Thus, $\liminf_{n\to\infty} (F + f_n) \ = \ F \ = \ \liminf_{n\to\infty} (F - f_n)$. Thus applying Fatou's Lemma:

$$\begin{aligned} \int_{\mathbf{X}} F\ d\mu \ &\leq \ \liminf_{n\to\infty} \int_{\mathbf{X}} (F + f_n)\ d\mu \ = \ \int_{\mathbf{X}} F\ d\mu \ + \ \liminf_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu, \\ \text{and} \quad \int_{\mathbf{X}} F\ d\mu \ &\leq \ \liminf_{n\to\infty} \int_{\mathbf{X}} (F - f_n)\ d\mu \ = \ \int_{\mathbf{X}} F\ d\mu \ + \ \liminf_{n\to\infty} \left(-\int_{\mathbf{X}} f_n\ d\mu\right) \\ &= \ \int_{\mathbf{X}} F\ d\mu \ - \ \limsup_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu. \end{aligned}$$

Subtract $\displaystyle\int_{\mathbf{X}} F\ d\mu$ to conclude:

$$0 \ \leq \ \liminf_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu \qquad \text{and} \qquad 0 \ \leq \ -\limsup_{n\to\infty} \int_{\mathbf{X}} f_n\ d\mu.$$

Thus,

$$\limsup_{n\to\infty} \int_{\mathbf{X}} f_n \, d\mu \quad \leq \quad 0 \quad \leq \quad \liminf_{n\to\infty} \int_{\mathbf{X}} f_n \, d\mu.$$

It follows that $\lim_{n\to\infty} \int_{\mathbf{X}} f_n \, d\mu = 0$

Case 2: ($f(x)$ *real-valued*) Since $\lim_{n\to\infty} f_n = f$, and $|f_n| < F$, it follows that $|f(x)| \leq F(x)$ for all $x \in \mathbf{X}$. Since F is integrable, it follows that f is also integrable, because:

$$\int_{\mathbf{X}} |f| \, d\mu \ \leq \ \int_{\mathbf{X}} F \, d\mu \quad < \quad \infty.$$

Observe that $\int_{\mathbf{X}} f \, d\mu \ - \ \int_{\mathbf{X}} f_n \, d\mu \quad = \quad \int_{\mathbf{X}} (f - f_n) \, d\mu$. Thus, it suffices to show that $\lim_{n\to\infty} \int_{\mathbf{X}} (f - f_n) \, d\mu \ = \ 0$. So, let $g_n = f - f_n$. Then:

- $\lim_{n\to\infty} g_n = 0$ everywhere.
- $|g_n(x)| < 2 \cdot F(x)$ for all x, and $2 \cdot F$ is also integrable.

Hence, applying Case 1, we conclude that $\lim_{n\to\infty} \int_{\mathbf{X}} g_n \quad d\mu = 0$, as desired. $\square$

Remark: The 'dominating' function F in the Dominated Convergence Theorem is necessary. To see this, let $\mathbf{X} = \mathbb{R}$, with the Lebesgue measure λ. Let $f_n(x) = \mathbb{1}_{[n,n+1]}$. Then clearly, $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$. However, $\int_{\mathbb{R}} f_n \, d\lambda = 1$ for all n, so that $\lim_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda \ = \ 1$. The Dominated Convergence Theorem 'fails' because there is no integrable function $F : \mathbb{R} \longrightarrow \mathbb{R}$ such that $|f_n| \ \leq \ F$ for all n.

3.3 Integration over Product Spaces

Prerequisites: §3.1 **Recommended:** §2.1(b)

Throughout this section, let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be two sigma-finite measure spaces, and let $(\mathbf{Z}, \mathcal{Z}, \zeta)$ be their product; ie.:

$$\mathbf{Z} \ = \ \mathbf{X} \times \mathbf{Y}, \qquad \mathcal{Z} \ = \ \mathcal{X} \otimes \mathcal{Y}, \qquad \text{and} \qquad \zeta \ = \ \xi \otimes \Upsilon.$$

A good example to keep in mind: if $\mathbf{X} = \mathbb{R} = \mathbf{Y}$ and ξ and Υ are Lebesgue measure, then $\mathbf{Z} = \mathbb{R}^2$, and ζ is the two-dimensional Lebesgue measure.

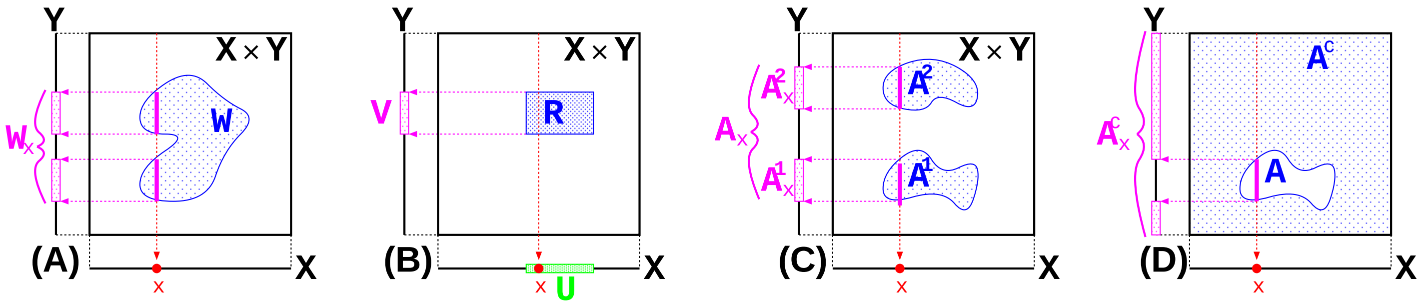


Figure 3.6: The fibre of a set.

We know from classical multivariate calculus that a two-dimensional Riemann integral can be computed via two consecutive one-dimensional integrals:

$$\int_{\mathbb{R}^2} f(\mathbf{x})\ d\mathbf{x} \quad = \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2)\ dx_1\ dx_2$$

We want to generalize this formula to an arbitrary product measure. We will need the following technical lemma.

Lemma 126 *Let $\mathcal{R} \subset \mathcal{S} \subset \mathcal{P}(\mathbf{Z})$ be two collections of subsets such that:*

1. *$\mathcal{S}$ contains the algebra generated by $\mathcal{R}$.*
2. *$\mathcal{S}$* **is closed under countable increasing unions:** *If $\mathbf{S}_1 \subset \mathbf{S}_2 \subset \ldots$ is an increasing sequence of elements in $\mathcal{S}$, then $\left(\bigcup_{n=1}^{\infty} \mathbf{S}_n\right) \in \mathcal{S}$.*
3. *$\mathcal{S}$* **is closed under countable decreasing intersections:** *If $\mathbf{S}_1 \supset \mathbf{S}_2 \supset \ldots$ is an decreasing sequence of elements in $\mathcal{S}$, then $\left(\bigcap_{n=1}^{\infty} \mathbf{S}_n\right) \in \mathcal{S}$.*

Then $\mathcal{S}$ contains the sigma-algebra generated by $\mathcal{R}$.

Proof: <u>**Exercise 108**</u> □

If $\mathbf{W} \subset \mathbf{Z}$, then for all $x \in \mathbf{X}$, the **fibre** of $\mathbf{W}$ over x is the subset of $\mathbf{Y}$ defined:

$$\mathbf{W}_x \quad = \quad \{y \in \mathbf{Y}\ ;\ (x, y) \in \mathbf{W}\} \qquad \text{(see Figure 3.6A)}$$

If $f : \mathbf{Z} \longrightarrow \mathbb{C}$, then for all $x \in \mathbf{X}$, the **fibre** of f over x is the function $f_x : \mathbf{Y} \longrightarrow \mathbb{C}$ defined:

$$f_x(y) \quad = \quad f(x, y) \qquad \text{(see Figure 3.7A)}$$

Theorem 127 (Fubini-Tonelli)

1. *Let* $\mathbf{W} \in \mathcal{Z}$. *Then:*
   (a) *For all* $x \in \mathbf{X}$, $\quad \mathbf{W}_x \in \mathcal{Y}$.
   (b) $\zeta[\mathbf{W}] \;=\; \int_{\mathbf{X}} \Upsilon(\mathbf{W}_x)\; d\xi[x]$.
2. *Let* $f : \mathbf{Z} \longrightarrow \mathbb{C}$ *be* $\mathcal{Z}$*-measurable.*
   (a) *For all* $x \in \mathbf{X}$, $\quad f_x : \mathbf{Y} \longrightarrow \mathbb{C}$ *is* $\mathcal{Y}$*-measurable.*
   (b) *For all* $x \in \mathbf{X}$, *let* $F(x) \;=\; \int_{\mathbf{Y}} f_x(y)\; d\Upsilon[y]$. *If* $f \in \mathbf{L}^1(\mathbf{Z})$, *then* $F \in \mathbf{L}^1(\mathbf{X})$, *and*

$$\int_{\mathbf{Z}} f(z)\; d\zeta[z] \quad = \quad \int_{\mathbf{X}} F(x)\; d\xi[x] \quad = \quad \int_{\mathbf{X}} \left(\int_{\mathbf{Y}} f_x(y)\; d\Upsilon[y] \right)\; d\xi[x] \qquad (3.8)$$

   If f *is nonnegative, then (3.8) holds also when* $\int_{\mathbf{Z}} f\; d\zeta \;=\; \infty$.
3. *Suppose that* $(\mathbf{X}, \mathcal{X}, \xi)$ *and* $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ *are complete.* $(\mathbf{Z}, \mathcal{Z}, \zeta)$ *is not necessarily complete, but let* $\widetilde{\mathcal{Z}}$ *be the* ζ*-completion of* $\mathcal{Z}$.
   (a) *If* $\mathbf{W} \in \widetilde{\mathcal{Z}}$, *then for* $\forall_\xi\; x \in \mathbf{X}$, $\;\mathbf{W}_x \in \mathcal{Y}$; *also,* **1(b)** *holds.*
   (b) *If* $f : \mathbf{Z} \longrightarrow \mathbb{C}$ *is* $\widetilde{\mathcal{Z}}$*-measurable, then for* $\forall_\xi\; x \in \mathbf{X}$, $\quad f_x$ *is* $\mathcal{Y}$*-measurable.*
   (c) *If* $f \in \mathbf{L}^1(\mathbf{Z}, \widetilde{\mathcal{Z}}, \zeta)$, *then for* $\forall_\xi\; x \in \mathbf{X}$, $\quad f_x \in \mathbf{L}^1(\mathbf{Y}, \mathcal{Y}, \Upsilon)$; *also,* **2(b)** *holds.*

Proof:

Proof of 1(a) Let $\mathcal{A} = \{\mathbf{A} \subset \mathbf{Z}\;;\; \mathbf{A}_x \in \mathcal{Y},$ for all $x \in \mathbf{X}\}$. We want to show $\mathcal{Z} \subset \mathcal{A}$. Recall that $\mathcal{Z} = \mathcal{X} \otimes \mathcal{Y}$ is the sigma-algebra generated by the set of measurable rectangles:

$$\mathcal{R} = \{\mathbf{U} \times \mathbf{V}\;;\; \mathbf{U} \in \mathcal{X} \;\text{ and }\; \mathbf{V} \in \mathcal{Y}\}.$$

Thus, it suffices to show that $\mathcal{R} \subset \mathcal{A}$, and that $\mathcal{A}$ is a sigma-algebra.

Claim 1: $\mathcal{R} \subset \mathcal{A}$.

Proof: Suppose $\mathbf{R} \in \mathcal{R}$, with $\mathbf{R} = \mathbf{U} \times \mathbf{V}$. Then

$$\mathbf{R}_x \;=\; \begin{cases} \mathbf{V} & \text{if } x \in \mathbf{U}; \\ \emptyset & \text{if } x \notin \mathbf{U}. \end{cases} \qquad \text{(see Figure 3.6B)} \qquad (3.9)$$

so $\mathbf{R}_x \in \mathcal{Y}$ for all x, so $\mathbf{R} \in \mathcal{A}$. .... □ `[Claim 1]`

Claim 2: $\mathcal{A}$ *is a sigma-algebra.*

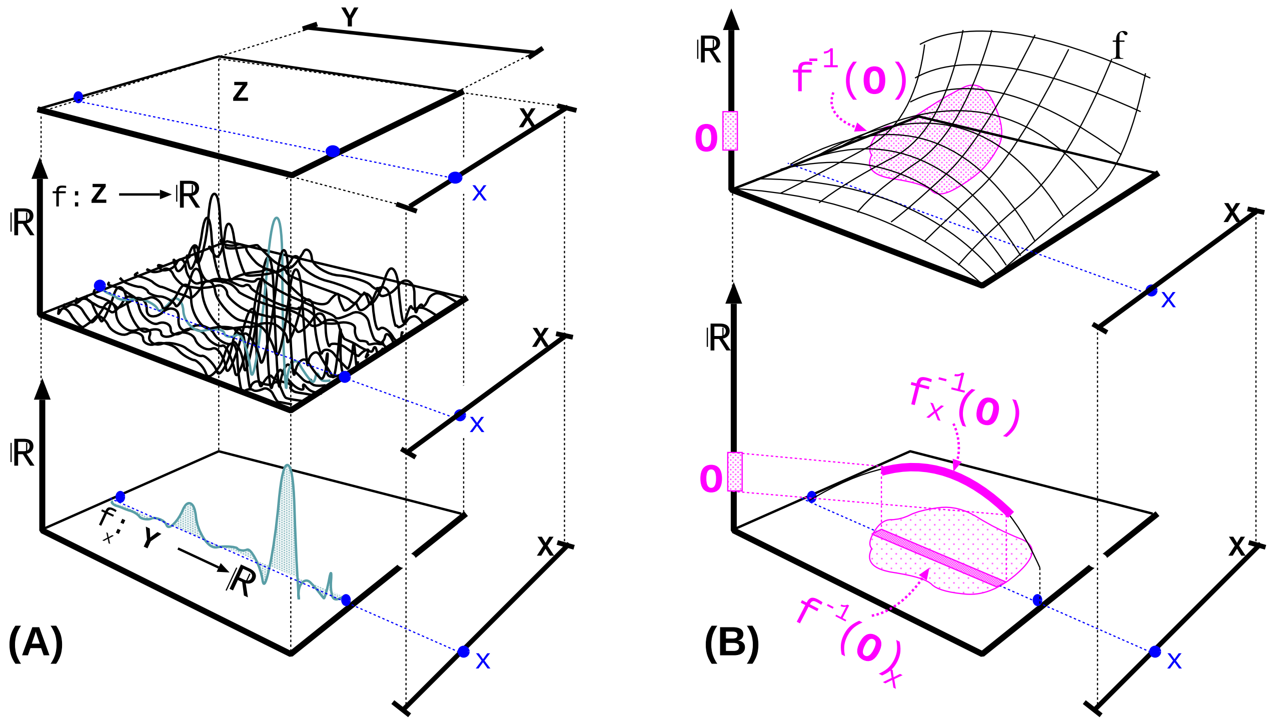


Figure 3.7: The fibre of a set.

Proof: It suffices to show that $\mathcal{A}$ is closed under complementation and countable union. Suppose $\mathbf{A}^{(n)} \in \mathcal{A}$ for all $n \in \mathbb{N}$, and let $\mathbf{A} = \bigcup_{n=1}^{\infty} \mathbf{A}^{(n)}$. Then for any $x \in \mathbf{X}$, $\mathbf{A}_x = \bigcup_{n=1}^{\infty} \mathbf{A}_x^{(n)}$ is a union of elements of $\mathcal{Y}$ (Figure 3.6C), so $\mathbf{A}_x \in \mathcal{Y}$. Hence $\mathbf{A} \in \mathcal{A}$. Likewise, if $\mathbf{A} \in \mathcal{A}$, then for any $x \in \mathbf{X}$, $\left(\mathbf{A}^{\complement}\right)_x = (\mathbf{A}_x)^{\complement} \in \mathcal{Y}$ (Figure 3.6D). Hence $\mathbf{A}^{\complement} \in \mathcal{A}$. □ `[Claim 2]`

Proof of 2(a) Fix $x \in \mathbf{X}$; for any open subset $\mathbf{O} \subset \mathbb{C}$, we want $f_x^{-1}(\mathbf{O}) \in \mathcal{Y}$. But

$$f_x^{-1}(\mathbf{O}) = \{y \in \mathbf{Y} \,;\, f_x(y) \in \mathbf{O}\} = \{y \in \mathbf{Y} \,;\, f(x,y) \in \mathbf{O}\} = \left\{y \in \mathbf{Y} \,;\, (x,y) \in f^{-1}(\mathbf{O})\right\}$$
$$= \left(f^{-1}(\mathbf{O})\right)_x \quad \text{(Figure 3.7B), and } (f^{-1}(\mathbf{O}))_x \in \mathcal{Y} \text{ by } \mathbf{1(a)}.$$

Proof of 1(b) **Case 1:** ($\mathbf{X}$ *and* $\mathbf{Y}$ *are finite*) Let $\mathcal{B} = \left\{\mathbf{B} \subset \mathbf{Z} \,;\, \zeta[\mathbf{B}] = \int_{\mathbf{X}} \Upsilon(\mathbf{B}_x)\, d\xi[x]\right\}$. We want to show that $\mathcal{Z} \subset \mathcal{B}$. Recall that $\mathcal{Z}$ is the sigma-algebra generated by $\mathcal{R}$; we will use Lemma 126 to show that $\mathcal{B}$ contains $\mathcal{Z}$.

Claim 3: $\mathcal{R} \subset \mathcal{A}$.

Proof: Suppose $\mathbf{R} \in \mathcal{R}$, with $\mathbf{R} = \mathbf{U} \times \mathbf{V}$. Then $\zeta[\mathbf{R}] = \xi[\mathbf{U}] \cdot \Upsilon[\mathbf{V}] = \int_{\mathbf{U}} \Upsilon[\mathbf{V}]\, d\mu =_{(*)} \int_{\mathbf{X}} \Upsilon[\mathbf{R}_x]\, d\mu[x]$, where $(*)$ follows from formula (3.9). . . . . □ `[Claim 3]`

Claim 4: *$\mathcal{B}$ is closed under finite disjoint unions.*

Proof: If $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \ldots, \mathbf{B}^{(N)} \in \mathcal{A}$ are disjoint, and $\mathbf{B} = \bigsqcup_{n=1}^{N} \mathbf{B}^{(n)}$, then for every $x \in \mathbf{X}$, $\mathbf{B}_x = \bigsqcup_{n=1}^{N} \mathbf{B}_x^{(n)}$ (Figure 3.6C). Then

$$\begin{aligned}
\int_{\mathbf{X}} \Upsilon(\mathbf{B}_x)\ d\xi[x] \quad &= \quad \int_{\mathbf{X}} \Upsilon\left(\bigsqcup_{n=1}^{N} \mathbf{B}_x^{(n)}\right)\ d\xi[x] \quad = \quad \sum_{n=1}^{N} \int_{\mathbf{X}} \Upsilon\left(\mathbf{B}_x^{(n)}\right)\ d\xi[x] \\
&= \quad \sum_{n=1}^{N} \zeta\left[\mathbf{B}^{(n)}\right] \quad = \quad \zeta[\mathbf{B}],
\end{aligned}$$

so $\mathbf{B} \in \mathcal{B}$ also. .... □ `[Claim 4]`

Claim 5: *$\mathcal{B}$ contains the algebra generated by $\mathcal{R}$.*

Proof: Example (61b)[1] says that the algebra generated by $\mathcal{R}$ is the set $\widetilde{\mathcal{R}}$ of all finite disjoint unions of rectangles. But $\widetilde{\mathcal{R}} \subset \mathcal{B}$ by Claims 3 and 4. .... □ `[Claim 5]`

Claim 6: *$\mathcal{B}$ is closed under countable increasing unions.*

Proof: Suppose $\mathbf{B}^{(1)} \subset \mathbf{B}^{(2)} \subset \ldots$ are in $\mathcal{B}$, and let $\mathbf{B} = \bigcup_{n=1}^{\infty} \mathbf{B}^{(n)}$. Define $F : \mathbf{X} \longrightarrow \mathbb{R}$ by $F(x) = \Upsilon(\mathbf{B}_x)$, and for all n, define $F_n : \mathbf{X} \longrightarrow \mathbb{R}$ by $F_n(x) = \Upsilon(\mathbf{B}_x^{(n)})$. Thus, $F_1 \leq F_2 \leq \ldots \leq F$, and $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbf{X}$ (by 'Upper Continuity' from Proposition 15 on page 17). Hence:

$$\int_{\mathbf{X}} F\ d\mu \quad =_{(a)} \quad \lim_{n\to\infty} \int_{\mathbf{X}} F_n\ d\mu \quad =_{(b)} \quad \lim_{n\to\infty} \zeta\left[\mathbf{B}^{(n)}\right] \quad =_{(c)} \quad \zeta\left[\bigcup_{n=1}^{\infty} \mathbf{B}^{(n)}\right] \quad = \quad \zeta[\mathbf{B}]$$

(a) by Monotone Convergence Theorem (page 77) (b) because $\mathbf{B}^{(n)} \in \mathcal{B}$. (c) by 'Upper Continuity' (Proposition 15 on page 17). .... □ `[Claim 6]`

Claim 7: *$\mathcal{B}$ is closed under countable decreasing intersections.*

Proof: Suppose $\mathbf{B}^{(1)} \supset \mathbf{B}^{(2)} \supset \ldots$ are in $\mathcal{B}$, and let $\mathbf{B} = \bigcap_{n=1}^{\infty} \mathbf{B}^{(n)}$. Define F and F_n as in Claim 6; then $F_1 \geq F_2 \geq \ldots \geq F$. Proceed as in Claim 6, but now apply the Dominated Convergence Theorem. .... □ `[Claim 7]`

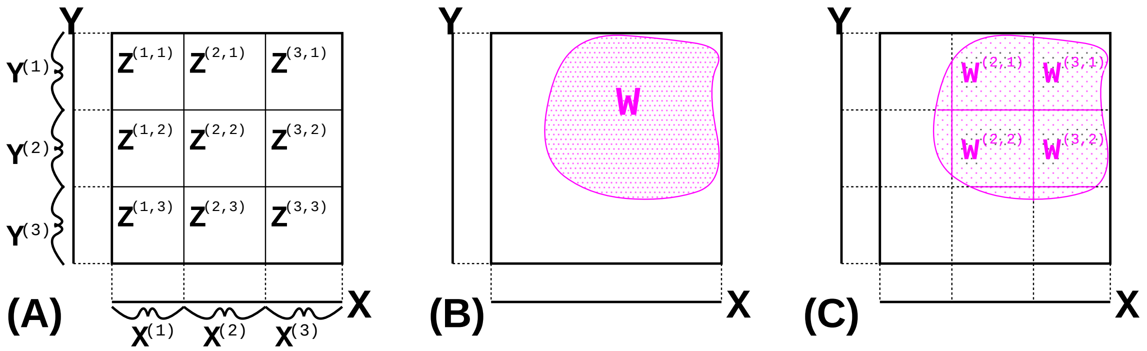


Figure 3.8: $\mathbf{Z}^{(n,m)} = \mathbf{X}^{(n)} \times \mathbf{Y}^{(m)}$.

By Claims 5, 6 and 7, and Lemma 126, we conclude that $\mathcal{Z} \subset \mathcal{B}$.

Case 2: ($\mathbf{X}$ *and* $\mathbf{Y}$ *are sigma-finite*) Write $\mathbf{X}$ and $\mathbf{Y}$ as disjoint unions of finite subspaces: $\mathbf{X} = \bigsqcup_{n=1}^{\infty} \mathbf{X}^{(n)}$ and $\mathbf{Y} = \bigsqcup_{m=1}^{\infty} \mathbf{Y}^{(m)}$. Then $\mathbf{Z} = \bigsqcup_{n,m=1}^{\infty} \mathbf{Z}^{(n,m)}$ where $\mathbf{Z}^{(n,m)} = \mathbf{X}^{(n)} \times \mathbf{Y}^{(m)}$ (Figure 3.8A). If $\mathbf{W} \subset \mathbf{Z}$ (Figure 3.8B), then let $\mathbf{W}^{(n,m)} = \mathbf{W} \cap \mathbf{Z}^{(n,m)}$ (Figure 3.8C). Thus,

$$\mathbf{W} \;=\; \bigsqcup_{n,m=1}^{\infty} \mathbf{W}^{(n,m)} \quad \text{and thus,} \quad \mathbf{W}_x \;=\; \bigsqcup_{n,m=1}^{\infty} \mathbf{W}_x^{(n,m)}, \quad \text{for all } x \in \mathbf{X}. \tag{3.10}$$

Thus,

$$\begin{aligned}
\int_{\mathbf{X}} \Upsilon(\mathbf{W}_x)\, d\xi[x] \;&=_{(a)}\; \int_{\mathbf{X}} \Upsilon\left(\bigsqcup_{n,m=1}^{\infty} \mathbf{W}_x^{(n,m)}\right) d\xi[x] \;=_{(b)}\; \sum_{n,m=1}^{\infty} \int_{\mathbf{X}_n} \Upsilon\left(\mathbf{W}_x^{(n,m)}\right) d\xi[x] \\
&=_{(c)}\; \sum_{n,m=1}^{\infty} \zeta\left[\mathbf{W}^{(n,m)}\right] \;=\; \zeta\left[\bigsqcup_{n,m=1}^{\infty} \mathbf{W}^{(n,m)}\right] \;=_{(d)}\; \zeta[\mathbf{W}]
\end{aligned}$$

(a) By (3.10). (b) By Monotone Convergence Theorem (page 77) and 'Upper Continuity' (Proposition 15 on page 17). (c) Apply Case 1 to $\mathbf{W}^{(n,m)} \subset \mathbf{Z}^{(n,m)}$ for all n and m. (d) By (3.10).

Proof of 2(b)

Case 1: (f *a characteristic function*) If $f = \mathbb{1}_{\mathbf{W}}$, then just apply part **1(b)** to get (3.8).

Case 2: (f *a simple function*) This follows immediately from Case 1.

Case 3: (f *is nonnegative*) By Corollary 116 on page 78, let $\phi^{(1)} \leq \phi^{(2)} \leq \ldots \leq f$ be a sequence of nonnegative simple functions increasing pointwise to f, such that

$$\lim_{n\to\infty} \int_{\mathbf{Z}} \phi^{(n)}(z)\, d\zeta[z] \;=\; \int_{\mathbf{Z}} f(z)\, d\zeta[z]. \tag{3.11}$$

[1] on page 45.

For all n, define $\Phi^{(n)} : \mathbf{X}\longrightarrow\mathbb{C}$ by: $\Phi^{(n)}(x) = \int_{\mathbf{Y}} \phi_x^{(n)}(y)\ d\Upsilon[y]$. Thus, by Case 2, we know that, for all $n \in \mathbb{N}$,

$$\int_{\mathbf{Z}} \phi^{(n)}(z)\ d\zeta[z] = \int_{\mathbf{X}} \Phi^{(n)}(x)\ d\mu[x]. \tag{3.12}$$

Claim 8: *For all $x \in \mathbf{X}$, $\Phi^{(1)}(x) \leq \Phi^{(2)}(x) \leq \ldots \leq F(x)$. Also, $\lim_{n\to\infty} \Phi^{(n)}(x) = F(x)$.*

Proof: For all $x \in \mathbf{X}$ and $y \in \mathbf{Y}$, clearly, $\phi_x^{(1)}(y) \leq \phi_x^{(2)}(y) \leq \ldots \leq f_x(y)$. Apply **Monotonicity** from Theorem 113 on page 74 and the Monotone Convergence Theorem (page 77). .... □ `[Claim 8]`

Thus,

$$\int_{\mathbf{Z}} f(z)\, d\zeta[z] \quad =_{(a)} \quad \lim_{n\to\infty} \int_{\mathbf{Z}} \phi^{(n)}(z)\ d\zeta[z] \quad =_{(b)} \quad \lim_{n\to\infty} \int_{\mathbf{X}} \Phi^{(n)}(x)\ d\mu[x] \quad =_{(c)} \quad \int_{\mathbf{X}} F(x)\ d\mu[x]$$

(a) By formula (3.11). (b) By formula (3.12). (b) By Monotone Convergence Theorem and Claim 8.

Case 4: (*f is real-valued*) f^+ and f^- are both nonnegative and have finite integrals. Thus,

$$\begin{aligned}
\int_{\mathbf{Z}} f(z)\ d\zeta[z] &= \int_{\mathbf{Z}} f^+(z)\ d\zeta[z] + \int_{\mathbf{Z}} f^-(z)\ d\zeta[z] \\
&=_{(a)} \int_{\mathbf{X}} \left(\int_{\mathbf{Y}} f_x^+(y)\ d\Upsilon[y]\right)\ d\xi[x] + \int_{\mathbf{X}} \left(\int_{\mathbf{Y}} f_x^-(y)\ d\Upsilon[y]\right)\ d\xi[x] \\
&=_{(b)} \int_{\mathbf{X}} \left(\int_{\mathbf{Y}} f_x^+(y)\ \Upsilon[y] + \int_{\mathbf{Y}} f_x^-(y)\ d\Upsilon[y]\right)\ d\xi[x] \\
&=_{(c)} \int_{\mathbf{X}} \int_{\mathbf{Y}} \left(f_x^+(y) + f_x^-(y)\right)\ d\Upsilon[y]\ d\xi[x] \\
&= \int_{\mathbf{X}} \int_{\mathbf{Y}} f_x(y)\ d\Upsilon[y]\ d\xi[x] \quad =_{(d)} \quad \int_{\mathbf{X}} F(x)\ d\xi[x]
\end{aligned}$$

(a) Apply Case 3 to f^+ and f^-. (b,c) By linearity of the integral. (d) By definition of F.

Case 5: (*f is complex-valued*) Apply Case 4 to f_{re} and f_{im}.

Proof of 3(a)

Claim 9: *Suppose $\mathbf{N} \in \mathcal{Z}$ and $\zeta[\mathbf{N}] = 0$. Then for ξ-almost all $x \in \mathbf{X}$, $\Upsilon[\mathbf{N}_x] = 0$.*

Proof: By part **1(b)**, we have $0 = \zeta[\mathbf{N}] = \int_{\mathbf{X}} \Upsilon[\mathbf{N}_x]\ d\mu[x]$. Hence, by the **Identity** property (Theorem 113 on page 74), conclude that $\Upsilon[\mathbf{N}_x] = 0$ for $\forall_\xi\ x$. . □ `[Claim 9]`

Now, let $\widetilde{\mathbf{W}} \in \widetilde{\mathcal{Z}}$. Thus,

$$\widetilde{\mathbf{W}} = \mathbf{W}^{(0)} \sqcup \mathbf{W}^{(1)}, \tag{3.13}$$

where $\mathbf{W}^{(1)} \in \mathcal{Z}$ and $\mathbf{W}^{(0)}$ is a null set; that is, $\mathbf{W}^{(0)} \subset \mathbf{N}$ where $\mathbf{N} \in \mathcal{Z}$ and $\zeta[\mathbf{N}] = 0$. Thus,

$$\text{For all } x \in \mathbf{X}, \qquad \widetilde{\mathbf{W}}_x \quad = \quad \mathbf{W}_x^{(0)} \sqcup \mathbf{W}_x^{(1)}. \tag{3.14}$$

Clearly, $\mathbf{W}_x^{(0)} \subset \mathbf{N}_x$. By Claim 9, $\Upsilon[\mathbf{N}_x] = 0$, for almost all $x \in \mathbf{X}$. Thus, $\mathbf{W}_x^{(0)}$ is a null set for almost all $x \in \mathbf{X}$. Also, by part **1(a)**, $\mathbf{W}_x^{(1)} \in \mathcal{Y}$ for all $x \in \mathbf{X}$. Thus, $\widetilde{\mathbf{W}}_x \;=\; \mathbf{W}_x^{(0)} \sqcup \mathbf{W}_x^{(1)}$ is $\mathcal{Y}$-measurable for almost all $x \in \mathbf{X}$. Furthermore,

$$\begin{aligned}
\widetilde{\zeta}[\widetilde{\mathbf{W}}] \;&=_{(a)}\; \zeta\left[\mathbf{W}^{(1)}\right] \;=_{(b)}\; \int_{\mathbf{X}} \Upsilon\left[\mathbf{W}_x^{(1)}\right]\, d\mu[x] \;=_{(c)}\; \int_{\mathbf{X}} \Upsilon\left[\mathbf{W}_x^{(1)}\right] + \Upsilon\left[\mathbf{W}_x^{(0)}\right]\, d\mu[x] \\
&= \int_{\mathbf{X}} \Upsilon\left[\mathbf{W}_x^{(1)} \sqcup \mathbf{W}_x^{(0)}\right]\, d\mu[x] \;=_{(d)}\; \int_{\mathbf{X}} \Upsilon\left[\widetilde{\mathbf{W}}_x\right]\, d\mu[x].
\end{aligned}$$

(a) By formula (3.13) and the definition of $\widetilde{\zeta}$. (b) Apply part **1(b)** to $\mathbf{W}^{(1)}$. (c) For almost all x, $\Upsilon\left[\mathbf{W}_x^{(0)}\right] = 0$, so apply the **Identity** property (Theorem 113 on page 74). (d) By formula (3.14)

Proof of 3(b,c):

Claim 10: *Suppose $f \in \mathbf{L}^1(\mathbf{Z}, \widetilde{\mathcal{Z}}, \zeta)$, and $f =_\zeta 0$. If $F : \mathbf{X} \longrightarrow \mathbb{C}$ is defined as in* **2(b)**, *then $F =_\xi 0$.*

Proof: <u>**Exercise 109**</u> .... □ [Claim 10]

<u>**Exercise 110**</u>: Apply Lemma 40 on page 31 to complete the proof. ________________□

<u>**Exercise 111**</u> Let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be the unit interval $[0,1]$ with Lebesgue measure λ and the (complete) Lebesgue sigma-algebra $\mathcal{L}$. Thus, $\mathbf{Z} = [0,1]^2$, $\mathcal{Z} = \mathcal{L} \otimes \mathcal{L}$ and $\zeta = \lambda \times \lambda$. Show that $\mathcal{L} \otimes \mathcal{L}$ is *not* complete by constructing an example of a set $\mathbf{W} \subset [0,1]^2$ of measure zero which has nonmeasurable subsets.

Theorem 127 concerns integration in a specific order over a product of two spaces, but this immediately generalizes to integration in any order, over any number of spaces...

Corollary 128 (Multifactor Fubini-Tonelli)

Let $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$, $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$, $\ldots$ $(\mathbf{X}_N, \mathcal{X}_N, \mu_N)$ be sigma-finite measure spaces, and let $(\mathbf{X}, \mathcal{X}, \mu)$ be their product, ie. $\mathbf{X} = \prod_{n=1}^{N} \mathbf{X}_n$, $\mathcal{X} = \bigotimes_{n=1}^{N} \mathcal{X}_n$, *and* $\mu = \bigotimes_{n=1}^{N} \mu_n$.

Let $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$. Then:

1. $\displaystyle\int_{\mathbf{X}} f(\mathbf{x})\, d\mu[\mathbf{x}] \;=\; \int_{\mathbf{X}_1} \int_{\mathbf{X}_2} \cdots \int_{\mathbf{X}_N} f(x_1, \ldots, x_N)\, d\mu_N[x_N] \ldots d\mu_2[x_2]\, d\mu_1[x_1]$

2. *If $\sigma : [1..N] \longrightarrow [1..N]$ is any permutation, then there is a natural isomorphism $(\mathbf{X}, \mathcal{X}, \mu) \cong (\widetilde{\mathbf{X}}, \widetilde{\mathcal{X}}, \widetilde{\mu})$, where* $\widetilde{\mathbf{X}} = \prod_{n=1}^{N} \mathbf{X}_{\sigma(n)}$, $\widetilde{\mathcal{X}} = \bigotimes_{n=1}^{N} \mathcal{X}_{\sigma(n)}$, *and* $\widetilde{\mu} = \bigotimes_{n=1}^{N} \mu_{\sigma(n)}$.

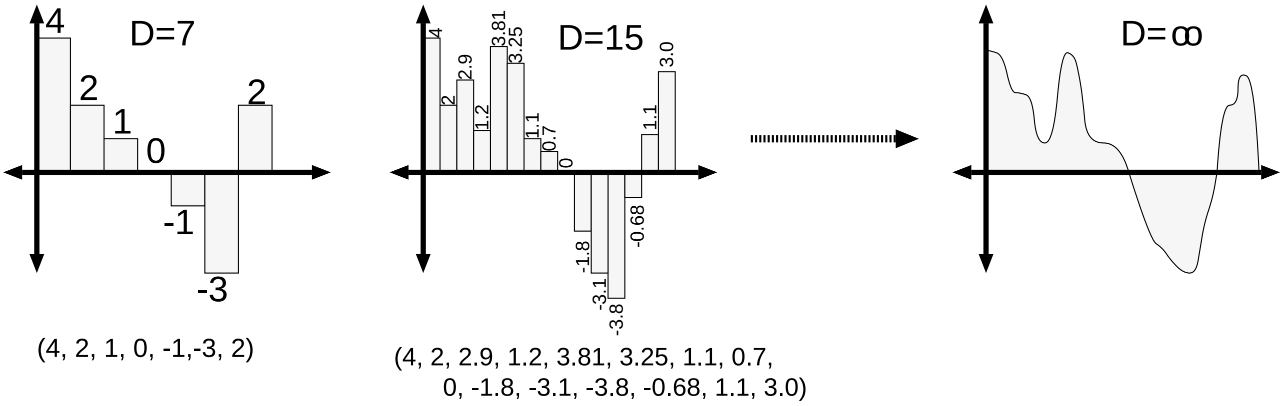


Figure 4.1: We can think of a function as an "infinite-dimensional vector"

3. $\displaystyle\int_{\mathbf{X}} f(\mathbf{x})\ d\mu[\mathbf{x}] \quad = \quad \int_{\mathbf{X}_{\sigma(1)}} \cdots \int_{\mathbf{X}_{\sigma(N)}} f(x_1, \ldots, x_N)\ d\mu_{\sigma(N)}[x_{\sigma(N)}]\ \ldots\ d\mu_{\sigma(1)}[x_{\sigma(1)}].$

Proof: **Exercise 112** □

4 Functional Analysis

4.1 Functions and Vectors

Vectors: If $\mathbf{v} = \begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -1.5 \\ 3 \\ 1 \end{bmatrix}$, then we can add these two vectors *componentwise*:

$$\mathbf{v} + \mathbf{w} \ = \ \begin{bmatrix} 2 - 1.5 \\ 7 + 3 \\ -3 + 1 \end{bmatrix} \ = \ \begin{bmatrix} 0.5 \\ 10 \\ -2 \end{bmatrix}.$$

In general, if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, then $\mathbf{u} = \mathbf{v} + \mathbf{w}$ is defined by:

$$u_n = v_n + w_n, \quad \text{for } n = 1, 2, 3 \tag{4.1}$$

(see Figure 4.2A)

Think of $\mathbf{v}$ as a function $v : \{1, 2, 3\} \longrightarrow \mathbb{R}$, where $v(1) = 2$, $v(2) = 7$, and $v(3) = -3$. In a similar fashion, any D-dimensional vector $\mathbf{u} = (u_1, u_2, \ldots, u_D)$ can be thought of as a function $u : [1...D] \longrightarrow \mathbb{R}$.

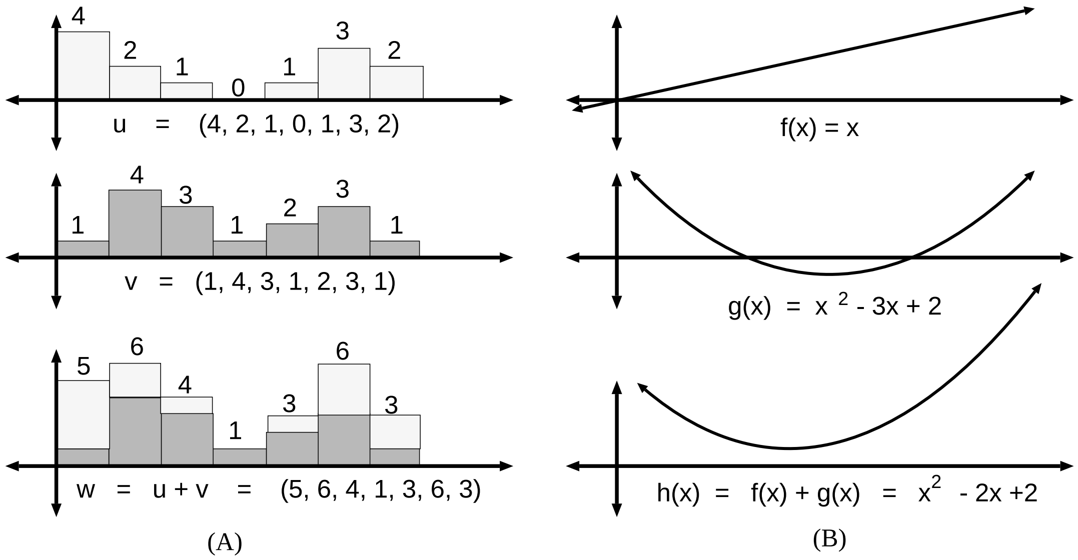


Figure 4.2: **(A)** We add vectors *componentwise*: If $\mathbf{u} = (4, 2, 1, 0, 1, 3, 2)$ and $\mathbf{v} = (1, 4, 3, 1, 2, 3, 1)$, then the equation "$\mathbf{w} = \mathbf{v} + \mathbf{w}$" means that $\mathbf{w} = (5, 6, 4, 1, 3, 6, 3)$. **(B)** We add two functions *pointwise*: If $f(x) = x$, and $g(x) = x^2 - 3x + 2$, then the equation "$h = f + g$" means that $h(x) = f(x) + g(x) = x^2 - 2x + 2$ for every x.

Functions as Vectors: Letting N go to infinity, we can imagine any function $f : \mathbb{R} \longrightarrow \mathbb{R}$ as a sort of "infinite-dimensional vector" (see Figure 4.1). Indeed, if f and g are two functions, we can add them *pointwise*, to get a new function $h = f + g$, where

$$h(x) = f(x) + g(x), \quad \text{for all } x \in \mathbb{R} \tag{4.2}$$

(see Figure 4.2B) Notice the similarity between formulae (4.2) and (4.1), and the similarity between Figures 4.2A and 4.2B.

The basic idea of functional analysis is that *functions are infinite-dimensional vectors*. Just as with finite vectors, we can add them together, act on them with linear operators, or represent them in different *coordinate systems* on infinite-dimensional space. Also, the vector space $\mathbb{R}^D$ has a natural geometric structure; we can identify a similar geometry in infinite dimensions.

Below are some examples of vector spaces of functions.

4.1(a) $\mathcal{C}$, the spaces of continuous functions

Let $\mathbf{X}$ be a topological space. We define:

$$\begin{aligned} \mathcal{C}(\mathbf{X}) &= \{f : \mathbf{X} \longrightarrow \mathbb{R}\ ;\ f \text{ is continuous}\} \\ \mathcal{C}(\mathbf{X};\ \mathbb{C}) &= \{f : \mathbf{X} \longrightarrow \mathbb{C}\ ;\ f \text{ is continuous}\} \\ \mathcal{C}(\mathbf{X};\ \mathbb{R}^n) &= \{f : \mathbf{X} \longrightarrow \mathbb{R}^n\ ;\ f \text{ is continuous}\} \end{aligned}$$

The **support** of f is the set $\mathsf{supp}[f] = \{x \in \mathbf{X} \ ; \ f(x) \neq 0\}$; if f is continuous, then $\mathsf{supp}[f]$ is a closed subset of $\mathbf{X}$. We say f has **compact support** if $\mathsf{supp}[f]$ is compact, and define

$$\mathcal{C}_c(\mathbf{X}) \quad = \quad \{f : \mathbf{X} \longrightarrow \mathbb{R} \ ; \ f \text{ is continuous, with compact support}\}.$$

We say that f **vanishes at** ∞ if, for any $\epsilon > 0$, the set $\{x \in \mathbf{X} \ ; \ |f(x)| \geq \epsilon\}$ is compact. Notice that this definition makes sense even when the space $\mathbf{X}$ has no well-defined "∞" point. We then define

$$\mathcal{C}_0(\mathbf{X}) \quad = \quad \{f : \mathbf{X} \longrightarrow \mathbb{R} \ ; \ f \text{ is continuous, and vanishes at } \infty\}.$$

Finally, say that f is **bounded** if $\sup_{x \in \mathbf{X}} |f(x)| \ < \ \infty$. Then define

$$\mathcal{C}_b(\mathbf{X}) \quad = \quad \{f : \mathbf{X} \longrightarrow \mathbb{R} \ ; \ f \text{ is continuous and bounded}\}.$$

Of course, we can also define $\mathcal{C}_0(\mathbf{X};\ \mathbb{C})$, etc. Observe that

$$\mathcal{C}_c(\mathbf{X}) \ \subset \ \mathcal{C}_0(\mathbf{X}) \ \subset \ \mathcal{C}_b(\mathbf{X}) \ \subset \ \mathcal{C}(\mathbf{X})$$

and, when $\mathbf{X}$ is compact, all of these spaces are equal (**Exercise 113**).

Exercise 114 Verify that $\mathcal{C}_c$, $\mathcal{C}_0$, $\mathcal{C}_b$ and $\mathcal{C}$ are all vector spaces —ie. that each is closed under the operations of pointwise addition and scalar multiplication.

4.1(b) $\mathcal{C}^n$, the spaces of differentiable functions

Let $\mathbf{X} \subset \mathbb{R}^n$ be some open subset. Recall that $f : \mathbf{X} \longrightarrow \mathbb{R}$ is **continuously differentiable** if f is everywhere differentiable and the derivative $\nabla f : \mathbf{X} \longrightarrow \mathbb{R}^n$ is continuous. More generally, $f : \mathbf{X} \longrightarrow \mathbb{R}^m$ is **continuously differentiable** if f is differentiable everywhere on $\mathbf{X}$ and the derivative $\mathcal{D}f : \mathbf{X} \longrightarrow \mathbb{R}^{n \times m}$ is continuous. Finally, recall that f is **analytic** if it has a power series that converges everywhere on $\mathbf{X}$. We define:

$$\begin{aligned}
\mathcal{C}^1(\mathbf{X}) \quad &= \quad \{f : \mathbf{X} \longrightarrow \mathbb{R} \ ; \ f \text{ is continuously differentiable}\}. \\
\mathcal{C}^1(\mathbf{X}; \mathbb{R}^m) \quad &= \quad \{f : \mathbf{X} \longrightarrow \mathbb{R}^m \ ; \ f \text{ is continuously differentiable}\}. \\
\mathcal{C}^k(\mathbf{X}; \mathbb{R}^m) \quad &= \quad \{f : \mathbf{X} \longrightarrow \mathbb{R}^m \ ; \ f \text{ is } k \text{ times continuously differentiable}\}. \\
\mathcal{C}^\infty(\mathbf{X}; \mathbb{R}^m) \quad &= \quad \{f : \mathbf{X} \longrightarrow \mathbb{R}^m \ ; \ f \text{ is infinitely often continuously differentiable}\}. \\
\mathcal{C}^\omega(\mathbf{X}; \mathbb{R}^m) \quad &= \quad \{f : \mathbf{X} \longrightarrow \mathbb{R}^m \ ; \ f \text{ is analytic}\}.
\end{aligned}$$

We say that $f : \mathbf{X} \longrightarrow \mathbb{R}$ is a **polynomial of degree** K if

$$f(x_1, x_2, \ldots, x_n) \quad = \quad \sum_{k_1 + k_2 + \ldots + k_n \leq K} a_{k_1 k_2 \ldots k_n} \ x_1^{k_1} x_1^{k_n} \ldots x_1^{k_n},$$

where $a_{k_1 k_2 \ldots k_n}$ are all real constants. We define

$$\begin{aligned} \mathcal{P}^k(\mathbf{X}) &= \{f : \mathbf{X} \longrightarrow \mathbb{R} \; ; \; f \text{ is a polynomial of degree at most } k\} \\ \text{and} \;\; \mathcal{P}(\mathbf{X}) &= \{f : \mathbf{X} \longrightarrow \mathbb{R} \; ; \; f \text{ is a polynomial of any finite degree}\} \end{aligned}$$

Observe that $\mathcal{P}^k(\mathbb{R})$ is a vector space of dimension $k+1$; more generally, if $\mathbf{X} \subset \mathbb{R}^n$, then $\mathcal{P}^k$ has a dimension that grows of order $\mathcal{O}(n^k)$. Thus, $\mathcal{P}(\mathbf{X})$ is infinite-dimensional. Also, note that

$$\mathcal{P}^1 \subset \mathcal{P}^2 \subset \mathcal{P}^3 \subset \ldots \subset \mathcal{P} \subset \mathcal{C}^\omega \subset \mathcal{C}^\infty \subset \ldots \subset \mathcal{C}^3 \subset \mathcal{C}^2 \subset \mathcal{C}^1 \subset \mathcal{C}$$

and all of these inclusions are proper.

Exercise 115 Verify that each of $\mathcal{P}^n$, $\mathcal{P}$, $\mathcal{C}^n$, $\mathcal{C}^\infty$ and $\mathcal{C}^\omega$ is a vector space —ie. that it is closed under the operations of pointwise addition and scalar multiplication.

4.1(c) L, the space of measurable functions

Prerequisites: §1.3(a)

Let $(\mathbf{X}, \mathcal{X})$ be a measurable space. We define

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathcal{X}) &= \{f : \mathbf{X} \longrightarrow \mathbb{R} \; ; \; f \text{ is measurable}\} \\ \mathbf{L}(\mathbf{X}, \mathcal{X}; \; \mathbb{C}) &= \{f : \mathbf{X} \longrightarrow \mathbb{C} \; ; \; f \text{ is measurable}\} \end{aligned}$$

etc. Suppose $\mathbf{X}$ is a topological space and $\mathcal{X}$ is the Borel sigma algebra. Then:

$$\mathcal{C}(\mathbf{X}) \subset \mathbf{L}(\mathbf{X}). \qquad (\textbf{Exercise 116})$$

Exercise 117 Verify that $\mathbf{L}(\mathbf{X}, \mathcal{X})$ is a vector space.

4.1(d) $\mathbf{L}^1$, the space of integrable functions

Prerequisites: §4.1(c), §**??**

If μ is a measure on $(\mathbf{X}, \mathcal{X})$, and $f \in \mathbf{L}(\mathbf{X}, \mathcal{X})$, then we define the **$\mathbf{L}^1$-norm** of f by:

$$\|f\|_1 = \int_{\mathbf{X}} |f|(x) \, d\mu[x] < \infty \qquad \text{(see Figure 4.3)}$$

and then define the space of **integrable functions:**

$$\begin{aligned} \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu) &= \{f \in \mathbf{L}(\mathbf{X}, \mathcal{X}) \; ; \; \|f\|_1 < \infty\} \\ \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu; \; \mathbb{C}) &= \{f \in \mathbf{L}(\mathbf{X}, \mathcal{X}; \; \mathbb{C}) \; ; \; \|f\|_1 < \infty\} \end{aligned}$$

Exercise 118 Verify that $\mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$ is a vector space.

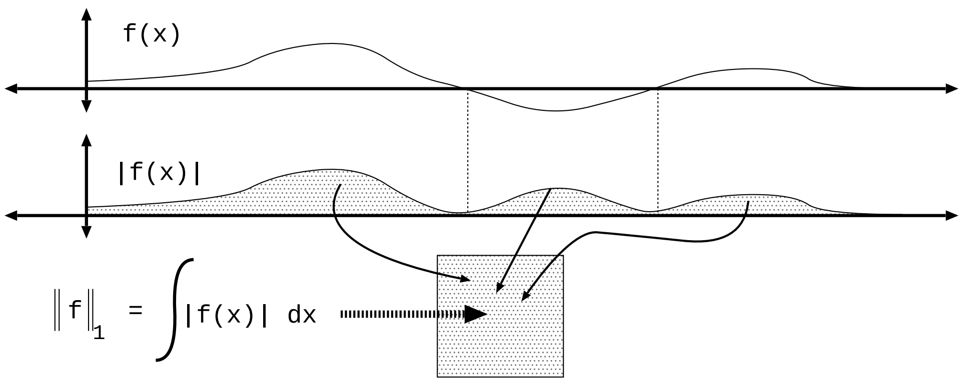


Figure 4.3: The $\mathbf{L}^1$ norm of f is defined: $\|f\|_1 \;=\; \int_{\mathbb{X}} |f(x)|\; d\mu[x]$.

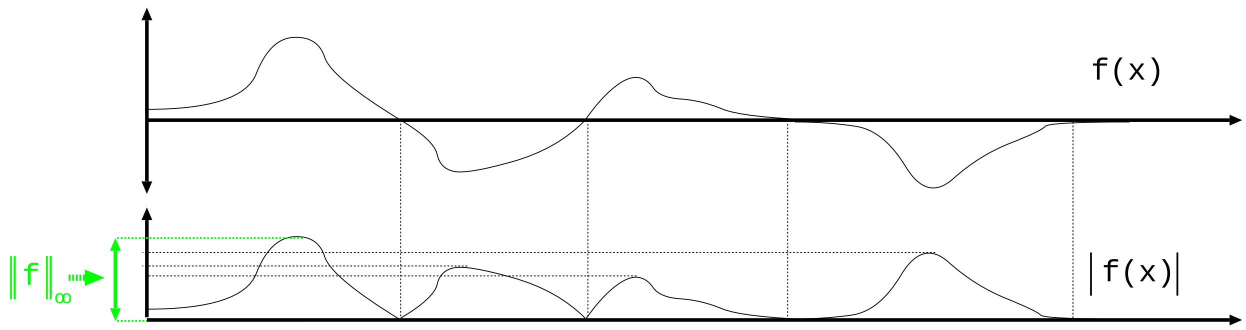


Figure 4.4: The $\mathbf{L}^\infty$ norm of f is defined: $\|f\|_\infty \;=\; \operatorname{ess\,sup}_{x\in\mathbb{X}} |f(x)|$.

Suppose that $\mathbf{X}$ is a topological space and $\mathcal{X}$ is the Borel sigma algebra. If μ is a finite measure, then

$$\mathcal{C}_b(\mathbf{X}) \subset \mathbf{L}^1(\mathbf{X}). \qquad (\textbf{Exercise 119})$$

If μ is an infinite measure, we say that μ is **locally finite** if $\mu(\mathbf{K}) < \infty$ for any compact $\mathbf{K} \subset \mathbf{X}$. In this case

$$\mathcal{C}_c(\mathbf{X}) \subset \mathbf{L}^1(\mathbf{X}). \qquad (\textbf{Exercise 120})$$

4.1(e) $\mathbf{L}^\infty$, the space of essentially bounded functions

Prerequisites: §4.1(c), §1.3(c)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and $f \in \mathbf{L}(\mathbf{X}, \mathcal{X})$. For any $c \in \mathbb{R}$, let $\mathbf{S}_c = \{x \in \mathbb{R} \;;\; f(x) > c\} = f^{-1}(c, \infty)$. If $\mu(\mathbf{S}_c) = 0$, then f is μ-almost everywhere less than c; we might say that f is **essentially** less than c. We define the **essential supremum** of

$$\operatorname*{ess\,sup}_{x\in\mathbf{X}} f(x) \;=\; \min\{c \in \mathbb{R} \;;\; \mu[\mathbf{S}_c] = 0\}$$

For example, suppose we define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by $f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}; \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$ Then, with respect to the Lebesgue measure, $\operatorname*{ess\,sup}_{x \in \mathbb{R}} f(x) = 0$, because $\mathbb{Q}$ has measure zero.

Most of the time, the essential supremum is the same as the supremum. For example, if $\mathbf{X}$ is an open subset of $\mathbb{R}^n$ with the Lebesgue measure, and $f : \mathbf{X} \longrightarrow \mathbb{R}$ is continuous, then $\operatorname*{ess\,sup}_{x \in \mathbf{X}} f(x) = \sup_{x \in \mathbf{X}} f(x)$ (**Exercise 121**).

More generally, of $\mathbf{X}$ is any measure space, and $\operatorname*{ess\,sup}_{x \in \mathbf{X}} f(x) = c$, then there is a function $\widetilde{f}$ such that $\widetilde{f} = f$ almost everywhere, and $\sup_{x \in \mathbf{X}} \widetilde{f}(x) = c$. Thus, in accord with the idea that sets of measure zero are 'negligible', we will often identify f with $\widetilde{f}$, and identify $\operatorname*{ess\,sup}_{x \in \mathbf{X}} f(x)$ as the 'supremum' of f.

If $f : \mathbf{X} \longrightarrow \mathbb{R}$ or $f : \mathbf{X} \longrightarrow \mathbb{C}$, we define the $\mathbf{L}^\infty$**-norm** of f by:

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x \in \mathbf{X}} |f(x)| \qquad \text{(see Figure 4.4)}$$

We then define the set of **essentially bounded** functions:

$$\begin{aligned} \mathbf{L}^\infty(\mathbf{X}, \mathcal{X}, \mu) &= \{f \in \mathbf{L}(\mathbf{X}, \mathcal{X}) \; ; \; \|f\|_\infty < \infty\} \\ \mathbf{L}^\infty(\mathbf{X}, \mathcal{X}, \mu; \; \mathbb{C}) &= \{f \in \mathbf{L}(\mathbf{X}, \mathcal{X}; \; \mathbb{C}) \; ; \; \|f\|_\infty < \infty\} \end{aligned}$$

Exercise 122 Verify that $\mathbf{L}^\infty(\mathbf{X}, \mathcal{X}, \mu)$ is a vector space.

When $\mathbf{X}$, $\mathcal{X}$, or μ are clear from context, we may drop one, two, or all three from the notation, writing, for example "$\mathbf{L}^\infty(\mathbf{X})$" or "$\mathbf{L}^1(\mu)$", depending on what is being emphasised.

Suppose that $\mathbf{X}$ is a topological space and $\mathcal{X}$ is the Borel sigma algebra. Then

$$\mathcal{C}_b(\mathbf{X}) \;\subset\; \mathbf{L}^\infty(\mathbf{X}, \mu). \qquad (\textbf{Exercise 123})$$

4.2 Inner Products (infinite-dimensional geometry)

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$, then the **inner product**[1] of $\mathbf{x}, \mathbf{y}$ is defined:

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_D y_D.$$

The inner product describes the geometric relationship between $\mathbf{x}$ and $\mathbf{y}$, via the formula:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos(\theta)$$

where $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are the *lengths* of vectors $\mathbf{x}$ and $\mathbf{y}$, and θ is the angle between them. (**Exercise 124** Verify this). In particular, if $\mathbf{x}$ and $\mathbf{y}$ are *perpendicular*, then $\theta = \pm\frac{\pi}{2}$, and then

[1] This is sometimes this is called the **dot product**, and denoted "$\mathbf{x} \bullet \mathbf{y}$".

$\langle \mathbf{x}, \mathbf{y} \rangle = 0$; we then say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal**. For example, $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\mathbf{y} = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ are orthogonal in $\mathbb{R}^2$, while

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

are all orthogonal to one another in $\mathbb{R}^4$. Indeed, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ also have unit norm; we call any such collection an **orthonormal set** of vectors. Thus, $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is an orthonormal set, but $\{\mathbf{x}, \mathbf{y}\}$ is not.

The **norm** of a vector satisfies the equation:

$$\|\mathbf{x}\| \;=\; \left(x_1^2 + x_2^2 + \ldots + x_D^2\right)^{1/2} \;=\; \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}.$$

If $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are a collection of mutually orthogonal vectors, and $\mathbf{x} = \mathbf{x}_1 + \ldots + \mathbf{x}_N$, then we have the generalized **Pythagorean formula**:

$$\|\mathbf{x}\|^2 \;=\; \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \ldots + \|\mathbf{x}_N\|^2$$

(**Exercise 125** Verify the Pythagorean formula.)

An **orthonormal basis** of $\mathbb{R}^D$ is any collection of mutually orthogonal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_D\}$, all of norm 1, so that, for any $\mathbf{w} \in \mathbb{R}^D$, if we define $\omega_d = \langle \mathbf{w}, \mathbf{v}_d \rangle$ for all $d \in [1..D]$, then:

$$\mathbf{w} = \omega_1 \mathbf{v}_1 + \omega_2 \mathbf{v}_2 + \ldots + \omega_D \mathbf{v}_D$$

In this case, the Pythagorean Formula becomes **Parseval's Equality:**

$$\|\mathbf{w}\|^2 = \omega_1^2 + \omega_2^2 + \ldots + \omega_D^2$$

(**Exercise 126** Deduce Parseval's equality from the Pythagorean formula.)

Example:

1. $\left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^D$.

2. If $\mathbf{v}_1 = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthonormal basis of $\mathbb{R}^2$.

   If $\mathbf{w} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, then $\omega_1 = \sqrt{3} + 2$ and $\omega_2 = 1 - 2\sqrt{3}$, so that

   $$\begin{bmatrix} 2 \\ 4 \end{bmatrix} \;=\; \omega_1 \mathbf{v}_1 + \omega_2 \mathbf{v}_2 \;=\; \left(\sqrt{3} + 2\right) \cdot \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix} \;+\; \left(1 - 2\sqrt{3}\right) \cdot \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}.$$

   Thus, $\|\mathbf{w}\|_2^2 \;=\; 2^2 + 4^2 \;=\; 20$, and also, by Parseval's equality, $20 \;=\; \omega_1^2 + \omega_2^2 \;=\; \left(\sqrt{3} + 2\right)^2 + \left(1 - 2\sqrt{3}\right)^2$. (**Exercise 127** Verify these claims.)

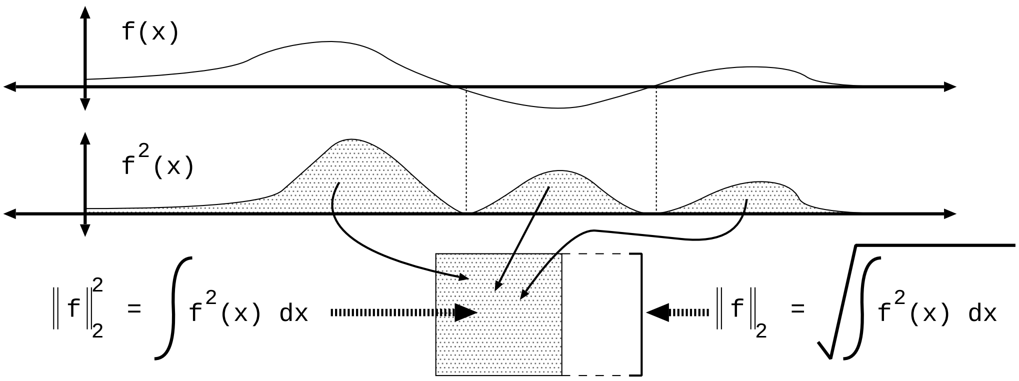


Figure 4.5: The $\mathbf{L}^2$ norm of f: $\|f\|_2 = \sqrt{\int_{\mathbf{X}} |f(x)|^2 \, d\mu[x]}$

4.3 $\mathbf{L}^2$ space

Prerequisites: §4.2, §**??**

All of this generalizes to spaces of functions. Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, and $f, g \in \mathbf{L}(\mathbf{X}, \mathcal{X}; \mathbb{C})$, then the **inner product** of f and g is defined:

$$\langle f, g \rangle \;=\; \frac{1}{M} \int_{\mathbf{X}} f(x) \cdot \overline{g(x)} \; d\mu[x]$$

Here, $\overline{g(x)}$ is the *complex conjugate* of $g(x)$. If $g(x) \in \mathbb{R}$, then of course $\overline{g(x)} = g(x)$. Meanwhile, $M = \displaystyle\int_{\mathbf{X}} 1 \; d\mathbf{x}$ is the *total mass* of $\mathbf{X}$.

For example, suppose $\mathbf{X} = [0, 3] = \{x \in \mathbb{R} \; ; \; 0 \le x \le 3\}$, with the Lebesgue measure λ. Thus $M = \lambda[0, 3] = 3$. If $f(x) = x^2 + 1$ and $g(x) = x$ for all $x \in [0, 3]$, then

$$\langle f, g \rangle \;=\; \frac{1}{3} \int_0^3 f(x) g(x) \; dx \;=\; \frac{1}{3} \int_0^3 (x^3 + x) \; dx \;=\; \frac{27}{4} + \frac{3}{2}.$$

The $\mathbf{L}^2$**-norm** of a function $f \in \mathbf{L}(\mathbf{X}, \mathcal{X})$ is defined

$$\|f\|_2 \;=\; \langle f, f \rangle^{1/2} \;=\; \left(\frac{1}{M} \int_{\mathbf{X}} |f|^2(x) \; d\mu[x] \right)^{1/2}. \tag{4.3}$$

(see Figure 4.5).

Note that the definition of $\mathbf{L}^2$-norm in (4.3) depends upon the choice of measure μ. Also, note that the integral in (4.3) *may not converge.* For example, if $f \in \mathbf{L}[0, 1]$ is defined: $f(x) = 1/x$, then $\|f\|_2 = \infty$ (relative to the Lebesgue measure).

The set of all measurable functions on $(\mathbf{X}, \mathcal{X})$ with finite $\mathbf{L}^2$-norm is denoted $\mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$, and called $\mathbf{L}^2$**-space**. For example, any bounded, continuous function $f : [0, 1] \longrightarrow \mathbb{R}$ is in $\mathbf{L}^2([0, 1], \lambda)$.

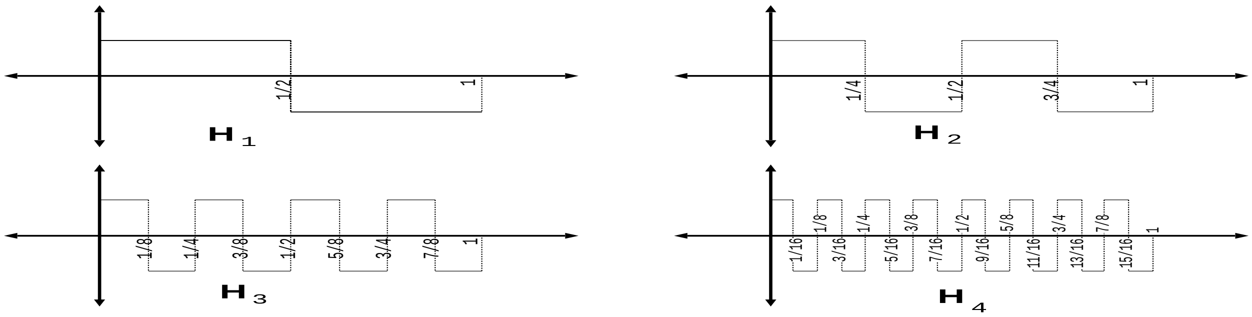


Figure 4.6: Four Haar basis elements: $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{H}_4$

4.4 Orthogonality

Prerequisites: §4.3

Two functions $f, g \in \mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$ are **orthogonal** if $\langle f, g \rangle = 0$. For example, if $\mathbf{X} = [-\pi, \pi]$ with Lebesgue measure, and sin and cos are treated as elements of $\mathbf{L}^2[-\pi, \pi]$, then they are orthogonal:

$$\langle \sin, \cos \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(x) \cos(x) \ dx \ = \ 0. \tag{4.4}$$

(**Exercise 128**) An **orthogonal set** of functions is a set $\{f_1, f_2, f_3, \ldots\}$ of elements in $\mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$ so that $\langle f_j, f_k \rangle = 0$ whenever $j \neq k$. If, in addition, $\|f_j\|_2 = 1$ for all j, then we say this is an **orthonormal set** of functions.

Example 129:

(a) Let $\mathbf{X} = [0, 1]$ with Lebesgue measure. Figure 4.6 portrays the **The Haar Basis**. We define $\mathbf{H}_0 \equiv 1$, and for any natural number $N \in \mathbb{N}$, we define the Nth **Haar function** $\mathbf{H}_N : [0, 1] \longrightarrow \mathbb{R}$ by:

$$\mathbf{H}_N(x) = \begin{cases} 1 & \text{if } \dfrac{2n}{2^N} \leq \ x \ < \dfrac{2n+1}{2^N}, \quad \text{for some } n \in \big[0...2^{N-1}\big); \\[2ex] -1 & \text{if } \dfrac{2n+1}{2^N} \leq \ x \ < \dfrac{2n+2}{2^N}, \quad \text{for some } n \in \big[0...2^{N-1}\big). \end{cases}$$

Then $\{\mathbf{H}_0, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \ldots\}$ is an orthonormal set in $\mathbf{L}^2[0, 1]$ (**Exercise 129**).

(b) Let $\mathbf{X} = [0, 1]$ with Lebesgue measure. Figure 4.7 portrays a **Wavelet Basis.** We define

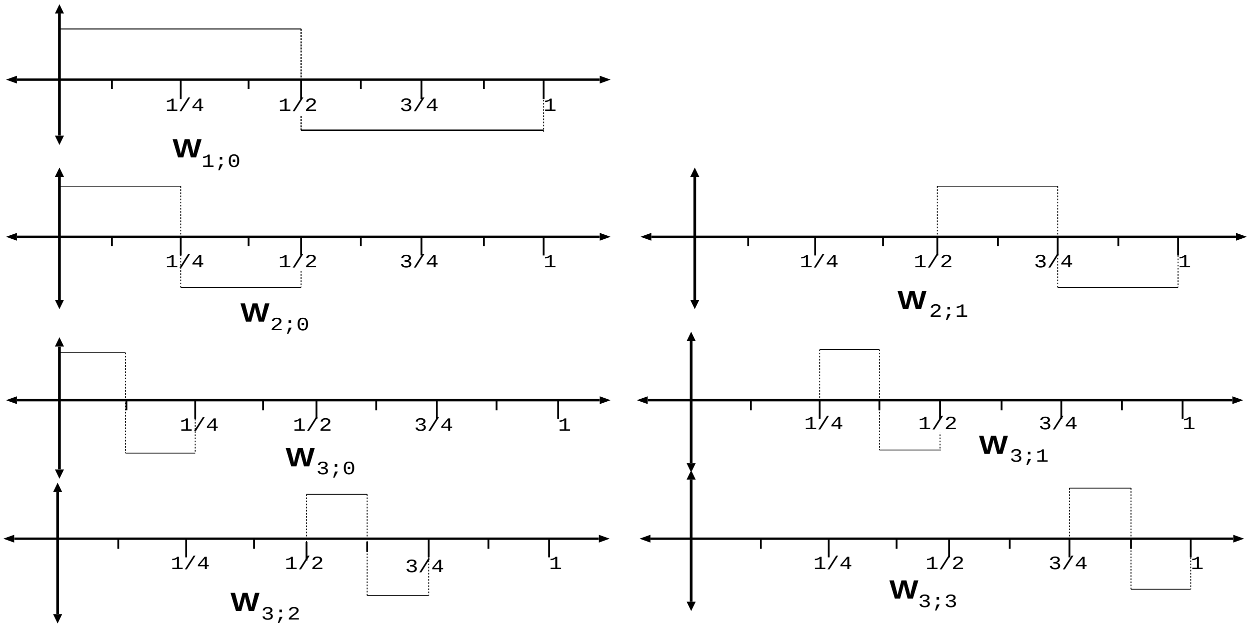


Figure 4.7: Seven Wavelet basis elements: $\mathbf{W}_{1,0}$; $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$; $\mathbf{W}_{3,0}$, $\mathbf{W}_{3,1}, \mathbf{W}_{3,2}$, $\mathbf{W}_{3,3}$

$\mathbf{W}_0 \equiv 1$, and for any $N \in \mathbb{N}$ and $n \in \left[0...2^{N-1}\right)$, we define

$$\mathbf{W}_{n;N}(x) = \begin{cases} 1 & \text{if } \dfrac{2n}{2^N} \leq x < \dfrac{2n+1}{2^N}; \\ -1 & \text{if } \dfrac{2n+1}{2^N} \leq x < \dfrac{2n+2}{2^N}; \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\{\mathbf{W}_0; \mathbf{W}_{1,0}; \mathbf{W}_{2,0}, \mathbf{W}_{2,1}; \mathbf{W}_{3,0}, \mathbf{W}_{3,1}, \mathbf{W}_{3,2}, \mathbf{W}_{3,3}; \mathbf{W}_{4,0}, \ldots, \mathbf{W}_{4,7}; \mathbf{W}_{5,0}, \ldots, \mathbf{W}_{5,15}; \ldots\}$$

is an *orthogonal* set in $\mathbf{L}^2[0,1]$, but is *not* orthonormal: for any N and n, we have $\|\mathbf{W}_{n;N}\|_2 = \frac{1}{2^{(N-1)/2}}$. (**Exercise 130**).

4.4(a) Trigonometric Orthogonality

Fourier analysis is based on the orthogonality of certain families of trigonometric functions. Formula (4.4) on page 102 was an example of this; this formula generalizes as follows....

Proposition 130: Trigonometric Orthogonality on $[-\pi, \pi]$

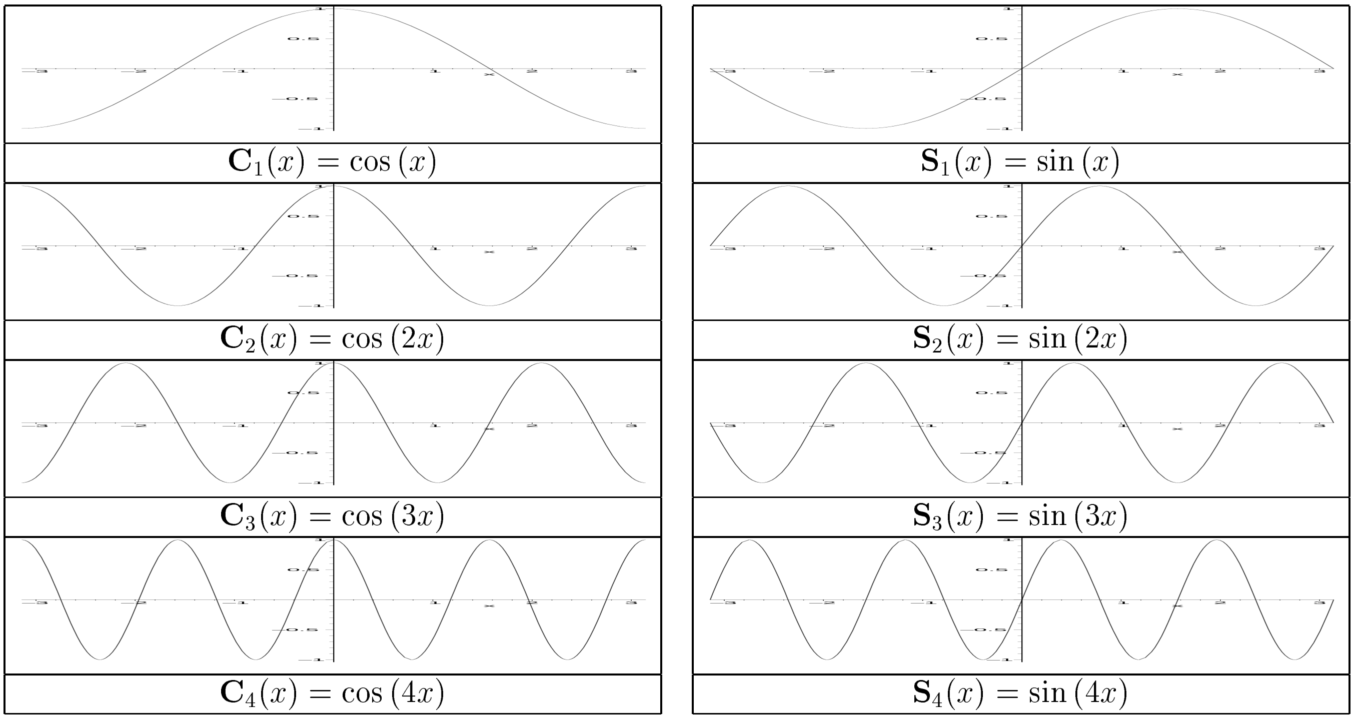


Figure 4.8: $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{C}_4$; $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, and $\mathbf{S}_4$

Let $\mathbf{X} = [-\pi, \pi]$ *with Lebesgue measure. For every* $n \in \mathbb{N}$, *define* $\mathbf{S}_n(x) = \sin(nx)$ *and* $\mathbf{C}_n(x) = \cos(nx)$. *(see Figure 4.8).*

The set $\{\mathbf{C}_0, \mathbf{C}_1, \mathbf{C}_2, \ldots; \mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots\}$ *is an* **orthogonal set** *of functions for* $\mathbf{L}^2[-\pi, \pi]$. *In other words:*

- $\langle \mathbf{S}_n, \mathbf{S}_m \rangle \;=\; \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \sin(nx)\sin(mx)\; dx \;=\; 0$, *whenever* $n \neq m$.
- $\langle \mathbf{C}_n, \mathbf{C}_m \rangle \;=\; \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \cos(nx)\cos(mx)\; dx \;=\; 0$, *whenever* $n \neq m$.
- $\langle \mathbf{S}_n, \mathbf{C}_m \rangle \;=\; \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \sin(nx)\cos(mx)\; dx = 0$, *for any* n *and* m.

However, these functions are not orthonormal, *because they do* not *have unit norm. Instead, for any* $n \neq 0$,

$$\|\mathbf{C}_n\|_2 \;=\; \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(nx)^2\; dx} \;=\; \frac{1}{\sqrt{2}}, \quad \text{and} \quad \|\mathbf{S}_n\|_2 \;=\; \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin(nx)^2\; dx} \;=\; \frac{1}{\sqrt{2}}.$$

Proof: **Exercise 131** Hint: Use the trigonometric identities: $2\sin(\alpha)\cos(\beta) = \sin(\alpha+\beta) + \sin(\alpha-\beta)$, $2\sin(\alpha)\sin(\beta) = \cos(\alpha-\beta) - \cos(\alpha+\beta)$, and $2\cos(\alpha)\cos(\beta) = \cos(\alpha+\beta) + \cos(\alpha-\beta)$. □

Remark: Notice that $\mathbf{C}_0(x) = 1$ is just the *constant* function.

It is important to remember that the statement, "f and g are orthogonal" depends upon the *domain* $\mathbf{X}$ which we are considering. For example, when $L = \pi$ in the following theorem, note that $\mathbf{S}_n(x) = \sin(nx)$ and $\mathbf{C}_n(x) = \cos(nx)$ just as in the previous theorem. However, the orthogonality relations are *different*, because the domain $\mathbf{X}$ is different.

Proposition 131: Trigonometric Orthogonality on $[0, L]$

Let $\mathbf{X} = [0, L]$ with Lebesgue measure. Let $L > 0$, and, for every $n \in \mathbb{N}$, define $\mathbf{S}_n(x) = \sin\left(\frac{n\pi x}{L}\right)$ and $\mathbf{C}_n(x) = \cos\left(\frac{n\pi x}{L}\right)$.

1. *The set $\{\mathbf{C}_0, \mathbf{C}_1, \mathbf{C}_2, \ldots\}$ is an* **orthogonal set** *of functions for $\mathbf{L}^2[0, L]$. In other words:* $\langle \mathbf{C}_n, \mathbf{C}_m \rangle \;=\; \frac{1}{L}\int_0^L \cos\left(\frac{n\pi}{L}x\right)\cos\left(\frac{m\pi}{L}x\right)\,dx \;=\; 0$, *whenever $n \neq m$.*
   However, these functions are **not** *orthonormal, because they do* not *have unit norm. Instead, for any $n \neq 0$,* $\|\mathbf{C}_n\|_2 \;=\; \sqrt{\frac{1}{L}\int_0^L \cos\left(\frac{n\pi}{L}x\right)^2\,dx} \;=\; \frac{1}{\sqrt{2}}$.

2. *The set $\{\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots\}$ is an* **orthogonal set** *of functions for $\mathbf{L}^2[0, L]$. In other words:* $\langle \mathbf{S}_n, \mathbf{S}_m \rangle \;=\; \frac{1}{L}\int_0^L \sin\left(\frac{n\pi}{L}x\right)\sin\left(\frac{m\pi}{L}x\right)\,dx \;=\; 0$, *whenever $n \neq m$.*
   However, these functions are **not** *orthonormal, because they do* **not** *have unit norm. Instead, for any $n \neq 0$,* $\|\mathbf{S}_n\|_2 \;=\; \sqrt{\frac{1}{L}\int_0^L \sin\left(\frac{n\pi}{L}x\right)^2\,dx} \;=\; \frac{1}{\sqrt{2}}$.

3. *The functions $\mathbf{C}_n$ and $\mathbf{S}_m$ are* **not** *orthogonal to one another on $[0, L]$. Instead:*

$$\langle \mathbf{S}_n, \mathbf{C}_m \rangle \;=\; \frac{1}{L}\int_0^L \sin\left(\frac{n\pi}{L}x\right)\cos\left(\frac{m\pi}{L}x\right)\,dx \;=\; \begin{cases} 0 & \text{if } n+m \text{ is even} \\ \dfrac{2n}{\pi(n^2-m^2)} & \text{if } n+m \text{ is odd.} \end{cases}$$

Proof: **Exercise 132** . □

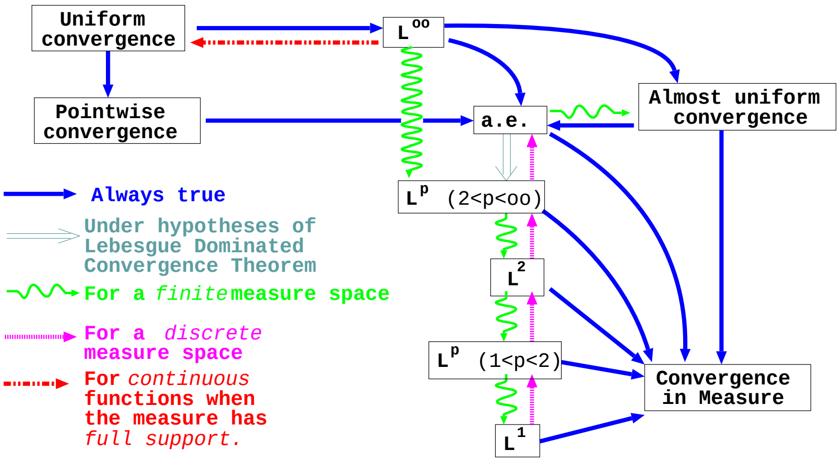


Figure 4.9: The lattice of convergence types. Here, "$A \Longrightarrow B$" means that convergence of type A implies convergence of type B, under the stipulated conditions.

4.5 Convergence Concepts

New

If $\{x_1, x_2, x_3, \ldots\}$ is a sequence of numbers, we know what it means to say "$\lim_{n\to\infty} x_n = x$". $\{f_1, f_2, f_3, \ldots\}$ was a sequence of functions, and f was some other function, then we might want to say that "$\lim_{n\to\infty} f_n = f$". Think of convergence as a kind of 'approximation'. Heuristically speaking, if the sequence $\{x_n\}_{n=1}^\infty$ converges to x, then, for very large n, the number x_n is *approximately equal to* x. Thus, heuristically speaking, if the sequence $\{f_n\}_{n=1}^\infty$ 'converges' to f, then, for very large n, the function f_n is a *good approximation* of f.

However, there are many ways we can interpret 'good approximation', and these lead to different notions of 'convergence'. Thus, convergence of *functions* is a much more subtle concept that convergence of *numbers*. We will now discuss the many different 'modes of convergence' for functions. Figure 4.9 illustrates their logical relationships.

4.5(a) Pointwise Convergence

Let $\mathbf{X}$ be any set, and let $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be functions for all $n \in \mathbb{N}$. We say the sequence $\{f_1, f_2, \ldots\}$ converges **pointwise** to $f : \mathbf{X} \longrightarrow \mathbb{R}$ if

$$\lim_{n\to\infty} f_n(x) \;=\; f(x), \qquad \text{for every } x \in \mathbf{X}. \qquad \text{(see Figure 4.10)}$$

Example 132: Let $\mathbf{X} = [0, 1]$.

(a) For all $n \in \mathbb{N}$, let $g_n(x) = \dfrac{1}{1 + n \cdot \left|x - \frac{1}{2n}\right|}$. Figure 4.11 on the facing page portrays functions g_1, g_5, g_{10}, g_{15}, g_{30}, and g_{50}; These picture strongly suggest that the sequence is con-

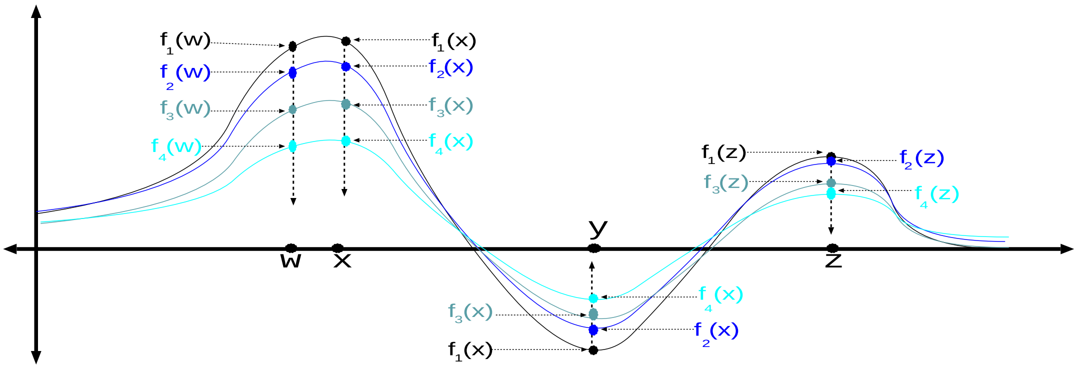
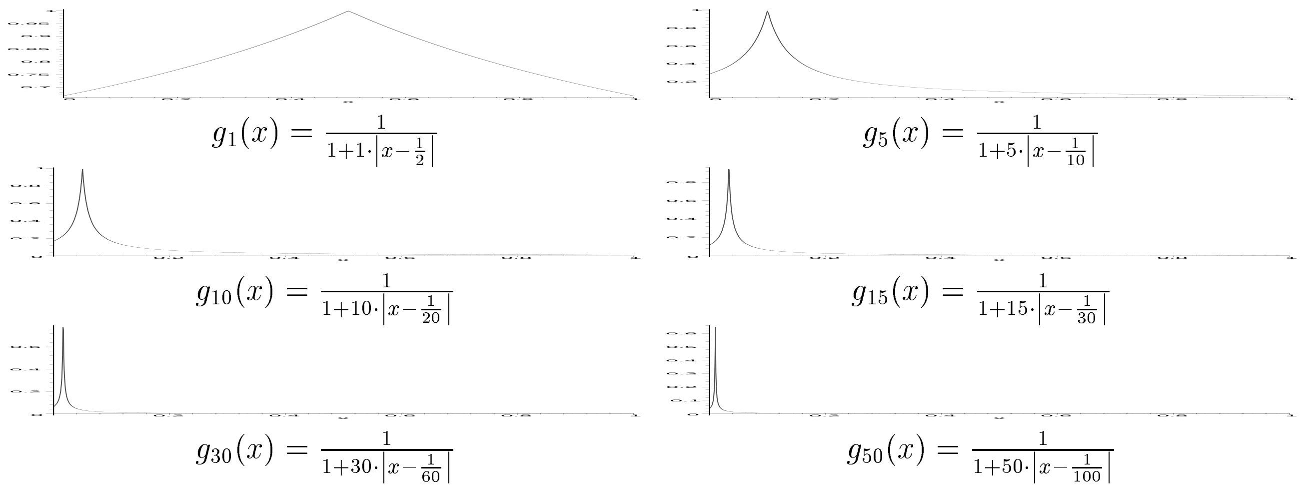


Figure 4.10: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges **pointwise** to the constant 0 function. Thus, if we pick some random points $w, x, y, z \in \mathbf{X}$, then we see that $\lim_{n\to\infty} f_n(w) = 0$, $\lim_{n\to\infty} f_n(x) = 0$, $\lim_{n\to\infty} f_n(y) = 0$, and $\lim_{n\to\infty} f_n(z) = 0$.



$g_1(x) = \frac{1}{1+1\cdot\left|x-\frac{1}{2}\right|}$ $\quad$ $g_5(x) = \frac{1}{1+5\cdot\left|x-\frac{1}{10}\right|}$

$g_{10}(x) = \frac{1}{1+10\cdot\left|x-\frac{1}{20}\right|}$ $\quad$ $g_{15}(x) = \frac{1}{1+15\cdot\left|x-\frac{1}{30}\right|}$

$g_{30}(x) = \frac{1}{1+30\cdot\left|x-\frac{1}{60}\right|}$ $\quad$ $g_{50}(x) = \frac{1}{1+50\cdot\left|x-\frac{1}{100}\right|}$

Figure 4.11: If $g_n(x) = \frac{1}{1+n\cdot\left|x-\frac{1}{2n}\right|}$, then the sequence $\{g_1, g_2, g_3, \ldots\}$ converges pointwise to the constant 0 function on $[0,1]$, and also converges to 0 in $\mathbf{L}^1[0,1]$, $\mathbf{L}^2[0,1]$, and $\mathbf{L}^\infty[0,1]$ but does *not* converge to 0 uniformly.

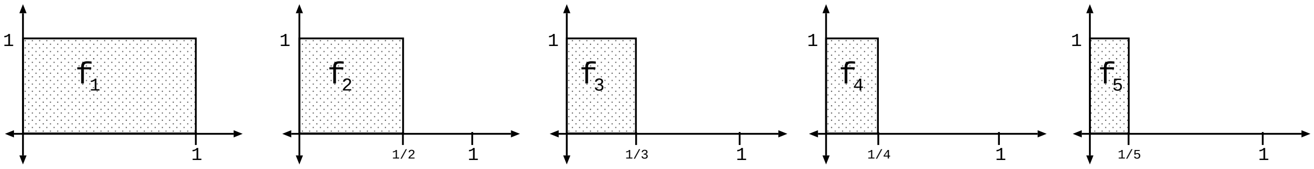


Figure 4.12: Example (132b).

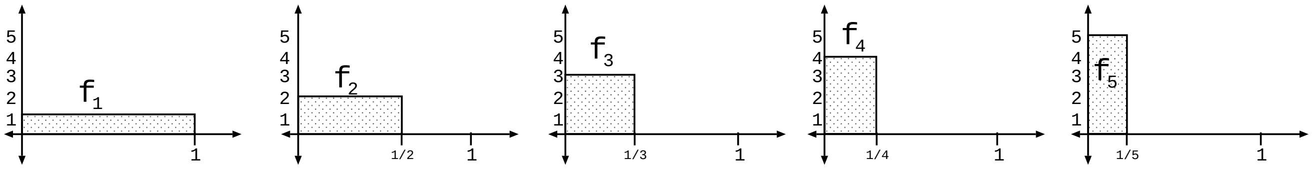


Figure 4.13: Example (132c).

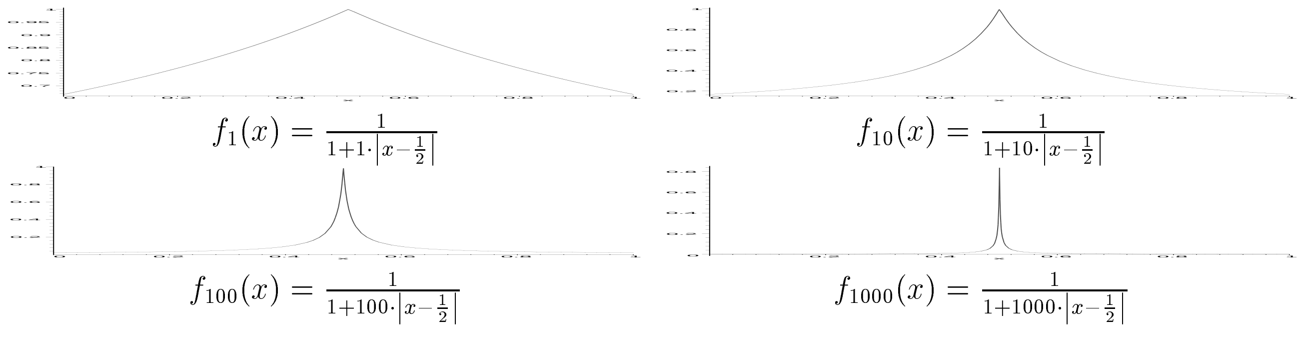
verging pointwise to the constant 0 function on $[0,1]$. The proof of this is **Exercise 133** .

(b) For each $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} 0 & \text{if } x = 0; \\ 1 & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$ (Figure 4.12).

This sequence converges pointwise to the constant 0 function on $[0,1]$ (**Exercise 134**).

(c) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} 0 & \text{if } x = 0; \\ n & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$ (Figure 4.13). Then this sequence converges pointwise to the constant 0 function.

(d) For each $n \in \mathbb{N}$, let $f_n(x) = \dfrac{1}{1 + n \cdot \left|x - \frac{1}{2}\right|}$ (see Figure 4.14). This sequence of does *not*



$f_1(x) = \frac{1}{1+1\cdot\left|x-\frac{1}{2}\right|}$ $f_{10}(x) = \frac{1}{1+10\cdot\left|x-\frac{1}{2}\right|}$

$f_{100}(x) = \frac{1}{1+100\cdot\left|x-\frac{1}{2}\right|}$ $f_{1000}(x) = \frac{1}{1+1000\cdot\left|x-\frac{1}{2}\right|}$

Figure 4.14: If $f_n(x) = \frac{1}{1+n\cdot\left|x-\frac{1}{2}\right|}$, then the sequence $\{f_1, f_2, f_3, \ldots\}$ converges to the constant 0 function in $\mathbf{L}^2[0,1]$.

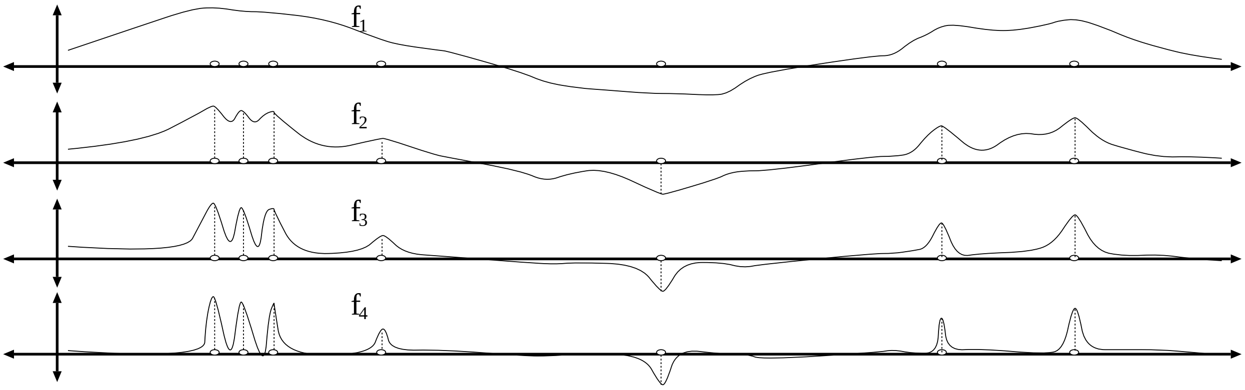


Figure 4.15: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges **almost everywhere** to the constant 0 function.

converge to zero pointwise, because $f_n(\frac{1}{2}) = 1$ for all $n \in \mathbb{N}$.

Let $(\mathbf{Y}, d)$ be a metric space, and suppose that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ and $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ for all $n \in \mathbb{N}$. If we define $g_n(x) = d\Big(f(x), f_n(x)\Big)$ for all $n \in \mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **pointwise** if $g_n \xrightarrow[n\to\infty]{} 0$ pointwise. Hence, to understand pointwise convergence in general, it is sufficient to understand pointwise convergence of nonnegative functions to the constant 0 function.

4.5(b) Almost-Everywhere Convergence

Prerequisites: §1.3(c) **Recommended:** §4.5(a)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure-space, and let $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be measurable functions for all $n \in \mathbb{N}$. We say the sequence $\{f_1, f_2, \ldots\}$ converges **almost everywhere** to $f : \mathbf{X} \longrightarrow \mathbb{R}$ if there is a set $\mathbf{X}_0 \subset \mathbf{X}$ such that $\mu[\mathbf{X} \setminus \mathbf{X}_0] = 0$, and

$$\lim_{n\to\infty} f_n(x) \;=\; f(x), \qquad \text{for every } x \in \mathbf{X}_0. \qquad \text{(see Figure 4.15)}$$

Example 133: Let $\mathbf{X} = [0, 1]$ and let μ be the Lebesgue measure.

(a) Let $f_n : [0, 1] \longrightarrow \mathbb{R}$ be as in Example (132d) on page 108. Then f_n does *not* converge to 0 pointwise, but f_n *does* converge to 0 almost everywhere (relative to Lebesgue measure).

(b) For all $n \in \mathbb{N}$, let $f_n(x) \;=\; \begin{cases} 1 & \text{if } x \in \mathbb{Q}; \\ \frac{1}{n} & \text{if } x \notin \mathbb{Q}. \end{cases}$ Then f_n does *not* converge to 0 pointwise, but f_n *does* converge to 0 almost everywhere (relative to Lebesgue measure).

Let $(\mathbf{Y}, d)$ be a metric space, and suppose that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ and $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ for all $n \in \mathbb{N}$. If we define $g_n(x) = d\Big(f(x), f_n(x)\Big)$ for all $n \in \mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **a.e.** if $g_n \xrightarrow[n\to\infty]{} 0$ a.e. Hence, to understand a.e. convergence in general, it is sufficient to understand a.e. convergence of nonnegative functions to the constant 0 function.

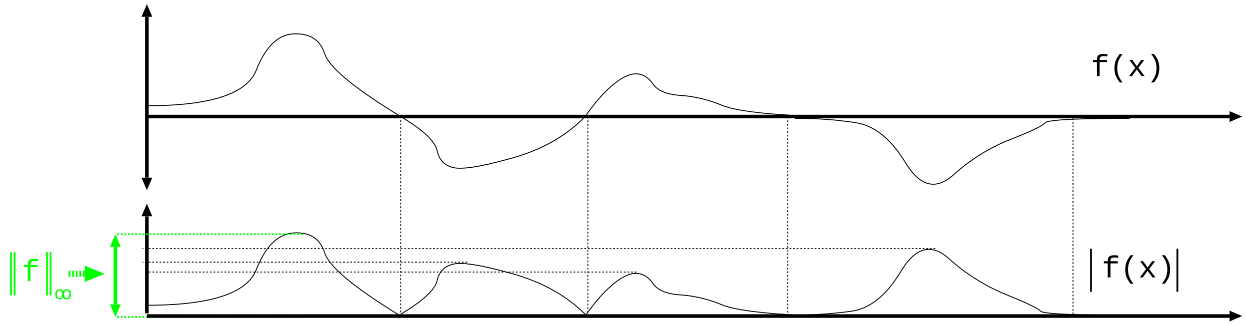


Figure 4.16: The uniform norm of f is given: $\|f\|_{\mathrm{u}} = \sup_{x \in \mathbf{X}} |f(x)|$.

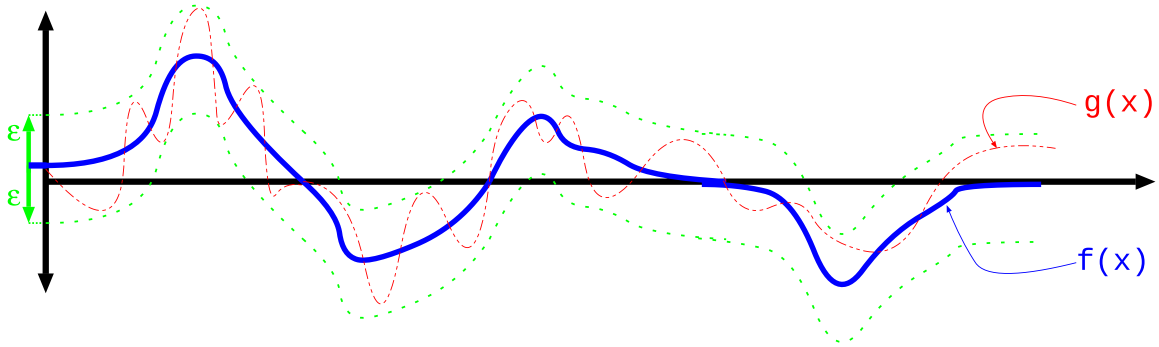


Figure 4.17: If $\|f-g\|_{\mathrm{u}} < \epsilon$, then $g(x)$ is confined within an 'ϵ-tube' around f.

4.5(c) Uniform Convergence

Prerequisites: §??

Let $\mathbf{X}$ be any set, and let $f : \mathbf{X} \longrightarrow \mathbb{R}$. The **uniform norm** of a function f is defined:

$$\|f\|_{\mathrm{u}} = \sup_{x \in \mathbf{X}} \left| f(x) \right|$$

This measures the *farthest deviation* of the function f from zero (see Figure 4.16).

Example: Suppose $\mathbf{X} = [0,1]$, and $f(x) = \frac{1}{3}x^3 - \frac{1}{4}x$ (as in Figure 4.18A). This function takes its minimum at $x = \frac{1}{2}$, where it has the value $\frac{-1}{12}$. Thus, $|f(x)|$ takes a maximum of $\frac{1}{12}$ at this point, so that $\|f\|_{\mathrm{u}} = \sup_{0 \leq x \leq 1} \left| \frac{1}{3}x^3 - \frac{1}{4}x \right| = \frac{1}{12}$.

The **uniform distance** between two functions f and g is then given by:

$$\|f-g\|_{\mathrm{u}} = \sup_{x \in \mathbf{X}} \left| f(x) - g(x) \right|$$

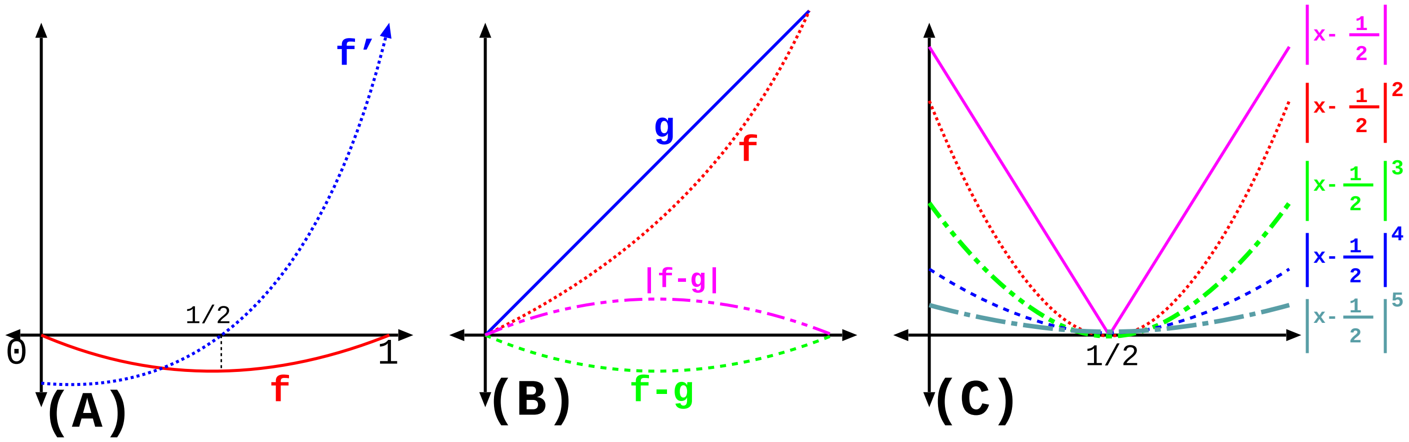


Figure 4.18: **(A)** The uniform norm of $f(x) = \frac{1}{3}x^3 - \frac{1}{4}x$. **(B)** The uniform distance between $f(x) = x(x+1)$ and $g(x) = 2x$. **(C)** $g_n(x) = \left|x - \frac{1}{2}\right|^n$, for $n = 1, 2, 3, 4, 5$.

One way to interpret this is portrayed in Figure 4.17. Define a "tube" of width ϵ around the function f. If $\|f - g\|_{\mathrm{u}} < \epsilon$, this means that $g(x)$ is confined within this tube for all $x \in \mathbf{X}$.

Example: Let $\mathbf{X} = [0,1]$, and suppose $f(x) = x(x+1)$ and $g(x) = 2x$ (as in Figure 4.18B). For any $x \in [0,1]$, $|f(x) - g(x)| = |x^2 + x - 2x| = |x^2 - x| = x - x^2$. This expression takes its maximum at $x = \frac{1}{2}$ (to see this, take the derivative), and its value at $x = \frac{1}{2}$ is $\frac{1}{4}$. Thus, we conclude: $\|f - g\|_{\mathrm{u}} = \sup_{x \in \mathbf{X}} \left|x(x-1)\right| = \dfrac{1}{4}$.

A sequence of functions $\{g_1, g_2, g_3, \ldots\}$ **converges uniformly** to f if $\lim_{n\to\infty} \|g_n - f\|_{\mathrm{u}} = 0$. This means not only that $\lim_{n\to\infty} g_n(x) = f(x)$ for every $x \in \mathbf{X}$, but furthermore, that the functions g_n converge to f everywhere at the same "speed". This is portrayed in Figure 4.19. For any $\epsilon > 0$, we can define a "tube" of width ϵ around f, and, no matter how small we make this tube, the sequence $\{g_1, g_2, g_3, \ldots\}$ will eventually enter this tube and remain there. To be precise: there is some N so that, for all $n > N$, the function g_n is confined within the ϵ-tube around f —ie. $\|f - g_n\|_{\mathrm{u}} < \epsilon$.

Example 134: Let $\mathbf{X} = [0,1]$.

(a) If $g_n(x) = 1/n$ for all $x \in [0,1]$, then the sequence $\{g_1, g_2, \ldots\}$ converges to zero uniformly on $[0,1]$ (**Exercise 135**).

(b) If $g_n(x) = \left|x - \frac{1}{2}\right|^n$ (see Figure 4.18C), then the sequence $\{g_1, g_2, \ldots\}$ converges to zero uniformly on $[0,1]$ (**Exercise 136**).

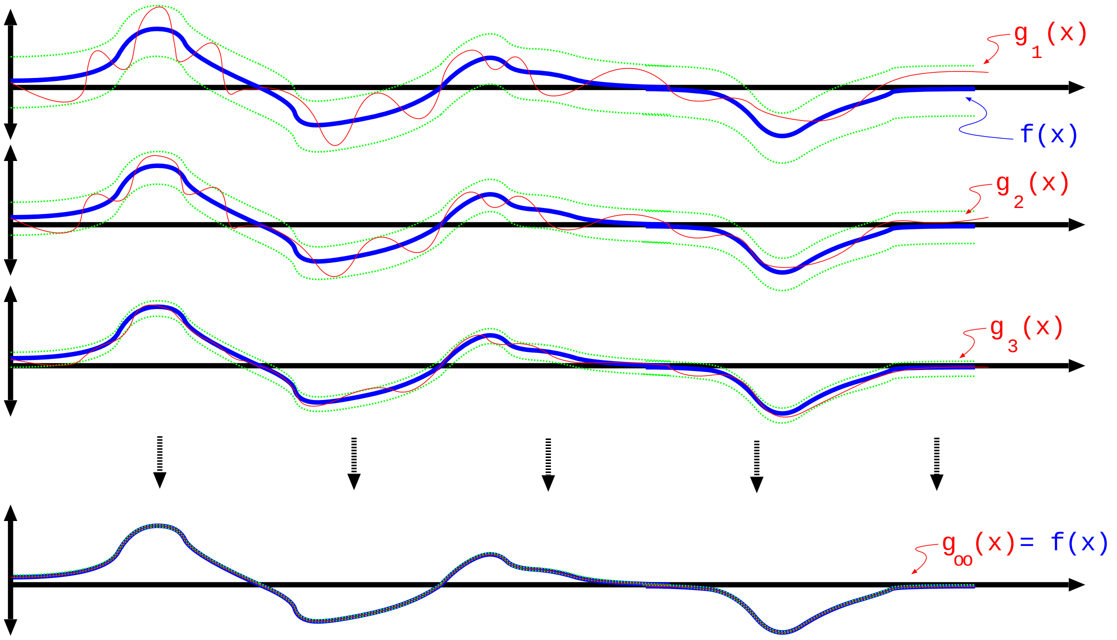


Figure 4.19: The sequence $\{g_1, g_2, g_3, \ldots\}$ converges **uniformly** to f.

(c) Recall the functions $g_n(x) = \frac{1}{1+n\cdot\left|x-\frac{1}{2n}\right|}$. from Example (132a) on page 106. The sequence $\{g_1, g_2, \ldots\}$ converges *pointwise* to the uniform zero function, but does *not* converge to zero uniformly on $[0,1]$. However, for any $\delta > 0$, the sequence $\{g_1, g_2, \ldots\}$ *does* converge to zero uniformly on $[\delta, 1]$ (**Exercise 137** Verify these claims.).

(d) Suppose, as in Example (132b) on page 108, that $g_n(x) = \begin{cases} 0 & \text{if } x = 0; \\ 1 & \text{if } 0 < x < \frac{1}{n}; \\ 0 & \text{otherwise.} \end{cases}$.

Then the sequence $\{g_1, g_2, \ldots\}$ converges *pointwise* to the uniform zero function, but does *not* converge to zero uniformly on $[0,1]$ (**Exercise 138**).

Proposition 135: *Suppose $f_n : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous for all $n \in \mathbb{N}$, and $f : \mathbb{R} \longrightarrow \mathbb{R}$. If $f_n \xrightarrow[n\to\infty]{} f$ uniformly, then f is also continuous.*

Proof: **Exercise 139** □

Let $(\mathbf{Y}, d)$ be a metric space, and suppose that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ and $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ for all $n \in \mathbb{N}$. If we define $g_n(x) = d\Big(f(x), f_n(x)\Big)$ for all $n \in \mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **uniformly** if $g_n \xrightarrow[n\to\infty]{} 0$ uniformly. Hence, to understand uniform convergence in general, it is sufficient to understand uniform convergence of nonnegative functions to the constant 0 function. In this context, Proposition 135 is a special case of:

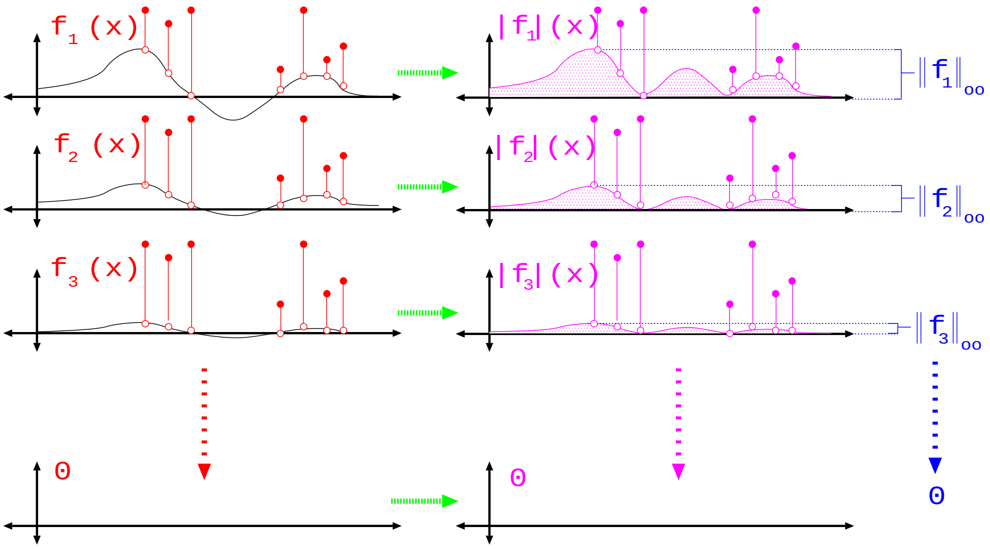


Figure 4.20: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges to the constant 0 function in $\mathbf{L}^\infty(\mathbf{X})$.

Proposition 136: *Let* $\mathbf{X}$ *be a topological space and let* $(\mathbf{Y}, d)$ *be a metric space. Suppose* $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ *is continuous for all* $n \in \mathbb{N}$*, and* $f : \mathbf{X} \longrightarrow \mathbf{Y}$*. If* $f_n \xrightarrow[n\to\infty]{} f$ *uniformly, then* f *is also continuous.*

Proof: **Exercise 140** □

4.5(d) $\mathbf{L}^\infty$ Convergence

Prerequisites: §4.1(e), §1.3(c) **Recommended:** §4.5(c)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f : \mathbf{X} \longrightarrow \mathbb{R}$ be measurable. Recall that the $\mathbf{L}^\infty$ **norm** of a function f is defined:

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x\in\mathbf{X}} \left|f(x)\right| \qquad \text{(see Figure 4.4 on page 98)}$$

This measures the farthest 'essential' deviation of the function f from zero.

The $\mathbf{L}^\infty$ **distance** between two functions f and g is then given by:

$$\|f - g\|_\infty \;=\; \operatorname*{ess\,sup}_{x\in\mathbf{X}} \left|f(x) - g(x)\right|$$

Suppose that $f_n \in \mathbf{L}^\infty(\mathbf{X}, \mathcal{X}, \mu)$ for all $n \in \mathbb{N}$. We say that the sequence $\{f_n\}_{n=1}^\infty$ **converges in** $\mathbf{L}^\infty$ to f if $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. This not only means that $\lim_{n\to\infty} f_n(x) = f(x)$ for μ-almost every $x \in \mathbf{X}$, but furthermore, that the functions f_n converge to f almost everywhere at the same "speed". This is portrayed in Figure 4.20. For any $\epsilon > 0$, we can define a "tube" of width ϵ around f, and, no matter how small we make this tube, the sequence $\{g_1, g_2, g_3, \ldots\}$ will eventually enter this tube and remain there. To be precise: there is some N so that, for all $n > N$, the function g_n is confined within the ϵ-tube around f almost everywhere. —ie. $\|f - g_n\|_\infty < \epsilon$.

Example 137: Let $\mathbf{X} = [0, 1]$ and let μ be the Lebesgue measure.

(a) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}; \\ \frac{1}{n} & \text{if } x \notin \mathbb{Q}. \end{cases}$ Then f_n does *not* converge to 0 uniformly (or even pointwise), f_n *does* converge to 0 in $\mathbf{L}^\infty[0, 1]$.

(b) Let $f_n : [0, 1] \longrightarrow \mathbb{R}$ be as in Examples (132a) or (132d) on page 108. Then f_n does *not* converge uniformly to 0, but *does* converge to 0 in $\mathbf{L}^\infty[0, 1]$.

Remark: Observe that the $\mathbf{L}^\infty$ norm is *not* the same as the uniform norm, and $\mathbf{L}^\infty$ convergence is *not* the same as uniform convergence. The difference is that, in $\mathbf{L}^\infty$ convergence, we allow convergence to *fail* on a set of measure zero. Thus,

$$\Big(\text{uniform convergence} \Big) \Longrightarrow \Big(\mathbf{L}^\infty \text{ convergence} \Big),$$

but the opposite is not true, in general. However, we do have:

Proposition 138: *Suppose $f_n : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuous for all $n \in \mathbb{N}$. and $f : \mathbf{X} \longrightarrow \mathbb{R}$. Then*

$$\Big(f_n \xrightarrow[n\to\infty]{} f \text{ uniformly} \Big) \iff \Big(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty(\mathbb{R}, \lambda) \Big)$$

Proof: <u>**Exercise 141**</u> □

Let $(\mathbf{Y}, d)$ be a metric space, and suppose that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ and $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ for all $n \in \mathbb{N}$. If we define $g_n(x) = d\Big(f(x), f_n(x)\Big)$ for all $n \in \mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **in** $\mathbf{L}^\infty$ if $g_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^\infty$. Hence, to understand $\mathbf{L}^\infty$ convergence in general, it is sufficient to understand $\mathbf{L}^\infty$ convergence of nonnegative functions to the constant 0 function. In this context, Proposition 138 is a special case of:

Proposition 139: *Let $\mathbf{X}$ be a topological space, and let μ be a Borel measure on $\mathbf{X}$ with full support. Let $(\mathbf{Y}, d)$ a metric space, and suppose $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ is continuous for all $n \in \mathbb{N}$, and $f : \mathbf{X} \longrightarrow \mathbf{Y}$. Then* $\Big(f_n \xrightarrow[n\to\infty]{} f \text{ uniformly} \Big) \iff \Big(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty(\mathbf{X}, \mu) \Big)$.

Proof: <u>**Exercise 142**</u> □

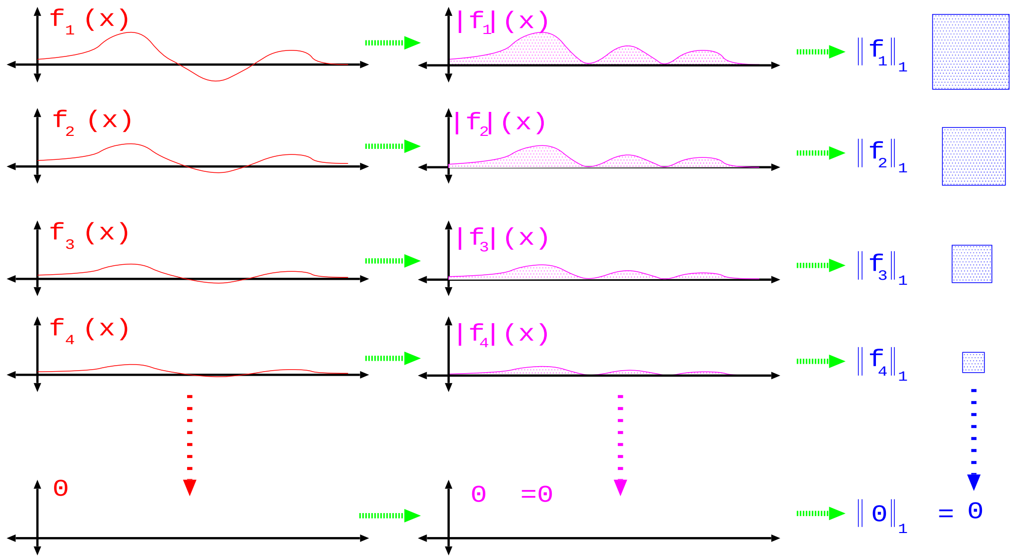


Figure 4.21: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges to the constant 0 function in $\mathbf{L}^1(\mathbf{X})$.

4.5(e) Almost Uniform Convergence

Prerequisites: §4.5(c), §1.3(c)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be functions for all $n \in \mathbb{N}$. We say the sequence $\{f_1, f_2, \ldots\}$ converges **almost uniformly** to $f : \mathbf{X} \longrightarrow \mathbb{R}$ if, for every $\epsilon > 0$, there is a subset $\mathbf{E} \subset \mathbf{X}$ such that

- $\mu[\mathbf{X} \setminus \mathbf{E}] < \epsilon$;
- $\displaystyle \lim_{n\to\infty} \left(\sup_{e\in\mathbf{E}} \left| f_n(e) - f(e) \right| \right) = 0$; in other words f_n converges uniformly to f inside $\mathbf{E}$.

Theorem 140: Egoroff

If $(\mathbf{X}, \mathcal{X}, \mu)$ is a finite measure space, then

$$\left(f_n \xrightarrow[n\to\infty]{} f \text{ almost everywhere} \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ almost uniformly} \right)$$

4.5(f) $\mathbf{L}^1$ convergence

Prerequisites: §4.1(d),§3.2 **Recommended:** §4.5(d)

If $f, g \in \mathbf{L}^1(\mathbf{X}, \mu)$, then the **$\mathbf{L}^1$-distance** between f and g is just

$$\|f - g\|_1 \;=\; \int_{\mathbf{X}} \left| f(x) - g(x) \right| \, d\mu[x]$$

If we think of f as an "approximation" of g, then $\|f - g\|_1$ measures the *average error* of this approximation. If $\{f_1, f_2, f_3, \ldots\}$ is a sequence of successive approximations of f, then we say the sequence **converges to f in $\mathbf{L}^1$** if $\lim_{n\to\infty} \|f_n - f\|_1 \;=\; 0$. See Figure 4.21.

Example 141: Let $\mathbf{X} = \mathbb{R}$, with the Lebesgue measure.

(a) For each $n \in \mathbb{N}$, let $f_n(x) \;=\; \begin{cases} 1 & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$ as in Example (132b) on page 108. Then $\|f_n\|_1 = \frac{1}{n}$, so $f_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^1$. Observe, however, that f_n does *not* converge uniformly to zero.

(b) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$, as in Example (132c) on page 108. This sequence converges *pointwise* to zero, but $\|f_n\|_1 = 1$ for all n, so the sequence does *not* converge to zero in $\mathbf{L}^1$.

(c) Perhaps $\mathbf{L}^1$-convergence fails in the previous example because the sequence is unbounded? For all $n \in \mathbb{N}$, let $f_n(x) \;=\; \begin{cases} 1 & \text{if } n < x < n+1; \\ 0 & \text{otherwise} \end{cases}$. This sequence is *bounded*, and converges *pointwise* to zero, but $\|f_n\|_1 = 1$ for all n, so the sequence does *not* converge to zero in $\mathbf{L}^1$.

(d) For any $n \in \mathbb{N}$, let $\ell(n) = \lfloor \log_2(n) \rfloor$ and find $k \in [0..2^m)$ so that $n = 2^{\ell(n)} + k$ (for example, $38 = 2^5 + 6$). Then define $\mathbf{U}_n \;=\; \left[\frac{k}{2^{\ell(n)}}, \frac{k+1}{2^{\ell(n)}}\right] \subset [0,1]$. (for example, $\mathbf{U}_{38} \;=\; \left[\frac{6}{32}, \frac{7}{32}\right]$), and define $f_n = \mathbb{1}_{\mathbf{U}_n}$ (Figure 4.22). Thus, $\|f_n\|_1 = \dfrac{1}{2^{\ell(n)}} \xrightarrow[n\to\infty]{} 0$, so that f_n converges to zero in $\mathbf{L}^1$. However, f_n does *not* converge to zero pointwise.

(e) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} \frac{1}{n} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$. This sequence converges *uniformly* to zero, but $\|f_n\|_1 = 1$ for all n, so the sequence does *not* converge to zero in $\mathbf{L}^1$.

(f) For all $n \in \mathbb{N}$, let $f_n(x) \;=\; \begin{cases} \frac{1}{n^2} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$. Then $\|f_n\|_1 \;=\; \frac{1}{n}$, so this sequence *does* converge to zero in $\mathbf{L}^1$.

Example 142: $\ell^1(\mathbb{N})$ vs. $\ell^\infty(\mathbb{N})$

Let $\mathbb{R}^{\mathbb{N}}$ be the set of all sequences of real numbers; we'll write such a sequence as $\mathbf{a} = [a_n]_{n=1}^\infty$. Recall that $\|\mathbf{a}\|_1 = \sum_{n=1}^{\infty} |a_n|$ and $\ell^1(\mathbb{N}) \;=\; \left\{\mathbf{a} \in \mathbb{R}^{\mathbb{N}} \;;\; \|\mathbf{a}\|_1 < \infty\right\}$. Similarly, $\|\mathbf{a}\|_\infty \;=\; \sup_{n\in\mathbb{N}} |a_n|$

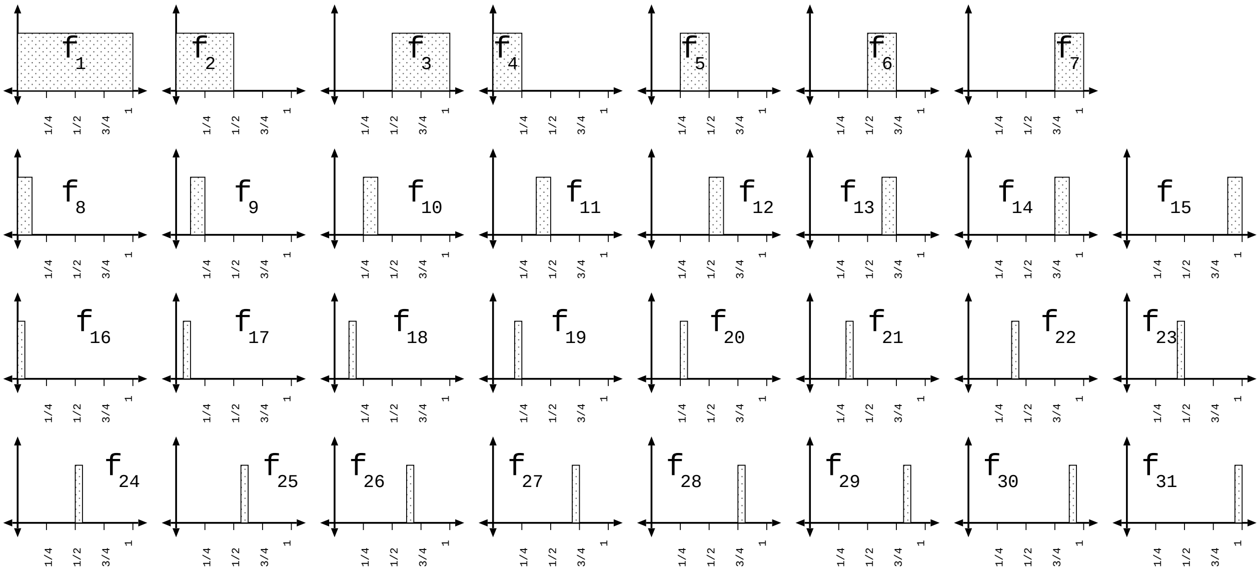


Figure 4.22: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges to the constant 0 function in $\mathbf{L}^1(\mathbf{X})$, but not pointwise.

and $\ell^\infty(\mathbb{N}) = \left\{\mathbf{a} \in \mathbb{R}^{\mathbb{N}} \ ; \ \|\mathbf{a}\|_\infty < \infty\right\}$. Clearly, for any sequence, $\sup_{n\in\mathbb{N}} |a_n| \ \leq \ \sum_{n=1}^{\infty} |a_n|$. It follows that $\ell^1(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$, and $\left(\mathbf{a}_n \xrightarrow[n\to\infty]{} \mathbf{a} \text{ in } \ell^1 \right) \implies \left(\mathbf{a}_n \xrightarrow[n\to\infty]{} \mathbf{a} \text{ in } \ell^\infty \right)$.

This is a special case of the following result:

Proposition 143: *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f : \mathbf{X} \longrightarrow \mathbb{R}$ and $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be integrable for all $n \in \mathbb{N}$.*

1. *If $(\mathbf{X}, \mathcal{X}, \mu)$ is* **discrete**, *and $m = \inf_{\substack{\mathbf{U}\in\mathcal{X} \\ \mu[\mathbf{U}]>0}} \mu[\mathbf{U}] \ > \ 0$, then:*

   (a) $\|f\|_\infty \ \leq \ \dfrac{1}{m}\|f\|_1$

   (b) $\mathbf{L}^1(\mathbf{X}, \mu) \ \subset \ \mathbf{L}^\infty(\mathbf{X}, \mu)$.

   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1 \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty \right)$.

2. *However, if $\inf_{\substack{\mathbf{U}\in\mathcal{X} \\ \mu[\mathbf{U}]>0}} \mu[\mathbf{U}] \ = \ 0$, then $\mathbf{L}^1(\mathbf{X}, \mu) \ \not\subset \ \mathbf{L}^\infty(\mathbf{X}, \mu)$, and 1(c) fails.*

Proof: 1(b-c) follows from 1(a); the proofs of 1(a) and 2 are **Exercise 143**. □

Example 144: $\mathbf{L}^1[0,1]$ vs. $\mathbf{L}^\infty[0,1]$

Let $[0,1]$ have the Lebesgue measure. If $f:[0,1]\longrightarrow\mathbb{C}$ is measurable, then $\|f\|_1 = \int_0^1 |f(x)|\,dx$, and $\|f\|_\infty = \sup_{0\le x\le 1} |f(x)|$. Clealry, $\int_0^1 |f(x)| \le \sup_{0\le x\le 1} |f(x)|$. It follows $\mathbf{L}^\infty[0,1] \subset \mathbf{L}^1[0,1]$, and $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1 \right)$.

This is a special case of the following result:

Proposition 145: *Let $(\mathbf{X},\mathcal{X},\mu)$ be a measure space, and let $f:\mathbf{X}\longrightarrow\mathbb{R}$ and $f_n:\mathbf{X}\longrightarrow\mathbb{R}$ be integrable for all $n\in\mathbb{N}$.*

1. *If $(\mathbf{X},\mathcal{X},\mu)$ is* **finite***, and $M=\mu(\mathbf{X})$, then:*
   (a) $\|f\|_1 \le M\cdot\|f\|_\infty$.
   (b) $\mathbf{L}^\infty(\mathbf{X},\mu) \subset \mathbf{L}^1(\mathbf{X},\mu)$.
   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \textit{ in } \mathbf{L}^\infty \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \textit{ in } \mathbf{L}^1 \right)$.
2. *If $(\mathbf{X},\mathcal{X},\mu)$ is* **infinite***, then $\mathbf{L}^\infty(\mathbf{X},\mu) \not\subset \mathbf{L}^1(\mathbf{X},\mu)$, and 3(c) fails.*
3. *However, if there is some $F\in\mathbf{L}^1(\mathbf{X},\mathcal{X},\mu)$ such that $\left|f_n(x)-f(x)\right| \le F(x)$ a.e., for all $n\in\mathbb{N}$, then $\left(f_n \xrightarrow[n\to\infty]{} f \textit{ in } \mathbf{L}^\infty \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \textit{ in } \mathbf{L}^1 \right)$.*

Proof: 1(b-c) follows from 1(a). The proofs of 1(a) and 2 are **Exercise 144**. Part 3 follows from Lebesgue's Dominated Convergence Theorem (page 84). □

Let $(\mathbf{Y},d)$ be a metric space, and suppose that $f:\mathbf{X}\longrightarrow\mathbf{Y}$ and $f_n:\mathbf{X}\longrightarrow\mathbf{Y}$ for all $n\in\mathbb{N}$. If we define $g_n(x) = d\left(f(x), f_n(x)\right)$ for all $n\in\mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **in $\mathbf{L}^1$** if $g_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^1$. Hence, to understand $\mathbf{L}^1$ convergence in general, it is sufficient to understand $\mathbf{L}^1$ convergence of nonnegative functions to the constant 0 function.

4.5(g) $\mathbf{L}^2$ convergence

Prerequisites: §4.3,§3.2 **Recommended:** §4.5(f),§4.5(d)

If $f,g\in\mathbf{L}^2(\mathbf{X},\mu)$, then the **$\mathbf{L}^2$-distance** between f and g is just

$$\|f-g\|_2 = \left(\int_{\mathbf{X}} \left|f(x)-g(x)\right|^2 \, d\mu[x]\right)^{1/2}$$

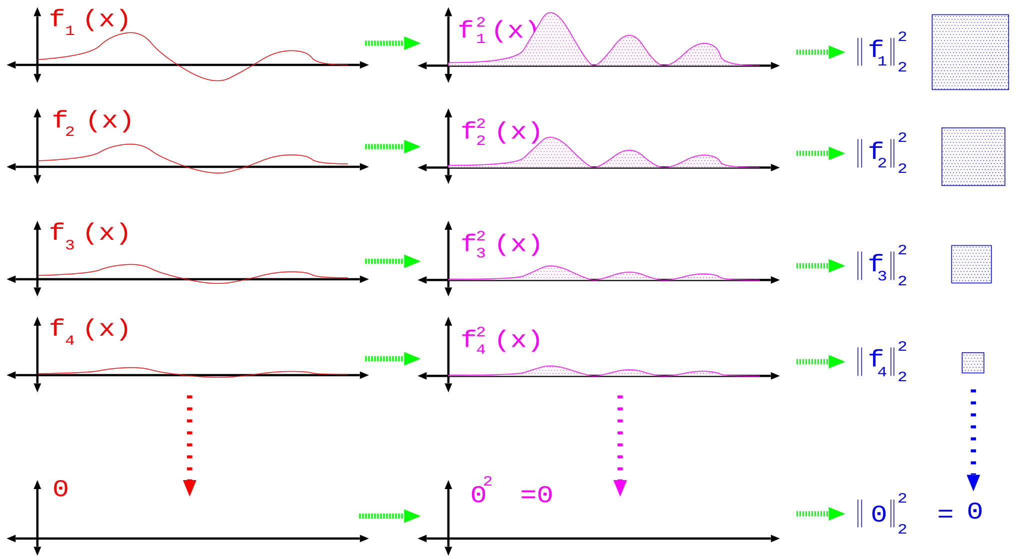


Figure 4.23: The sequence $\{f_1, f_2, f_3, \ldots\}$ converges to the constant 0 function in $\mathbf{L}^2(\mathbf{X})$.

If we think of f as an "approximation" of g, then $\|f-g\|_2$ measures the *root-mean-squared error* of this approximation. If $\{f_1, f_2, f_3, \ldots\}$ is a sequence of successive approximations of f, then we say the sequence **converges to** f **in** $\mathbf{L}^2$ if $\lim_{n\to\infty} \|f_n - f\|_2 = 0$. See Figure 4.23.

Example 146: Let $\mathbf{X} = \mathbb{R}$, with the Lebesgue measure.

(a) For each $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} 1 & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$ as in Example (141a) on page 116. Then $\|f_n\|_2 = \frac{1}{\sqrt{n}}$, so $f_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^2$.

(b) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$, as in Example (132c) on page 108. This sequence converges *pointwise* to zero, but $\|f_n\|_2 = \sqrt{n}$ for all n, so the sequence does *not* converge to zero in $\mathbf{L}^2$.

(c) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} \frac{1}{n} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$, as in Example (141e) on page 116. This sequence does *not* converge to zero in $\mathbf{L}^1$. However, $\|f_n\|_2 = \frac{1}{\sqrt{n}}$ for all n, so the sequence *does* converge to zero in $\mathbf{L}^2$.

(d) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} \frac{1}{\sqrt{n}} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$. This sequence converges *uniformly* to zero. However, $\|f_n\|_2 = 1$ for all n, so this sequence does *not* converge to zero in $\mathbf{L}^2$.

Example 147: $\ell^1(\mathbb{N})$ vs. $\ell^2(\mathbb{N})$ vs. $\ell^\infty(\mathbb{N})$

Let $\mathbb{R}^{\mathbb{N}}$ be the set of all sequences of real numbers; we'll write such a sequence as $\mathbf{a} = [a_n]_{n=1}^{\infty}$. Recall that $\|\mathbf{a}\|_2 = \sqrt{\sum_{n=1}^{\infty} |a_n|^2}$ and $\ell^2(\mathbb{N}) = \{\mathbf{a} \in \mathbb{R}^N atur \ ; \ \|\mathbf{a}\|_2 < \infty\}$. We next will show that $\sup_{n\in\mathbb{N}} |a_n| \leq \sqrt{\sum_{n=1}^{\infty} |a_n|^2} \leq \sum_{n=1}^{\infty} |a_n|$. It follows that $\ell^1(\mathbb{N}) \subset \ell^2(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$.

Proposition 148: *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f : \mathbf{X} \longrightarrow \mathbb{R}$ and $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be integrable for all $n \in \mathbb{N}$.*

1. *If $(\mathbf{X}, \mathcal{X}, \mu)$ is **discrete**, and $m = \inf_{\substack{\mathbf{U}\in\mathcal{X} \\ \mu[\mathbf{U}]>0}} \mu[\mathbf{U}] > 0$, then:*
   (a) $\|f\|_\infty \leq \frac{1}{\sqrt{m}}\|f\|_2$ *and* $\|f\|_2 \leq \frac{1}{\sqrt{m}}\|f\|_1$.
   (b) $\mathbf{L}^1(\mathbf{X}, \mu) \subset \mathbf{L}^2(\mathbf{X}, \mu) \subset \mathbf{L}^\infty(\mathbf{X}, \mu)$.
   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1 \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^2 \right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty \right)$.
2. *However, if $\inf_{\substack{\mathbf{U}\in\mathcal{X} \\ \mu[\mathbf{U}]>0}} \mu[\mathbf{U}] = 0$, then $\mathbf{L}^1(\mathbf{X}, \mu) \not\subset \mathbf{L}^2(\mathbf{X}, \mu) \not\subset \mathbf{L}^\infty(\mathbf{X}, \mu)$, and 1(c) fails.*

Proof: 1(b-c) and follow from 1(a);

Proof of 1(a): The first inequality is **Exercise 145** . We will prove the second. Assume without loss of generality that $\{x\}$ is measurable and that $\mu\{x\} > 0$ for all $x \in \mathbf{X}$. Thus, $\|f\|_2 = \left(\sum_{x\in\mathbf{X}} |f(x)|^2 \cdot \mu\{x\}\right)^{1/2}$ and $\|f\|_1 = \sum_{x\in\mathbf{X}} |f(x)| \cdot \mu\{x\}$. Thus

$$\begin{aligned}
\|f\|_1{}^2 &= \left(\sum_{x\in\mathbf{X}} |f(x)| \cdot \mu\{x\}\right)^2 \quad \geq_{(a)} \quad \sum_{x\in\mathbf{X}} \Big(|f(x)| \cdot \mu\{x\}\Big)^2 \\
&= \sum_{x\in\mathbf{X}} |f(x)|^2 \cdot \mu\{x\}^2 \quad = \quad \sum_{x\in\mathbf{X}} |f(x)|^2 \cdot \mu\{x\} \cdot \mu\{x\} \\
&\geq \left(\sum_{x\in\mathbf{X}} |f(x)|^2 \cdot \mu\{x\}\right) \cdot \inf_{x\in\mathbf{X}} \mu\{x\} \quad =_{(b)} \quad \|f\|_2{}^2 \cdot m
\end{aligned}$$

(a) Because the function $x \longrightarrow x^2$ is convex –ie. $(x + y)^2 \geq x^2 + y^2$ if $x, y \geq 0$.

(b) By definition, $m = \inf_{x\in\mathbf{X}} \mu\{x\}$.

Thus, taking the square root on both sides, we get $\|f\|_1 \geq \|f\|_2 \cdot m^{\frac{1}{2}}$. Divide both sides by $m^{\frac{1}{2}}$ to get $m^{-\frac{1}{2}} \cdot \|f\|_1 \geq \|f\|_2$, as desired.

Proof of 2: **Exercise 146** Hint: Find examples of functions violating each inclusion. □

Example 149: $\mathbf{L}^1[0,1]$ vs. $\mathbf{L}^2[0,1]$ vs. $\mathbf{L}^\infty[0,1]$

Let $[0,1]$ have the Lebesgue measure. If $f:[0,1]\longrightarrow\mathbb{C}$ is measurable, then $\|f\|_2 = \sqrt{\int_0^1 |f(x)|^2\,dx}$.

We next will show that $\int_0^1 |f(x)| \leq \sqrt{\int_0^1 |f(x)|^2\,dx} \leq \sup_{0\leq x\leq 1} |f(x)|$. It follows $\mathbf{L}^\infty[0,1] \subset \mathbf{L}^2[0,1] \subset \mathbf{L}^1[0,1]$.

Proposition 150: *Let $(\mathbf{X},\mathcal{X},\mu)$ be a measure space, and let $f:\mathbf{X}\longrightarrow\mathbb{R}$ and $f_n:\mathbf{X}\longrightarrow\mathbb{R}$ be integrable for all $n\in\mathbb{N}$.*

1. *If $(\mathbf{X},\mathcal{X},\mu)$ is* **finite**, *and $M=\mu(\mathbf{X})$, then:*
   (a) $\|f\|_1 \leq \sqrt{M}\cdot\|f\|_2$, *and* $\|f\|_2 \leq \sqrt{M}\cdot\|f\|_\infty$.
   (b) $\mathbf{L}^\infty(\mathbf{X},\mu) \subset \mathbf{L}^2(\mathbf{X},\mu) \subset \mathbf{L}^1(\mathbf{X},\mu)$.
   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty\right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^2\right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1\right)$.
2. *If $(\mathbf{X},\mathcal{X},\mu)$ is* **infinite**, *then $\mathbf{L}^\infty(\mathbf{X},\mu) \not\subset \mathbf{L}^2(\mathbf{X},\mu) \not\subset \mathbf{L}^1(\mathbf{X},\mu)$, and 3(c) fails.*
3. *However, if there is some $F\in\mathbf{L}^2(\mathbf{X},\mathcal{X},\mu)$ such that $\left|f_n(x)-f(x)\right| \leq F(x)$ a.e., for all $n\in\mathbb{N}$, then* $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty\right) \Longrightarrow \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^2\right)$.

Proof: 1(b-c) follow from 1(a)

Proof of 1(a): Is there a way to do this without Hölder ?

Part 5 follows from Lebesgue's Dominated Convergence Theorem (page 84).

Proof of 2: **Exercise 147** Hint: Find examples of functions violating each inclusion. □

Let $(\mathbf{Y},d)$ be a metric space, and suppose that $f:\mathbf{X}\longrightarrow\mathbf{Y}$ and $f_n:\mathbf{X}\longrightarrow\mathbf{Y}$ for all $n\in\mathbb{N}$. If we define $g_n(x) = d\left(f(x), f_n(x)\right)$ for all $n\in\mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **in $\mathbf{L}^2$** if $g_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^2$. Hence, to understand $\mathbf{L}^2$ convergence in general, it is sufficient to understand $\mathbf{L}^2$ convergence of nonnegative functions to the constant 0 function.

4.5(h) $\mathbf{L}^p$ convergence

Prerequisites: §3.2 **Recommended:** §4.5(f),§4.5(g),§4.5(d)

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $p \in [1, \infty)$. If $f : \mathbf{X} \longrightarrow \mathbb{R}$ is measurable, then we define the **$\mathbf{L}^p$-norm** of f:

$$\|f\|_p \quad = \quad \left(\int_{\mathbf{X}} \left| f(x) \right|^p \, d\mu[x] \right)^{1/p}$$

Observe that, if $p = 2$, this is just the $\mathbf{L}^2$ norm from §**??**, while if $p = 1$, it is the $\mathbf{L}^1$ norm from §**??**.

If $f, g \in \mathbf{L}^p(\mathbf{X}, \mu)$, then the **$\mathbf{L}^p$-distance** between f and g is just

$$\|f - g\|_p \;=\; \left(\int_{\mathbf{X}} |f(x) - g(x)|^p \, d\mu[x] \right)^{1/p}$$

Again, if $p = 2$ or 1, this agrees with the $\mathbf{L}^2$ or $\mathbf{L}^1$ distances.

If $\{f_1, f_2, f_3, \ldots\}$ is a sequence of elements in $\mathbf{L}^p$, of f, then we say the sequence **converges to f in $\mathbf{L}^p$** if $\lim_{n\to\infty} \|f_n - f\|_p \;=\; 0$. See Figure 4.23.

Example 151: Let $\mathbf{X} = \mathbb{R}$, with the Lebesgue measure.

(a) For each $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} 1 & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$ as in Example (141a) on page 116. Then $\|f_n\|_p = \frac{1}{n^{1/p}}$, so $f_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^p$.

(b) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n; \\ 0 & \text{otherwise} \end{cases}$, as in Example (132c) on page 108. This sequence converges *pointwise* to zero, but $\|f_n\|_p = n^{\frac{p-1}{p}}$ for all n, and $\frac{p-1}{p} > 0$, so the sequence does *not* converge to zero in $\mathbf{L}^p$.

(c) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} \frac{1}{n} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$, as in Example (141e) on page 116. This sequence does *not* converge to zero in $\mathbf{L}^1$. However, $\|f_n\|_p = n^{\frac{1-p}{p}}$ for all n, and $\frac{1-p}{p} < 0$, so the sequence *does* converge to zero in $\mathbf{L}^p$ for any $p > 1$.

(d) For all $n \in \mathbb{N}$, let $f_n(x) = \begin{cases} \frac{1}{n^{1/p}} & \text{if } 0 < x < n; \\ 0 & \text{otherwise} \end{cases}$. This sequence converges *uniformly* to zero. However, $\|f_n\|_p = 1$ for all n, so this sequence does *not* converge to zero in $\mathbf{L}^p$.

Proposition 152: *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $f : \mathbf{X} \longrightarrow \mathbb{R}$ and $f_n : \mathbf{X} \longrightarrow \mathbb{R}$ be integrable for all $n \in \mathbb{N}$. Let $1 < p < q < \infty$.*

1. *If $(\mathbf{X}, \mathcal{X}, \mu)$ is* **discrete**, *and $m = \inf_{\substack{\mathbf{U} \in \mathcal{X} \\ \mu[\mathbf{U}] > 0}} \mu[\mathbf{U}] > 0$, then:*

   (a) $\|f\|_\infty \leq m^{-\frac{1}{q}} \cdot \|f\|_q$, $\|f\|_q \leq m^{\frac{1}{q}-\frac{1}{p}} \cdot \|f\|_p$, *and* $\|f\|_p \leq m^{\frac{1}{p}-1} \cdot \|f\|_1$.

   (b) $\mathbf{L}^1(\mathbf{X}, \mu) \subset \mathbf{L}^p(\mathbf{X}, \mu) \subset \mathbf{L}^q(\mathbf{X}, \mu) \subset \mathbf{L}^\infty(\mathbf{X}, \mu)$.

   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1 \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^p \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^q \right)$
   $\implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty \right)$.

2. *However, if $\inf_{\substack{\mathbf{U} \in \mathcal{X} \\ \mu[\mathbf{U}] > 0}} \mu[\mathbf{U}] = 0$, then $\mathbf{L}^1(\mathbf{X}, \mu) \not\subset \mathbf{L}^p(\mathbf{X}, \mu) \not\subset \mathbf{L}^q(\mathbf{X}, \mu) \not\subset \mathbf{L}^\infty(\mathbf{X}, \mu)$, and 1(c) fails.*

3. *If $(\mathbf{X}, \mathcal{X}, \mu)$ is* **finite**, *and $M = \mu(\mathbf{X})$, then:*

   (a) $\|f\|_1 \leq M^{\left(1-\frac{1}{p}\right)} \cdot \|f\|_p$, $\|f\|_p \leq M^{\left(\frac{1}{p}-\frac{1}{q}\right)} \cdot \|f\|_q$, and $\|f\|_q \leq M^{\frac{1}{q}} \cdot \|f\|_\infty$.

   (b) $\mathbf{L}^\infty(\mathbf{X}, \mu) \subset \mathbf{L}^q(\mathbf{X}, \mu) \subset \mathbf{L}^p(\mathbf{X}, \mu) \subset \mathbf{L}^1(\mathbf{X}, \mu)$.

   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^\infty \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^q \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^p \right)$
   $\implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^1 \right)$.

4. *However, if $(\mathbf{X}, \mathcal{X}, \mu)$ is* **infinite**, *then $\mathbf{L}^\infty(\mathbf{X}, \mu) \not\subset \mathbf{L}^q(\mathbf{X}, \mu) \not\subset \mathbf{L}^p(\mathbf{X}, \mu) \not\subset \mathbf{L}^1(\mathbf{X}, \mu)$, and 3(c) fails.*

5. *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be any measure space, and suppose $1 \leq p < q < r \leq \infty$. Let $\lambda = \frac{pr-pq}{qr-pq}$. (thus $\lambda \cdot \frac{1}{p} + (1-\lambda) \cdot \frac{1}{r} = \frac{1}{q}$).*

   (a) $\|f\|_q \leq \|f\|_p{}^\lambda \cdot \|f\|_r{}^{1-\lambda}$.

   (b) $\mathbf{L}^p(\mathbf{X}, \mu) \cap \mathbf{L}^r(\mathbf{X}, \mu) \subset \mathbf{L}^q(\mathbf{X}, \mu) \subset \mathbf{L}^p(\mathbf{X}, \mu) + \mathbf{L}^r(\mathbf{X}, \mu)$.

   (c) $\left(f_n \xrightarrow[n\to\infty]{} f \text{ in both } \mathbf{L}^p \text{ and } \mathbf{L}^r \right) \implies \left(f_n \xrightarrow[n\to\infty]{} f \text{ in } \mathbf{L}^q \right)$.

Proof: 1(b,c) and 3(b,c) follow respectively from 1(a) and 3(a).

Proof of 1(a): Assume without loss of generality that $\mathcal{X} = \mathcal{P}(\mathbf{X})$, and that $\mu\{x\} > 0$ for all $x \in \mathbf{X}$. Thus, $\|f\|_p = \left(\sum_{x \in \mathbf{X}} |f(x)|^p \cdot \mu\{x\} \right)^{1/p}$ and similarly for $\|f\|_q$. Thus,

$$\|f\|_p{}^q = \left(\sum_{x \in \mathbf{X}} |f(x)|^p \cdot \mu\{x\} \right)^{\frac{q}{p}} \geq_{(a)} \sum_{x \in \mathbf{X}} \left(|f(x)|^p \cdot \mu\{x\} \right)^{\frac{q}{p}}$$

$$= \sum_{x\in\mathbf{X}} |f(x)|^q \cdot \mu\{x\}^{\frac{q}{p}} \quad = \quad \sum_{x\in\mathbf{X}} |f(x)|^q \cdot \mu\{x\} \cdot \mu\{x\}^{\frac{q}{p}-1}$$

$$\geq \left(\sum_{x\in\mathbf{X}} |f(x)|^q \cdot \mu\{x\}\right) \cdot \inf_{x\in\mathbf{X}} \mu\{x\}^{\frac{q}{p}-1} \quad =_{(b)} \quad \|f\|_q{}^q \cdot m^{\frac{q}{p}-1}$$

(a) Because $q > p$, so that $\frac{q}{p} > 1$, so the function $x \longrightarrow x^{\frac{q}{p}}$ is convex –ie. $(x+y)^{\frac{q}{p}} \geq x^{\frac{q}{p}} + y^{\frac{q}{p}}$.

(b) By definition, $m = \inf_{x\in\mathbf{X}} \mu\{x\}$.

Thus, taking the qth root on both sides, we get $\|f\|_p \geq \|f\|_q \cdot m^{\frac{1}{p}-\frac{1}{q}}$. Divide both sides by $m^{\frac{1}{p}-\frac{1}{q}}$ to get $m^{\frac{1}{q}-\frac{1}{p}} \cdot \|f\|_p \geq \|f\|_q$, as desired.

The case involving $\|f\|_p$ vs. $\|f\|_1$ is obtained by applying the '$\|f\|_{q'}$ vs. $\|f\|_{p'}$' case with $q' = p$ and $p' = 1$. The case involving $\|f\|_\infty$ vs. $\|f\|_q$ is **Exercise 148**.

Proof of 3(a): $\boxed{\text{Is there a way to do this without Hölder ?}}$

Proof of 2,4: **Exercise 149** Hint: Find examples of functions violating each inclusion.

Proof of 5(a): $\boxed{\text{Is there a way to do this without Hölder ?}}$

5(c) follows from **5(a)**.

Proof of 5(b): If $f \in \mathbf{L}^p(\mathbf{X},\mu) \cap \mathbf{L}^r(\mathbf{X},\mu)$, it follows immediately from **5(a)** that $f \in \mathbf{L}^q(\mathbf{X},\mu)$. To see the other inclusion, suppose $f \in \mathbf{L}^q$, and define

$$f_p(x) = \begin{cases} f(x) & \text{if } |f(x)| \geq 1 \\ 0 & \text{if } |f(x)| < 1 \end{cases} \quad \text{and} \quad f_r(x) = \begin{cases} 0 & \text{if } |f(x)| \geq 1 \\ f(x) & \text{if } |f(x)| < 1 \end{cases}$$

Observe that $f(x) = f_p(x) + f_r(x)$. It is **Exercise 150** to verify that $f_p \in \mathbf{L}^p(\mathbf{X},\mu)$ and $f_r \in \mathbf{L}^r(\mathbf{X},\mu)$. □

Let $(\mathbf{Y}, d)$ be a metric space, and suppose that $f : \mathbf{X} \longrightarrow \mathbf{Y}$ and $f_n : \mathbf{X} \longrightarrow \mathbf{Y}$ for all $n \in \mathbb{N}$. If we define $g_n(x) = d\big(f(x), f_n(x)\big)$ for all $n \in \mathbb{N}$, then we say $f_n \xrightarrow[n\to\infty]{} f$ **in** $\mathbf{L}^p$ if $g_n \xrightarrow[n\to\infty]{} 0$ in $\mathbf{L}^p$. Hence, to understand $\mathbf{L}^p$ convergence in general, it is sufficient to understand $\mathbf{L}^p$ convergence of nonnegative functions to the constant 0 function.

5 Information

5.1 Sigma Algebras as Information

Prerequisites: §1.1

Consider the following commonplace statements:

- "In retrospect it was a bad idea to buy tech stocks."
- "Alice knows more about Chinese history than Bob does."
- "Xander and Yvonne each know things about Zoroastrianism which the other does not."
- "You learn something new every day."
- "In 1944, the Germans did not know that the Allies had broken the Enigma cypher."

Each of these is a statement about *knowledge*, or the lack thereof. In particular, each compares the *states of knowledge* of two people (or the same person at different times). This idea of a *state of knowledge* is ubiquitous in everyday parlance. How can we mathematically model this?

The appropriate mathematical model of a knowledge-state is a sigma-algebra. Imagine we are curious about some system $\mathcal{S}$, whose (unknown) internal state is an element of a statespace $\mathbf{X}$. For example, suppose $\mathcal{S}$ is a lost treasure; in this case, the location of the treasure is some point on the Earth's (spherical) surface, so $\mathbf{X} = \mathbb{S}^2$.

Let $\mathcal{X}$ be a sigma-algebra over $\mathbf{X}$, and suppose $x \in \mathbf{X}$ is the (unknown) state of $\mathcal{S}$. The *knowledge represented by* $\mathcal{X}$ is the information about $\mathcal{S}$ that one obtains from knowing, for every $\mathbf{U} \in \mathcal{X}$, whether or not $x \in \mathbf{U}$. Thus, if $\mathcal{Y} \supset \mathcal{X}$ is a larger sigma algebra, then $\mathcal{Y}$ contains 'more' information than $\mathcal{X}$. The power set $\mathcal{P}(\mathbf{X})$ represents a state of total omniscience; the null algebra $\{\emptyset, \mathbf{X}\}$ represents a state of total ignorance.

Example 153: Dyadic Partitions of the Interval

Let $\mathbb{I} = [0, 1)$. Any element $\alpha \in \mathbb{I}$ has a unique[1] **binary expansion** $\alpha = 0.a_1a_2a_3a_4\ldots$ such that

$$\alpha = \sum_{n=0}^{\infty} \frac{a_n}{2^n}$$

[1]Actually the dyadic rationals (numbers of the form $\frac{a}{2^n}$) are an exception, and have *two* binary expansions. However, these form a set of measure zero, so we can ignore them.

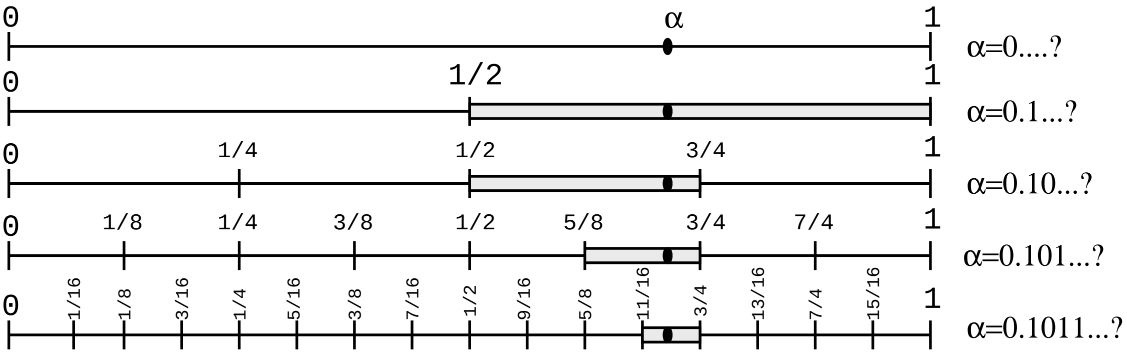


Figure 5.1: Finer dyadic partitions reveal more binary digits of α

Consider the following sequence of **dyadic partitions**, illustrated in Figure 5.1:

$$
\begin{aligned}
\mathcal{P}_0 &= \{\mathbb{I}\} \\
\mathcal{P}_1 &= \left\{\left[0,\frac{1}{2}\right),\ \left[\frac{1}{2},1\right)\right\} \\
\mathcal{P}_2 &= \left\{\left[0,\frac{1}{4}\right),\ \left[\frac{1}{4},\frac{1}{2}\right),\ \left[\frac{1}{2},\frac{3}{4}\right),\ \left[\frac{3}{4},1\right)\right\} \\
\mathcal{P}_3 &= \left\{\left[0,\frac{1}{8}\right),\ \left[\frac{1}{8},\frac{1}{4}\right),\ \left[\frac{1}{4},\frac{3}{8}\right),\ \left[\frac{3}{8},\frac{1}{2}\right),\ \left[\frac{1}{2},\frac{5}{8}\right),\ \left[\frac{5}{8},\frac{3}{4}\right),\ \left[\frac{3}{4},\frac{7}{8}\right),\ \left[\frac{7}{8},1\right)\right\} \\
\vdots\ &\ \vdots\ \ \vdots
\end{aligned}
$$

Suppose α is unknown. Then note that:

$$\Big(\text{ knowledge of digits } a_1, a_2, \ldots, a_n \Big) \iff \Big(\text{ knowledge of which element of } \mathcal{P}_n \text{ contains } \alpha \Big)$$

In other words, $\mathcal{P}_n$ 'contains' the information about the first n binary digits of α.

Notice that $\mathcal{P}_0 \prec \mathcal{P}_1 \prec \mathcal{P}_2 \ldots$. Thus, *finer* partitions contain *more* information.

Example: *Partitions of a Square*

Let $\mathbb{I}^2$ be a square, and let $\mathcal{P}$ be a partition of $\mathbb{I}^2$, with associated sigma-algebra $\sigma(\mathcal{P})$. The information contained in $\sigma(\mathcal{P})$ is the information about $x \in \mathbb{I}^2$ you obtain from knowing which atom of $\mathcal{P}$ the point x lies in.

Recall that partition $\mathcal{Q}$ **refines** $\mathcal{P}$ if every element of $\mathcal{P}$ can be written as a union of atoms in $\mathcal{P}$, as shown in Figure 1.5 on page 4. Thus, $\sigma(\mathcal{Q})$ contains *more* information than $\sigma(\mathcal{P})$, because it specifies the location of x with greater 'precision'. For example, suppose $\mathcal{P}$ and $\mathcal{Q}$ were *grids* on $\mathbb{I}^2$, as in Figure 5.2. If $\mathcal{P} \prec \mathcal{Q}$, then $\mathcal{Q}$ is a *higher resolution* grid, providing proportionately better information about spatial position.

Example 154: Projections of the Cube

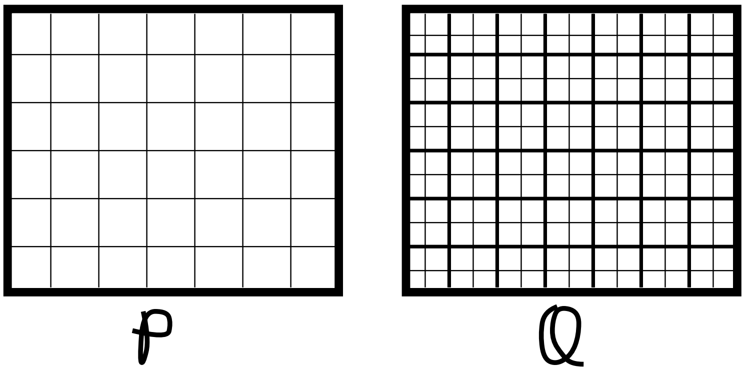


Figure 5.2: A higher resolution grid corresponds to a finer partition.

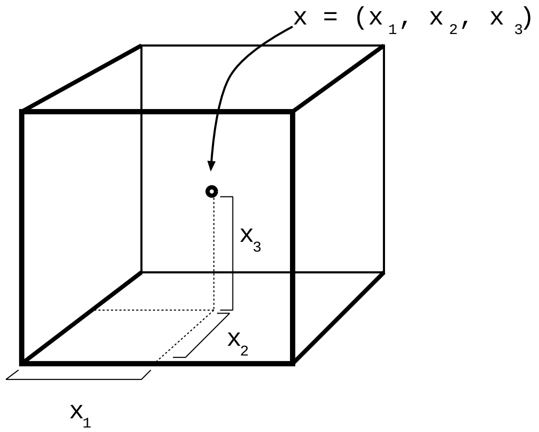


Figure 5.3: Our position in the cube is completely specified by three coordinates.

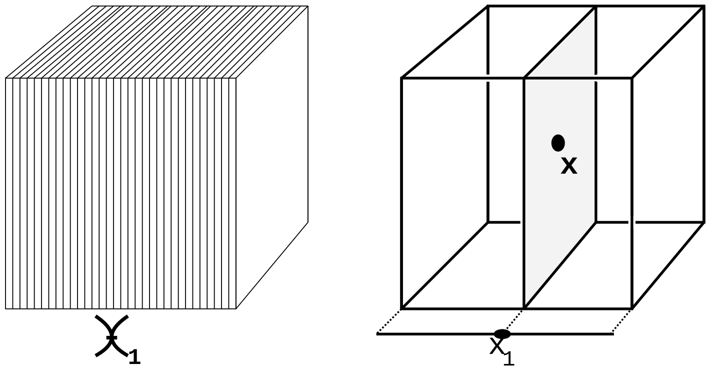


Figure 5.4: The sigma algebra $\mathcal{X}_1$ consists of "vertical sheets", and specifies the x_1 coordinate.

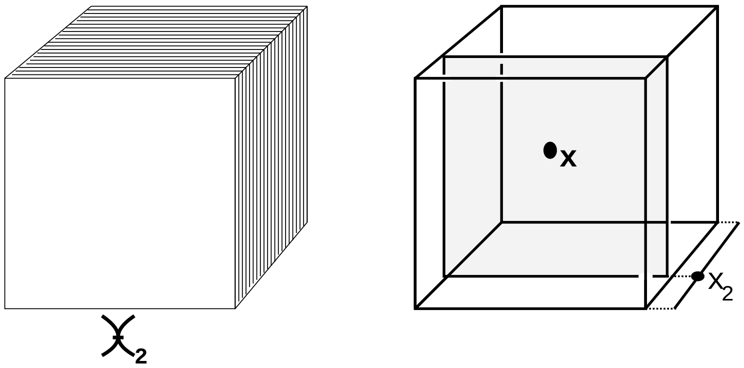


Figure 5.5: The sigma algebra $\mathcal{X}_2$ consists of "vertical sheets", and specifies the x_2 coordinate.

Consider the **unit cube**, $\mathbb{I}^3 := \mathbb{I} \times \mathbb{I} \times \mathbb{I}$, where $\mathbb{I} := [0,1]$ is the unit interval (see Figure 5.3). Let $\mathbb{I}^3$ have the Borel sigma-algebra $\mathcal{I}^3$. As shown in Figure 5.3, the position of $x \in \mathbb{I}^3$ is completely specified by three coordinates (x_1, x_2, x_3). The information embodied by each coordinate corresponds to a certain sigma-algebra.

Consider the projection onto the first coordinate, $\mathbf{pr}_1 : \mathbb{I}^3 \longrightarrow \mathbb{I}$. Thus, if $\mathbf{x} := (x_1, x_2, x_3) \in \mathbb{I}^3$, then $\mathbf{pr}_1(\mathbf{x}) = x_1 \in \mathbb{I}$. Consider the *pulled back* sigma algebra $\mathcal{X}_1 := \mathbf{pr}_1{}^{-1}(\mathcal{I})$, (where $\mathcal{I}$ is the Borel sigma-algebra on $\mathbb{I}$). Roughly speaking, $\mathcal{X}_1$ consists of all "vertical sheets" in the cube (Figure 5.4). That is:

$$\mathcal{X}_1 \quad = \quad \left\{ \mathbf{pr}_1{}^{-1}(\mathbf{U}) \; ; \; \mathbf{U} \in \mathcal{I}^2 \right\} \quad = \quad \left\{ \mathbf{U} \times \mathbb{I}^2 \; ; \; \mathbf{U} \in \mathcal{I} \right\} \quad = \quad \mathcal{I} \otimes \{\mathbb{I}^2\}$$

Knowing the coordinate x_1 is equivalent to knowing which vertical sheet $\mathbf{x}$ lies in.

Next, consider the projection onto the second coordinate, $\mathbf{pr}_2 : \mathbb{I}^3 \longrightarrow \mathbb{I}$. Thus, if $\mathbf{x} := (x_1, x_2, x_3) \in \mathbb{I}^3$, then $\mathbf{pr}_2(\mathbf{x}) = x_2 \in \mathbb{I}$. The pulled back sigma algebra $\mathcal{X}_2 := \mathbf{pr}_2{}^{-1}(\mathcal{I})$ consists of the 'vertical sheets' shown in Figure 5.5:

$$\mathcal{X}_2 \quad = \quad \left\{ \mathbf{pr}_2{}^{-1}(\mathbf{U}) \; ; \; \mathbf{U} \in \mathcal{I} \right\} \quad = \quad \{ \mathbb{I} \times \mathbf{U} \times \mathbb{I} \; ; \; \mathbf{U} \in \mathcal{I} \} \quad = \quad \{\mathbb{I}\} \otimes \mathcal{I} \otimes \{\mathbb{I}\}$$

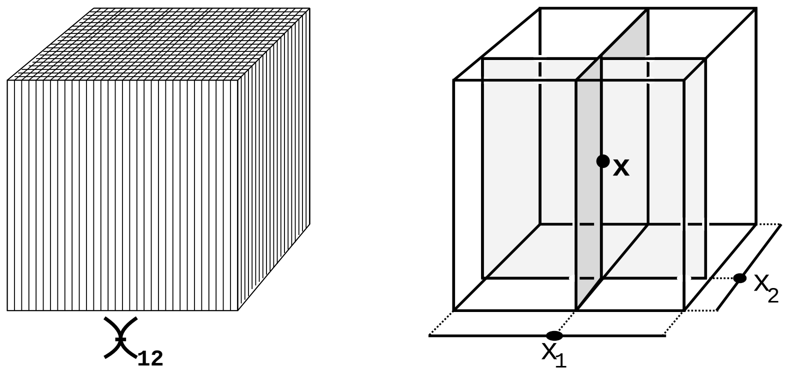


Figure 5.6: The sigma algebra $\mathcal{X}_{12}$ completely specifies (x_1, x_2) coordinates of $\mathbf{x}$.

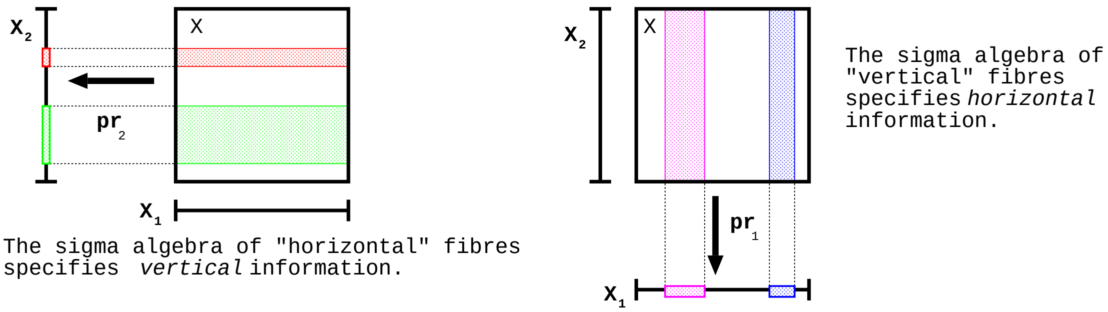


Figure 5.7: "Vertical" and "horizontal" sigma algebras in a product space.

Knowing the coordinate x_2 is equivalent to knowing which of these sheets $\mathbf{x}$ lies in.

Finally, Let $\mathbb{I}^2$ be the unit square, and consider the projection onto the first two coordinates, $\mathbf{pr}_{1,2} : \mathbb{I}^3 \longrightarrow \mathbb{I}^2$; if $\mathbf{x} := (x_1, x_2, x_3) \in \mathbb{I}^3$, then $\mathbf{pr}_{1,2}(\mathbf{x}) = (x_1, x_2) \in \mathbb{I}^2$. Let $\mathcal{X}_{12} := \mathbf{pr}_{1,2}{}^{-1}(\mathcal{I}^2)$ (where $\mathcal{I}^2$ is the Borel sigma-algebra on $\mathbb{I}^2$). As discussed in Example 28 on page 26, elements of $\mathcal{X}_{12}$ look like "vertical fibres" in the cube (Figure 5.6). That is: $\mathcal{X}_{12} \quad = \quad \mathcal{I}^2 \otimes \{\mathbb{I}\}$. Alternately, we could write:

$$\mathcal{X}_{12} \quad = \quad \mathcal{X}_1 \otimes \mathcal{X}_2$$

Thus, $\mathcal{X}_{12}$-related information specifies exactly which vertical fibre-sets $\mathbf{x}$ is in, and exactly which vertical fibre sets we are *not* in. From this, we can reconstruct arbitrarily accurate information about the coordinates x_1 and x_2. In other words, as illustrated in Figure 5.6:

> *The information contained in $\mathcal{X}_{12}$ is exactly the same information contained in the (x_1, x_2) coordinates of* $\mathbf{x}$.

Example: *Product Spaces*

Let $(\mathbf{X}_1, \mathcal{X}_1)$ and $(\mathbf{X}_2, \mathcal{X}_2)$ be measurable spaces, and consider the product space, $(\mathbf{X}, \mathcal{X}) = (\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{X}_1 \otimes \mathcal{X}_2)$. For $k = 1$ or 2, let $\mathbf{pr}_k : \mathbf{X} \longrightarrow \mathbf{X}_k$ be projection onto the kth coordinate, and consider the pulled-back sigma-algebra

$$\widehat{\mathcal{X}}_k := \mathbf{pr}_k^{-1}(\mathcal{X}_k)$$

Then the sigma algebra $\widehat{\mathcal{X}}_k$ contains exactly the same information as the coordinate $\mathbf{pr}_k$ contains. As illustrated in Figure 5.7, one can imagine $\widehat{\mathcal{X}}_1$ as the sigma-algebra of all 'vertical' fibres, so it contains all 'horizontal' information. Conversely, $\widehat{\mathcal{X}}_2$ consists of all 'horizontal' fibres, and thus, specifies 'vertical' information.

Example: *A measurement*

Imagine that the space $\mathbf{X}$ represents the state of some system $\mathcal{S}$, and imagine we perform a 'measurement' on $\mathcal{S}$ with an apparatus, which yields output in some set $\mathbf{Y}$. For example, if the apparatus yields *numerical* values, then $\mathbf{Y} = \mathbb{R}$.

Thus, we can represent the measurement procedure with a function

$$f : \mathbf{X} \longrightarrow \mathbf{Y}$$

Suppose that $\mathbf{Y}$ is a measurable space, with sigma-algebra $\mathcal{Y}$, Then the pulled-back sigma algebra $f^{-1}\mathcal{Y}$ is a sigma algebra on $\mathbf{X}$, and contains the information about $\mathcal{S}$ that is provided by the measurement.

5.1(a) Refinement and Filtration

Prerequisites: §5.1

If the sigma algebra $\mathcal{X}$ contains strictly *more* information than the sigma-algebra $\mathcal{Y}$ (in the sense that we could recover every bit of information about $\mathcal{Y}$ by looking at $\mathcal{X}$), then we say that $\mathcal{X}$ **refines** $\mathcal{Y}$. In a sense, once we know the information contained in $\mathcal{X}$, the information contained in $\mathcal{Y}$ is *redundant*. Formally, we have the following definition.

Definition 155 Refinement

*Let $\mathbf{X}$ be a set, and let $\mathcal{X}$ and $\mathcal{Y}$ be two sigma-algebras on $\mathbf{X}$. We say that $\mathcal{X}$ **refines** $\mathcal{Y}$ if $\mathcal{Y}$ is a subset of $\mathcal{X}$.*

In other words, if $\mathbf{U} \in \mathcal{Y}$, then $\mathbf{U} \in \mathcal{X}$ also. We write this: "$\mathcal{Y} \subset \mathcal{X}$".

Example 156:

(a) **Partitions:** Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions of $\mathbf{X}$. Then $\sigma(\mathcal{Q})$ refines $\sigma(\mathcal{P})$ if and only if $\mathcal{Q}$ refines $\mathcal{P}$ (<u>**Exercise 151**</u>).

(b) **The power set:** The sigma-algebra $\mathcal{P}(\mathbf{X})$ contains the "most" information about $\mathbf{X}$, because it refines every sigma-subalgebra $\mathcal{Y}$. Thus, $\mathcal{P}(\mathbf{X})$ represents the state of 'omniscience'.

(c) The **trivial** sigma-algebra $\mathcal{X}_\emptyset := \{\emptyset, \mathbf{X}\}$ contains the *least* information, because every other sigma algebra on the space refines $\mathcal{X}_\emptyset$. Thus, $\mathcal{X}_\emptyset$ represents the state of 'total ignorance'.

(d) **The cube:** Let $\mathbb{I}^3$ be the unit cube, with Borel sigma-algebra $\mathcal{I}^3$. Recall that $\mathcal{X}_{12} = {\mathbf{pr}_{1,2}}^{-1}(\mathcal{I}^2)$ contained all information about the (x_1, x_2) coordinates of an unknown point $\mathbf{x} \in \mathbf{X}$. However, the sigma-algebra $\mathcal{X}_1 := {\mathbf{pr}_1}^{-1}(\mathcal{I}^1)$ only contained information about the x_1 coordinate.

$$\mathcal{X}_1 \subset \mathcal{X}_{12} \subset \mathcal{I}^3$$

$\mathcal{X}_{12}$ contains less information than $\mathcal{I}^3$ because $\mathcal{X}_{12}$ tells us nothing about the x_3 coordinate. Further, $\mathcal{X}_1$ contains only only half the horizontal information contained in $\mathcal{X}_{12}$, because $\mathcal{X}_1$ contains only information x_1 coordinate, but says nothing about the x_2 coordinate.

Example 157: Stock market

Imagine you are monitoring the price of a certain stock over a 30 day period. Hence, the behaviour of the stock will be described by an (as-yet unknown) 30-element sequence of numbers $\mathbf{p} = (p_1, p_2, \ldots, p_{30}) \in \mathbb{R}^{30}$. On the first day, you learn the value of p_1, on the second day, that of p_2, and so on. Thus, after the first n days, you know the value of $p_1, \ldots, p_n$.

Let $\mathbf{pr}_n : \mathbb{R}^{30} \longrightarrow \mathbb{R}^n$ be the projection: $\mathbf{pr}_n(\mathbf{p}) = (p_1, \ldots, p_n)$. Let $\mathcal{B}^n$ be the Borel sigma-algebra of $\mathbb{R}^n$, and let $\mathcal{X}_n = {\mathbf{pr}_n}^{-1}(\mathcal{B}_n)$. Thus, your knowledge state on day n is contained in the sigma-algebra $\mathcal{X}_n$, and we have a *refining sequence* of sigma-algebras:

$$\mathcal{X}_\emptyset \subset \mathcal{X}_1 \subset \mathcal{X}_2 \subset \ldots \subset \mathcal{X}_{30}$$

where $\mathcal{X}_\emptyset = \{\emptyset, \mathbb{R}\}$ is the trivial sigma-algebra, and $\mathcal{X}_{30} = \mathcal{B}_{30}$ is the whole Borel sigma-algebra of $\mathbb{R}^{30}$.

Thus, a progressive revelation of information corresponds to a a refining sequence of sigma-algebras...

Definition 158 Filtration

Let $(\mathbf{X}, \mathcal{X})$ be a measurable space, and let $\mathbb{T}$ be a linearly ordered set. A **$\mathbb{T}$-indexed filtration** *is a collection $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ of sigma subalgebras of $\mathcal{X}$ so that, for any $s, t \in \mathbb{T}$.*

$$\Big(s < t \Big) \Longrightarrow \Big(\mathcal{F}_s \subset \mathcal{F}_t \Big)$$

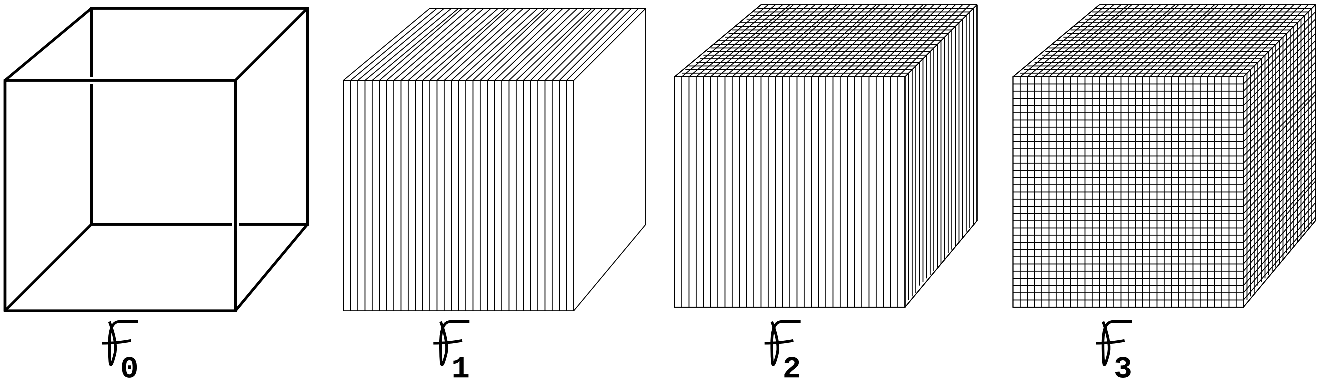


Figure 5.8: $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ is a filtration of the Borel-sigma algebra of the cube.

Normally, $\mathbb{T}$ is interpreted as *time*. Thus, a filtration represents a revelation of information unfolding in time.

Example 159:

(a) In the stockmarket of Example 157, $\mathbb{T} = [0..30]$, and $\mathcal{F}_t = \mathcal{X}_t$.

(b) Recall the dyadic partitions of Example 153. Here, $\mathbb{T} = \mathbb{N}$, and $\mathcal{F}_n = \sigma(\mathcal{P}_n)$ is a sigma-algebra containing information about the first n binary digits of an unknown real number $\alpha \in [0, 1]$.

(c) Recall the cube projections of Example 154. Let $\mathbb{T} = \{0, 1, 2, 3\}$, and define

$$\mathcal{F}_0 = \{\emptyset, \mathbb{I}^3\}; \qquad \mathcal{F}_1 = \mathcal{X}_1; \qquad \mathcal{F}_2 = \mathcal{X}_{1,2}; \qquad \mathcal{F}_3 = \mathcal{I}^3;$$

(see Figure 5.8). Then $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ is a four-element filtration.

5.1(b) Conditional Probability and Independence

Prerequisites: §5.1

Opposite to the concept of refinement is the concept of *independence*. If $\mathcal{Y}$ contains *no* information about $\mathcal{X}$, and $\mathcal{X}$ contains *no* information about $\mathcal{Y}$, then the information contained in the two sigma-algebras is *complementary*; each one tells us things the other one doesn't tell us. Heuristically, if we knew about both the $\mathcal{X}$-related information *and* the $\mathcal{Y}$-related information, then we would have 'twice' as much information as if we knew only one or the other. We say that $\mathcal{X}$ and $\mathcal{Y}$ are **independent** of one another.

To make this concept precise, we must fix a measure upon $\mathbf{X}$, which determines whether or not two peices of information are probabilistically *correlated* with each other. Depending upon the measure we place on $\mathbf{X}$, sigma algebras $\mathcal{X}$ and $\mathcal{Y}$ may range from being totally independent of one another, to being 'effectively identical'.

First, we must introduce the notion of conditional probability. Suppose that, over a historical period of 10000 days:

- It rained in Toronto on 4500 days.
- It rained in Montréal on 3500 days.
- It rained in Toronto *and* Montréal on 3000 days.

Assuming this sample accurately reflects the underlying statistics, we can conclude, for example:

> The probability that it will rain in Montréal on any given day is $\dfrac{3500}{10000} \;=\; 0.35.$

This probability estimate is made assuming 'total ignorance'. Suppose, however, that you aready *knew* it was raining in Toronto. This might modify your wager about Montréal. Out of the 4500 days during which it rained in Toronto, it rained in Montréal on 3500 of those days. Thus,

> *Given* that it is raining in Toronto, the probability that it will *also* rain in Montréal, is $\dfrac{3000}{4500} = \dfrac{2}{3} \;=\; 0.6666\ldots.$

If $\mathbf{T}$ is the event "It is raining in Toronto", and $\mathbf{M}$ is the event "It is raining in Montréal", then $\mathbf{M} \cap \mathbf{T}$ is the event "It is raining in Toronto *and* Montréal". What we have just concluded is:

$$\mathsf{Prob}\,[\mathbf{M} \text{ given } \mathbf{T}] \quad = \quad \frac{\mathsf{Prob}\,[\mathbf{M} \cap \mathbf{T}]}{\mathsf{Prob}\,[\mathbf{T}]}$$

Definition 160 Conditional Probability

Let $(\mathbf{X}, \mathcal{X}, \mu)$ *be a measure space, and* $\mathbf{U}, \mathbf{V} \in \mathcal{X}$. *The* **conditional probability**[2] *of* $\mathbf{U}$, *given* $\mathbf{V}$ *is defined:*

$$\mu[\mathbf{U} \;\langle\!\langle\; \mathbf{V}] \quad = \quad \frac{\mu[\mathbf{U} \cap \mathbf{V}]}{\mu[\mathbf{V}]}$$

In the previous example, meteorological information about Toronto modified our wager about Montréal. Suppose instead that the statistics were as follows: over 10000 days,

- It rained in Toronto on 5000 days.
- It rained in Montréal on 3000 days.
- It rained in Toronto *and* Montréal on 1500 days.

Then we conclude:

- The probability that it will rain in Montréal on any given day is $\dfrac{3000}{10000} \;=\; 0.3.$

[2] This terminology obviously refers to the case when μ is a probability measure, but we will apply it even when μ is not.

- *Given* that it is raining in Toronto, the probability that it will *also* rain in Montréal, is $\frac{1500}{5000} = 0.3$.

In other words, the rain in Toronto has *no influence* on the rain in Montréal. Meteorological information from Toronto is *useless* to predicting Montréal precipitation. If $\mathbf{M}$ and $\mathbf{T}$ are as before, we have:

$$\frac{\mathsf{Prob}\left[\mathbf{M}\cap\mathbf{T}\right]}{\mathsf{Prob}\left[\mathbf{T}\right]} \quad = \quad \mathsf{Prob}\left[\mathbf{M}\right].$$

Another way to write this:

$$\mathsf{Prob}\left[\mathbf{M}\cap\mathbf{T}\right] \quad = \quad \mathsf{Prob}\left[\mathbf{M}\right]\cdot\mathsf{Prob}\left[\mathbf{T}\right].$$

Definition 161 Independence of sets

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and $\mathbf{U}, \mathbf{V} \in \mathcal{X}$. We say $\mathbf{U}$ and $\mathbf{V}$ are **independent** *if*

$$\mu\left[\mathbf{U}\cap\mathbf{V}\right] \;=\; \mu[\mathbf{U}]\cdot\mu[\mathbf{V}]$$

Instead of examining single sets, however, we might look at an entire sigma-algebra. For example, suppose $\mathcal{T}$ was a sigma-algebra representing *all possible weather events* in Toronto, and $\mathcal{M}$ was a sigma-algebra representing all possible weather events in Montréal. Suppose we found that, not only was the *precipitation* between the two cities unrelated, but in fact, *all* weather events were unrelated. For example, if $\mathcal{T}$ might contains sets representing statements like "There is a tornado in Toronto", and $\mathcal{M}$ might contain, "It is freezing in Montréal." To say that the weather in the two cities is totally unrelated is to say that *every* set $\mathbf{T} \in \mathcal{T}$ is independent of *every* set $\mathbf{M} \in \mathcal{M}$.

Definition 162 Independence of sigma-algebras

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and and let $\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{X}$ be two sigma subalgebras.

$\mathcal{Y}_1$ and $\mathcal{Y}_2$ are **independent** *(with respect to μ) if every element of $\mathcal{Y}_1$ is* independent *of every element of $\mathcal{Y}_2$, with respect to the measure μ. That is,*

$$\textit{For all } \mathbf{U}_1 \in \mathcal{Y}_1 \textit{ and } \mathbf{U}_2 \in \mathcal{Y}_2, \qquad \mu\left[\mathbf{U}_1\cap\mathbf{U}_2\right] \quad = \quad \mu[\mathbf{U}_1]\cdot\mu[\mathbf{U}_2].$$

We write this: "$\mathcal{Y}_1 \perp \mathcal{Y}_2$".

Example 163:

(a) **The Square with Lebesgue Measure:**

Consider the unit square $\mathbb{I}^2$, with sigma-algebra $\mathcal{I}^2$, and the Lebesgue probability measure λ^2. Let:

$$\begin{aligned} \mathcal{X}_1 &= \mathbf{pr}_1\mathcal{I} && \text{('horizontal' information)} \\ \mathcal{X}_2 &= \mathbf{pr}_2\mathcal{I} && \text{('vertical' information)} \end{aligned}$$

Then $\mathcal{X}_1$ and $\mathcal{X}_2$ are independent. (**Exercise 152**)

(see Figure 5.9 on the facing page)

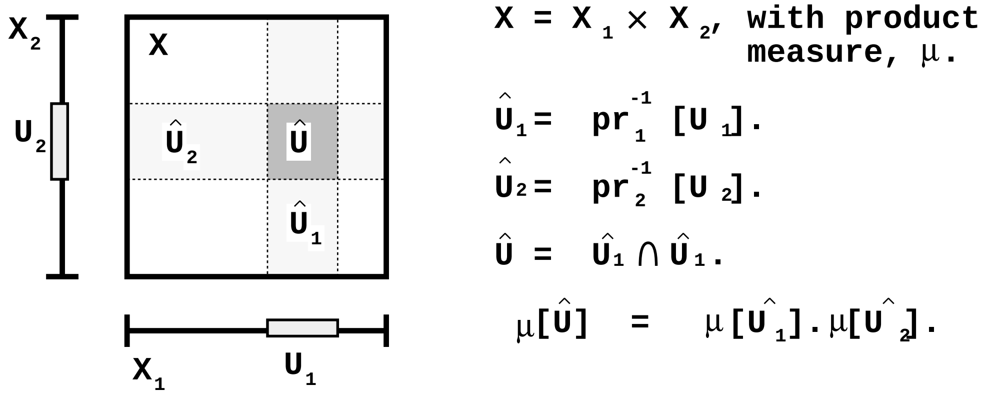


Figure 5.9: With respect to the **product measure**, the horizontal and vertical sigma-algebras are **independent**.

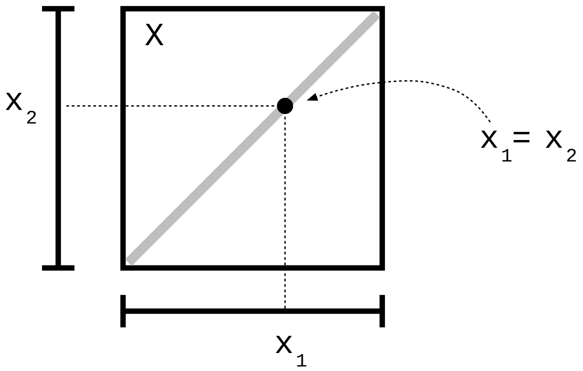


Figure 5.10: The measure μ is supported only on the diagonal, and thus, "correlates" horizontal and vertical information.

(b) **The Square with Diagonal Measure:** Again consider $(\mathbb{I}^2, \mathcal{I}^2)$, but now with the *diagonal* measure λ_Δ defined:

$$\lambda_\Delta(\mathbf{U}) \;=\; \mu\{x \in \mathbb{I} \;;\; (x,x) \in \mathbf{U}\}$$

(see Figure 1.14 on page 23). Let $\Delta = \{(x,x) \in \mathbb{I}^2\}$ be the 'diagonal' subset of the square. Then $\lambda_\Delta(\mathbf{U})$ simply measures the size of $\mathbf{U} \cap \Delta$. In particular, notice that:

$$\lambda_\Delta\left[\mathbb{I}^2 \setminus \Delta\right] \;=\; 0$$

In other words, the set of all points (x,y) where $x \neq y$ has *probability zero* with respect to λ_Δ. To put it another way:

With probability one, the first coordinate and second coordinate are equal.

(see Figure 5.10 on the preceding page)

Thus if you have all $\mathcal{X}_1$ information, you *automatically* have all $\mathcal{X}_2$ information as well, because λ_Δ tells you that the first coordinate and second coordinate must be equal. —

Definition 164 Effective Refinement

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{X}$ be two sigma subalgebras.

$\mathcal{Y}_2$ **effectively refines** $\mathcal{Y}_1$ *(with respect to μ) if every element of $\mathcal{Y}_1$ can be "approximated" by some element of $\mathcal{Y}_2$, so that the difference set has measure zero according to μ. That is,*

$$\textit{For all } \mathbf{U}_1 \in \mathcal{Y}_1 \textit{ there exists } \mathbf{U}_2 \in \mathcal{Y}_2, \textit{ so that } \mu\left[\mathbf{U}_1 \triangle \mathbf{U}_2\right] \;= 0.$$

We write this: "$\mathcal{Y}_1 \subset_\mu \mathcal{Y}_2$".

For example, in Example 163b, each of $\mathcal{X}_1$ and $\mathcal{X}_2$ effectively refines the other (**Exercise 153**).

5.2 Conditional Expectation

Prerequisites: §1.1 **Recommended:** §5.1

5.2(a) Blurred Vision...

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and $f : \mathbf{X} \longrightarrow \mathbb{C}$ a $\mathcal{X}$ -measurable function. Let $\mathcal{Y} \subset \mathcal{X}$ be a *sigma-subalgebra*. In general, f will *not* be measurable with respect to $\mathcal{Y}$; we want to 'approximate' f as closely as possible with a $\mathcal{Y}$-measurable function, g. We want g to have two properties:

(CE1) g is $\mathcal{Y}$-measurable.

(CE2) For any subset $\mathbf{U} \in \mathcal{Y}$, we have $\int_{\mathbf{U}} g \ d\mu \ = \ \int_{\mathbf{U}} f \ d\mu$.

To understand property **(CE2)**, recall that $\mu_{|_{\mathcal{Y}}} : \mathcal{Y} \longrightarrow [0, \infty]$ is a measure on $\mathcal{Y}$, so that the $\left(\mathbf{X}, \mathcal{Y}, \mu_{|_{\mathcal{Y}}}\right)$ is *also* a measure space. So, since g is $\mathcal{Y}$-integrable by **(CE1)**, it makes sense to talk about its integral with respect to μ, as long as we only integrate over subsets $\mathbf{U}$ which are in $\mathcal{Y}$.

If such a function g exists, it is essentially unique...

Claim: *If g_1 and g_2 both satisfy* **(CE1)** *and* **(CE2)**, *then $g_1 = g_2$, a.e.$[\mu]$. In other words, if $\Delta = \{x \in \mathbf{X} \ ; \ g_1(x) \neq g_2(x)\}$, then Δ is a $\mathcal{Y}$-measurable and $\mu(\Delta) = 0$.*

Proof: <u>**Exercise 154**</u> $\square$

The function g is a sort of 'blurring' of f. It represents the 'best estimate possible' of the value of f, given the limited information provided by $\mathcal{Y}$.

Definition 165 Conditional Expectation

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathcal{Y}$ be a sigma-subalgebra of $\mathcal{X}$. Let $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$. The **conditional expectation** *of f with respect to $\mathcal{Y}$ is the unique function satisfying conditions* **(CE1)** *and* **(CE2)** *above. This function is denoted "$\mathbb{E}_{\mathcal{Y}}[f]$".*

Definition 166 Conditional Measure

If $\mathbf{U} \subset \mathbf{X}$ is measurable, the **conditional measure** *of $\mathbf{U}$ with respect to $\mathcal{Y}$ is the* conditional expectation *of $\mathbb{1}_{\mathbf{U}}$.*

If $(\mathbf{X}, \mathcal{X}, \mu)$ is a probability space, we call this the **conditional probability** *of $\mathbf{U}$.*

The conditional measure of $\mathbf{U}$ is written as "$\mu[\mathbf{U} \ \langle\!\langle \ \mathcal{Y}]$" (not as "$\mu_{\mathcal{Y}}[\mathbf{U}]$", since this could cause confusion). Formally: $\mu[\mathbf{U} \ \langle\!\langle \ \mathcal{Y}] \ := \ \mathbb{E}_{\mathcal{Y}}[\mathbb{1}_{\mathbf{U}}]$.

The notation for conditional measure is confusing. Do not forget the fact that $\mu[\mathbf{U} \ \langle\!\langle \ \mathcal{Y}]$ is a *function*, not just a number.

5.2(b) Some Examples

Prerequisites: §5.2(a),§1.3

We have not yet shown that a function $\mathbb{E}_{\mathcal{Y}}[f]$ satisfying **(CE1)** and **(CE2)** exists. The examples in this section prove the existence of $\mathbb{E}_{\mathcal{Y}}[f]$ in certain special cases, and also provide intuition about its properties.

Example 167: Conditional Expectation with respect to a Partition

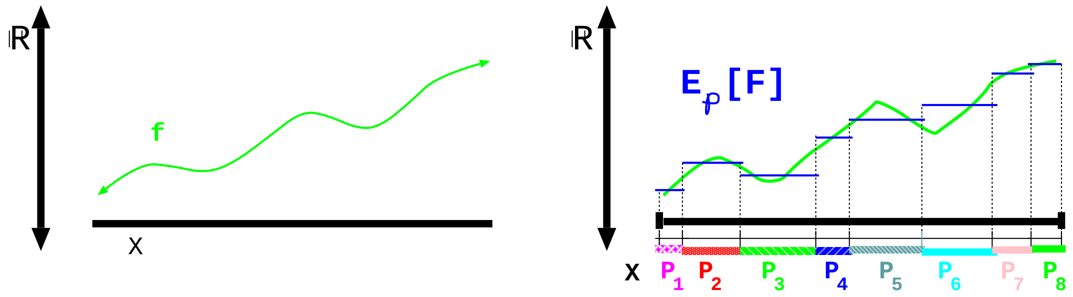


Figure 5.11: The conditional expectation of f is constant on each atom of the partition $\mathcal{P}$. (A one-dimensional example)

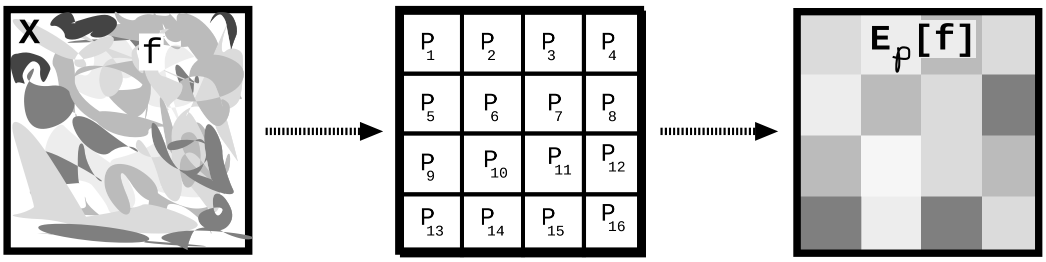


Figure 5.12: The conditional expectation of f is constant on each atom of the partition $\mathcal{P}$. (A two-dimensional example)

Suppose that $\mathcal{P}$ is a partition of $\mathbf{X}$, and let $\sigma(\mathcal{P})$ be the corresponding sigma-algebra. We then use the notation:

$$\mathbb{E}_{\mathcal{P}}[f] \quad := \quad \mathbb{E}_{\sigma(\mathcal{P})}[f] \qquad \text{and} \qquad \mu[\mathbf{U} \ll \mathcal{P}] \quad := \quad \mu[\mathbf{U} \ll \sigma(\mathcal{P})]$$

Consider an atom $\mathbf{P} \in \mathcal{P}$. It is left as **Exercise 155** to prove the following properties:

1. The conditional expectation $\mathbb{E}_{\mathcal{Y}}[f]$ must be **constant** on $\mathbf{P}$. (see Figures 5.11 and 5.12)

2. The value of $\mathbb{E}_{\mathcal{Y}}[f]$ on $\mathbf{P}$ is given by: $\mathbb{E}_{\mathcal{Y}}[f] \quad = \quad \dfrac{1}{\mu[\mathbf{P}]} \displaystyle\int_{\mathbf{P}} f \;\; d\mu.$

3. For any set $\mathbf{U} \subset \mathbf{X}$, the value of $\mu[\mathbf{U} \ll \mathcal{P}]$ on $\mathbf{P}$ is given by: $\mu[\mathbf{U} \ll \mathbf{P}] \quad = \quad \dfrac{\mu[\mathbf{U} \cap \mathbf{P}]}{\mu[\mathbf{P}]}$, that is, the **conditional probability** of $\mathbf{U}$ on $\mathbf{P}$. This quantity is the answer to the question:

   > "Given that you already know you are inside $\mathbf{P}$, what is the probability that you are *also* in $\mathbf{U}$?"

   If $\mathcal{P}$ as defines a *grid* on the space $\mathbf{X}$, then the function $\mu[\mathbf{U} \ll \mathcal{P}]$ is a 'pixelated' version of the characteristic function of $\mathbf{U}$. (See Figure 5.13) x

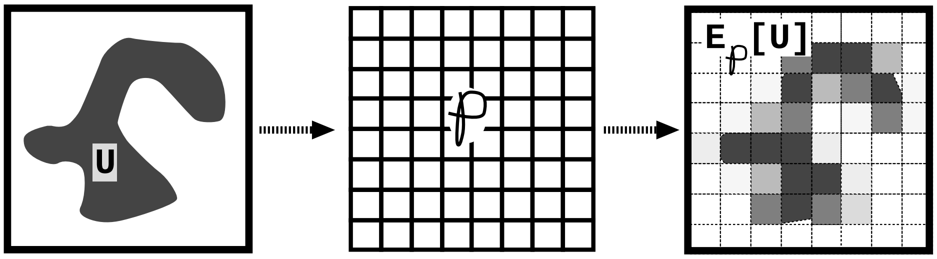


Figure 5.13: The conditional measure of the set **U**, with respect to the "grid" partition $\mathcal{P}$, looks like a low-resolution "pixelated" version of the characteristic function of **U**.

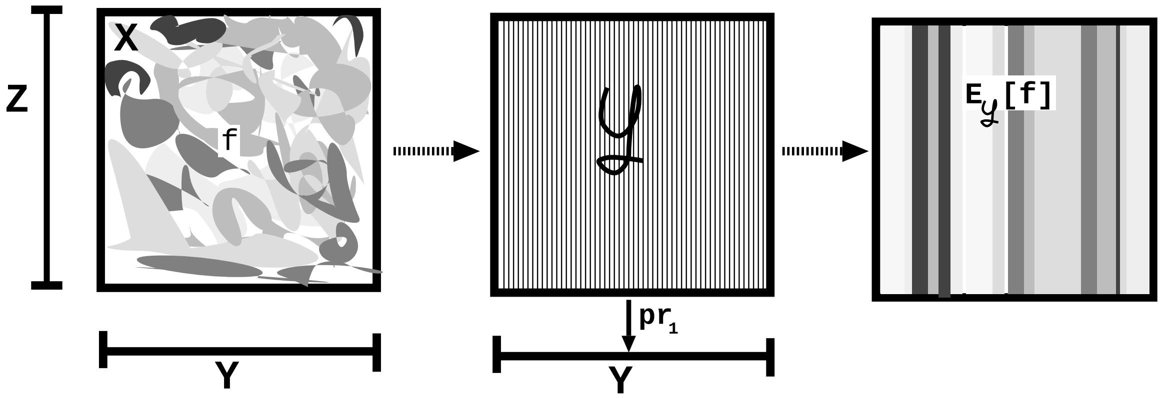


Figure 5.14: The conditional expectation is constant on each fibre.

Example 168: A Product Space

Let $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ and $(\mathbf{Z}, \mathcal{Z}, \zeta)$ be measure spaces, and consider the product space $(\mathbf{X}, \mathcal{X}, \xi)$:

$$\mathbf{X} \;=\; \mathbf{Y} \times \mathbf{Z}, \qquad \mathcal{X} \;=\; \mathcal{Y} \otimes \mathcal{Z}, \qquad \text{and} \quad \xi \;=\; \Upsilon \times \zeta,$$

Consider the sigma-algebra of "vertical" subsets of $\mathbf{X}$,

$$\widehat{\mathcal{Y}} \;:=\; \mathcal{Y} \otimes \{\mathbf{Z}\} \;=\; \{\mathbf{V} \times \mathbf{Z}\,;\; \mathbf{V} \in \mathcal{Y}\}$$

and let $f : \mathbf{X} \longrightarrow \mathbb{R}$. For any $y \in \mathbf{Y}$, let $\mathbf{F}_y \;:=\; \{y\} \times \mathbf{Z} \subset \mathbf{X}$ be the **fibre** over y. It is left as **Exercise 156** to prove the following properties:

1. $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ is constant on each fibre $\mathbf{F}_y$ (see Figure 5.14). Thus, there is a unique function $f_{\mathbf{Y}} \in \mathbf{L}^1(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ so that:
   (a) The value of $f_{\mathbf{Y}}(y)$ is equal to the (constant) value of $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ on $\mathbf{F}_y$.
   (b) For any $\mathbf{V} \subset \mathbf{Y}$, if $\mathbf{U} = \mathbf{V} \times \mathbf{Z}$, then $\displaystyle\int_{\mathbf{V}} f_{\mathbf{Y}} \; d\Upsilon \;=\; \int_{\mathbf{U}} \mathbb{E}_{\widehat{\mathcal{Y}}}[f] \; d\xi \;=\; \int_{\mathbf{U}} f \; d\xi.$
2. Suppose $\mathbf{U} \subset \mathbf{X}$. For any $y \in \mathbf{Y}$, we can identify $\mathbf{F}_y$ with $\mathbf{Z}$, and endow it with a **fibre measure** ζ_y, identical to ζ. Then the value of $\mu[\mathbf{U} \;\langle\!\langle\; \widehat{\mathcal{Y}}]$ on $\mathbf{F}_y$ is:

   $$\zeta_y\left[\mathbf{U} \cap \mathbf{F}_x\right].$$

   Metaphorically speaking,the conditional probability $\mu[\mathbf{U} \;\langle\!\langle\; \widehat{\mathcal{Y}}]$ measures the 'shadow' that $\mathbf{U}$ casts down upon $\mathbf{Y}$ (see Figure 5.15).

Thus we can identify $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ with $f_{\mathbf{Y}}$, and think of it as a sort of "projection" of the function f down to the factor space $\mathbf{Y}$.

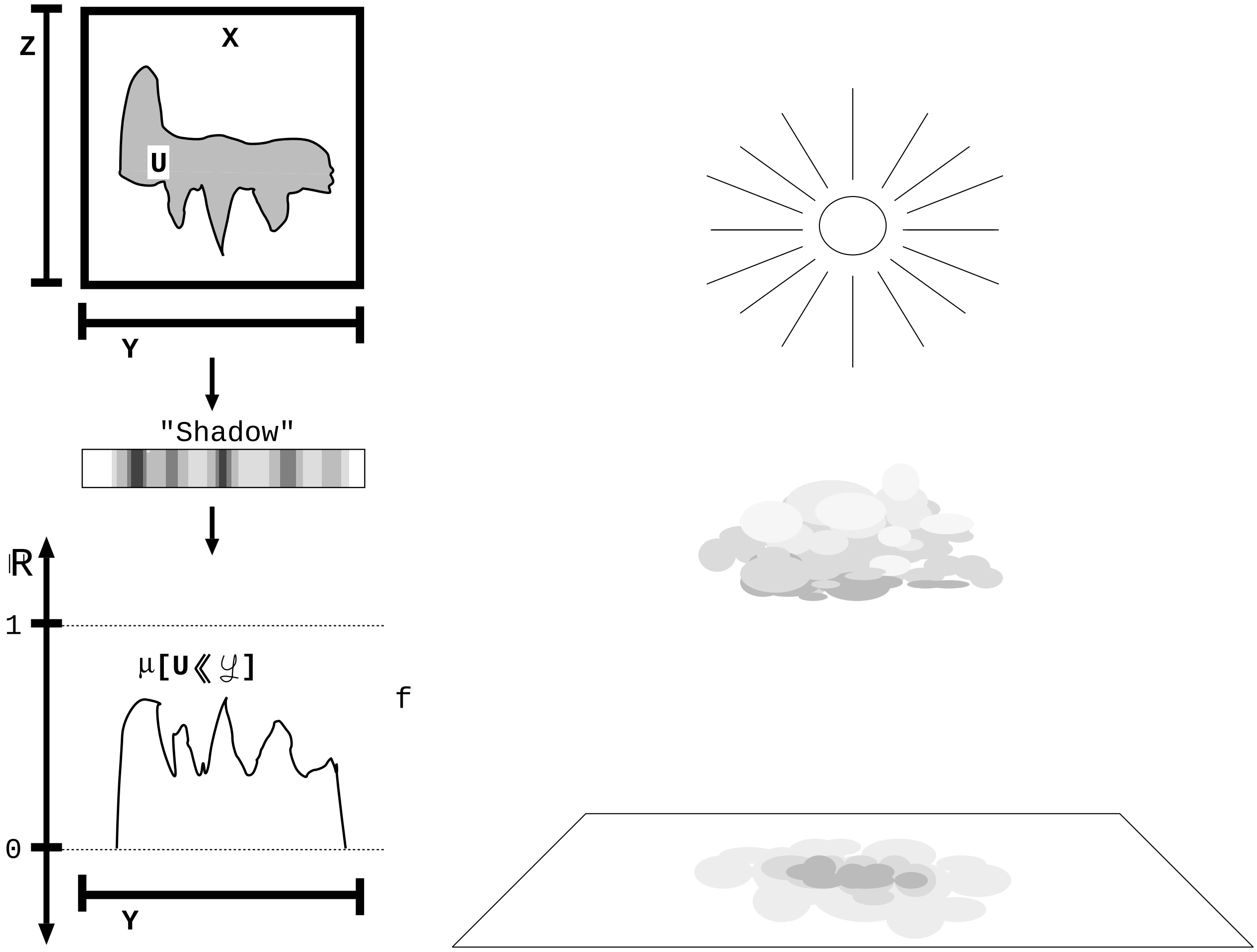
Example 169: Projection through a morphism

Let $(\mathbf{X}, \mathcal{X}, \xi)$ and $(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be measure spaces, and let $P : (\mathbf{X}, \mathcal{X}, \xi) \longrightarrow (\mathbf{Y}, \mathcal{Y}, \Upsilon)$ be a measure-preserving map. Define

$$\widehat{\mathcal{Y}} \;:=\; P^{-1}(\mathcal{Y}) \;\subset\; \mathcal{X}.$$

If $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \xi)$, then we can consider its conditional expectation with respect to $\widehat{\mathcal{Y}}$. This is just a generalization of the previous example. For any $y \in \mathbf{Y}$, let $\mathbf{F}_y \;:=\; P^{-1}\{y\}$ be the **fibre** over y; then it is **Exercise 157** to prove the following properties:

1. $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ is constant on each fibre of P. Thus, there is a unique function $f_{\mathbf{Y}} \in \mathbf{L}^1(\mathbf{Y}, \mathcal{Y}, \Upsilon)$ so that:
   (a) $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ is constant on $\mathbf{F}_y$, and equal to the value of $f_{\mathbf{Y}}(y)$.
   (b) For any $\mathbf{V} \subset \mathbf{Y}$, if $\mathbf{U} = P^{-1}(\mathbf{V})$, then $\displaystyle\int_{\mathbf{V}} f_{\mathbf{Y}} \; d\Upsilon \;=\; \int_{\mathbf{U}} \mathbb{E}_{\widehat{\mathcal{Y}}}[f] \; d\xi \;=\; \int_{\mathbf{U}} f \; d\xi.$



(A) (B)

Figure 5.15: **(A)** The conditional probability of **U** is like a "shadow" cast upon **Y**. **(B)** The shadow of a cloud.

We call $\mathbb{E}_{\widehat{\mathcal{Y}}}[f]$ the conditional expectation *onto* $\mathbf{Y}$, or, alternately, as the conditional expectation *through* P, and write "$\mathbb{E}_{\mathbf{Y}}[f]$" or "$\mathbb{E}_P[f]$".

Example 170: The Shadow of a Cloud

Figure 5.15(B) shows a cloud is floating in a clear sky; let $f : \mathbb{R}^3 \longrightarrow [0, \infty)$ be a function describing the *density* of the cloud at any point in space. Let $\mathbf{P} : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be the orthogonal projection down to the ground (which we assume is flat). Suppose that the sun is directly overhead, and suppose further that the sunlight passing through the cloud is attenuated at a rate proportional to the cloud density. Then the *shadow* cast by the cloud upon the ground will be the conditional expectation of f through P.

5.2(c) Existence in $\mathbf{L}^2$

Prerequisites: §5.2(a),[Hilbert Spaces]

We here establish the existence of $\mathbb{E}_{\mathcal{Y}}[f]$ for any function $f \in \mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$.

Since $(\mathbf{X}, \mathcal{Y}, \mu|_{\mathcal{Y}})$ is a measure space in its own right, we can consider the associated Hilbert space $\mathbf{L}^2(\mathbf{X}, \mathcal{Y}, \mu|_{\mathcal{Y}})$, which consists of square-integrable, $\mathcal{Y}$-measurable functions on $\mathbf{X}$.

Lemma 171: *$\mathbf{L}^2(\mathbf{X}, \mathcal{Y}, \mu|_{\mathcal{Y}})$ is a closed linear subspace of $\mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$.*

Proof: **Exercise 158** □

Theorem 172: *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and $\mathcal{Y}$ a sigma-subalgebra of $\mathcal{X}$. Let $\mathbf{P} : \mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu) \longrightarrow \mathbf{L}^2(\mathbf{X}, \mathcal{Y}, \mu)$ be the orthogonal projection map.*

Then for any $f \in \mathbf{L}^2(\mathbf{X}, \mathcal{X}, \mu)$, we have: $\mathbb{E}_{\mathcal{Y}}[f] \;=\; \mathbf{P}(f)$.

Proof: **Exercise 159** Hint: It suffices to show that $\mathbf{P}(f)$ satisfies **(CE1)** and **(CE2)**. □

5.2(d) Existence in $\mathbf{L}^1$

Prerequisites: §5.2(a),[Radon-Nikodym theorem]

We here establish the existence of $\mathbb{E}_{\mathcal{Y}}[f]$ for any function $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$.

Recall that f defines a measure μ_f on $\mathcal{X}$ by:

$$\mu_f[\mathbf{U}] \;:=\; \int_U f \; d\mu.$$

By construction, μ_f is absolutely continous with respect to μ, with Radon-Nikodym derivative:

$$\frac{d\mu_f}{d\mu} \;=\; f.$$

Let $\mu|_{\mathcal{Y}}$ be the restriction of μ to a measure on $\mathcal{Y}$, and let $(\mu_f)|_{\mathcal{Y}}$ be the restriction of μ_f.

Theorem 173: *Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and $\mathcal{Y}$ a sigma-subalgebra of $\mathcal{X}$. Let $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$.*

The measure $(\mu_f)\,|_{\mathcal{Y}}$ is absolutely continuous with respect to $\mu|_{\mathcal{Y}}$. The conditional expectation of f is then the Radon-Nikodym derivative of $(\mu_f)\,|_{\mathcal{Y}}$ relative to $\mu|_{\mathcal{Y}}$:

$$\mathbb{E}_{\mathcal{Y}}[f] \quad := \quad \frac{d\,(\mu_f)\,|_{\mathcal{Y}}}{\mu|_{\mathcal{Y}}}.$$

Proof: **Exercise 160** Hint: It suffices to show that $\frac{d(\mu_f)|_{\mathcal{Y}}}{\mu|_{\mathcal{Y}}}$ satisfies **(CE1)** and **(CE2)**. □

5.2(e) Properties of Conditional Expectation

Prerequisites: §5.2(a)

Throughout this section, $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, $f \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$, and $\mathcal{Y} \subset \mathcal{X}$ is a sigma-subalgebra.

Theorem 174:

1. *If f is $\mathcal{Y}$-measurable, then* $\mathbb{E}_{\mathcal{Y}}[f] \;=\; f$.
2. *In particular,* $\mathbb{E}_{\mathcal{Y}}\Big[\mathbb{E}_{\mathcal{Y}}[f]\Big] \;=\; \mathbb{E}_{\mathcal{Y}}[f]$.
3. *More generally, if $\mathcal{Y}_1 \subset \mathcal{Y}_2$, then* $\mathbb{E}_{\mathcal{Y}_1}\Big[\mathbb{E}_{\mathcal{Y}_2}[f]\Big] \;=\; \mathbb{E}_{\mathcal{Y}_1}[f]$.

Proof: **Exercise 161** □

Example 175: Fubini-Tonelli Theorem

Consider a product space $(\mathbf{X}, \mathcal{X}, \mu)$, where $\mathbf{X} = \mathbf{Y} \times \mathbf{Z}$, $\mathcal{X} = \mathcal{Y} \otimes \mathcal{Z}$, and $\mu = \Upsilon \times \zeta$, as in Example 168. Let $\widehat{\mathcal{Y}} := \mathcal{Y} \otimes \{\mathbf{Z}\}$ as in that example.

Set $\mathcal{Y}_2 = \widehat{\mathcal{Y}}$ and $\mathcal{Y}_1 = \{\emptyset, \mathbf{X}\}$ be the trivial algebra. Observe that part (3) of Theorem 174 is then equivalent to the Fubini-Tonelli theorem. (**Exercise 162**)

Part (3) of the Theorem 174 is one of the most often-used facts about conditional expectation, and comes up constantly in probability theory. It says this: if first you learn fact **A**, and then you later learn fact **B** (which happens to *imply* fact **A** as a corollary), then your final state of knowledge is basically the same as if you had just been told fact **B** to begin with.

Theorem 176: Orthogonal Sigma algebras

1. *If f is* **independent** *of $\mathcal{Y}$ then $\mathbb{E}_{\mathcal{Y}}[f]$ is* **constant**, *with value* $\displaystyle\int_X f \, d\mu$.
2. *In particular, if $\mathcal{Y}_1 \perp \mathcal{Y}_2$, then $\mathbb{E}_{\mathcal{Y}_1}[\mathbb{E}_{\mathcal{Y}_2}[f]]$ is constant, with value* $\displaystyle\int_X f \, d\mu$.

Proof: **Exercise 163** □

Theorem 177: Algebraic Properties

1. *The conditional expectation operator is linear. That is, for all $f_1, f_2 \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$ and $c_1, c_2 \in \mathbb{C}$,*
$$\mathbb{E}_{\mathcal{Y}}\left[c_1 \cdot f_1 + c_2 \cdot f_2\right] \quad = \quad c_1 \cdot \mathbb{E}_{\mathcal{Y}}[f_1] + c_2 \cdot \mathbb{E}_{\mathcal{Y}}[f_2].$$
2. *If f is $\mathcal{Y}$ -measurable, then for any $g \in \mathbf{L}^1(\mathbf{X}, \mathcal{X}, \mu)$,*
$$\mathbb{E}_{\mathcal{Y}}[f \cdot g] \quad = \quad f \cdot \mathbb{E}_{\mathcal{Y}}[g]$$

Proof: **Exercise 164** □

Part (2) of Theorem 177 is one of the other most often used fact about conditional expectation. To appreciate the meaning of this result, think about its implications in some of the concrete example discussed in §5.2(b). For example, consider the case when $\mathcal{Y}$ is generated by a partition (thus, f is constant on each element of the partition), or when $\mathcal{Y}$ is the pulled-back sigma algebra of some measure-preserving map (so that f is constant on each fibre).

Theorem 178: Banach Space Properties

For every $p \in [1..\infty]$, the conditional expectation operator induces a **bounded linear map**
$$\mathbb{E}_{\mathcal{Y}} : \mathbf{L}^p(\mathbf{X}, \mathcal{X}, \mu) \longrightarrow \mathbf{L}^p(\mathbf{X}, \mathcal{Y}, \mu)$$
of norm 1.

Proof: **Exercise 165** □

Definition 179 Convex Function

Let $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$. Then Φ is **convex** *if, for all $x_1, \ldots, x_N \in \mathbb{R}$, and any $\lambda_1, \ldots \lambda_N \in [0,1]$ with $\displaystyle\sum_{n=1}^{N} \lambda_n = 1$, we have:* $\displaystyle\Phi\left(\sum_{n=1}^{N} \lambda_n \cdot x_n\right) \quad \leq \quad \sum_{n=1}^{N} \lambda_n \cdot \Phi(x_n)$. *(see Figure 5.16)*

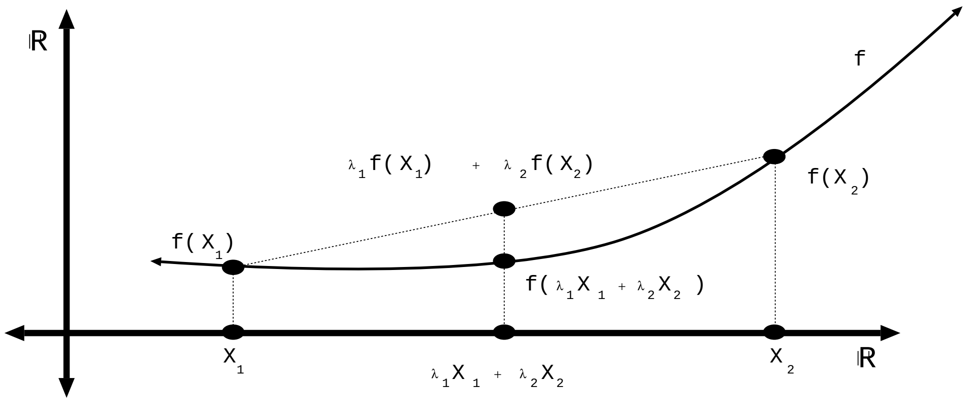


Figure 5.16: A convex function.

Theorem 180: Jensen's Inequality

If $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$ *is any integrable, convex function, then* $\Phi\left(\mathbb{E}_{\mathcal{Y}}[f]\right) \quad \leq \quad \mathbb{E}_{\mathcal{Y}}[\Phi \circ f]$.

Proof: **Exercise 166** $\square$

Chapter 6

Measure Algebras

6.1 Sigma Ideals

Prerequisites: §1.1

Recommended: §1.3(c)

In the algebraic theory of rings, an ideal is the kernel of a ring homomorphism. In other words, an ideal is something that you *mod out* by; it is a set of objects which are identified with zero.

Sometimes a particular construction (or assertion) is not well-defined (or true) everywhere in a domain $\mathbf{X}$, but only 'almost' everywhere. There is perhaps some 'bad' set $\mathbf{B} \subset \mathbf{X}$ where the construction (or assertion) fails. However, if $\mathbf{B}$ is 'small enough', then this may not matter; the set $\mathbf{B}$ can be considered 'negligible'. One example of this is in §1.3(c), where we showed how sets of measure zero can be treated as 'negligible'.

The mathematical model of this is a *sigma-ideal*; a collection of subsets of a space deemed 'negligible'. As in ring theory, an ideal is a set of objects identified with zero, by which we can 'mod out'.

Definition 181 Sigma-ideal

Let $\mathbf{X}$ be a set. A **sigma ideal** *over $\mathbf{X}$ is a collection $\mathcal{Z}$ of subsets of $\mathbf{X}$ with the following properties:*

1. *$\mathcal{Z}$ is closed under* **countable unions.** *In other words, if $\mathbf{U}_1, \mathbf{U}_2, \ldots$, are in $\mathcal{Z}$, then their union $\displaystyle\bigcup_{n=1}^{\infty} \mathbf{U}_n$ is also in $\mathcal{Z}$.*
2. *If $\mathbf{U}$ is in $\mathcal{Z}$, and $\mathbf{A}$ is any subset of $\mathbf{X}$, then $\mathbf{A} \cap \mathbf{U}$ is also in $\mathcal{Z}$.*

Example 182:

(a) If $\mathbf{X}$ is uncountable, then the collection of all *finite and countable* subsets is a sigma-ideal.

(b) If $(\mathbf{X}, \mathcal{X}, \mu)$ is a measure space, then the collection of all sets of *measure zero* is a sigma-ideal.

(c) If $\mathbf{X}$ is a topological space, then the collection of all *meager* sets is a sigma-ideal (**Exercise 167** Recall: a subset $\mathbf{N} \subset \mathbf{X}$ is **nowhere dense** if $\text{int}\left(\text{cl}(\mathbf{N})\right) = \emptyset$, and $\mathbf{M} \subset \mathbf{X}$ is **meager** if $\mathbf{M} = \bigcup_{j=1}^{\infty} \mathbf{N}_j$ where $\mathbf{N}_j$ are nowhere dense)

(d) Let $\mathfrak{M}$ be a collection of measures on $\mathbf{X}$, and define

$$\mathcal{Z} = \{\mathbf{U} \in \mathcal{X} \ ; \ \mu(\mathbf{U}) = 0 \ \text{ for some } \mu \in \mathfrak{M}\}.$$

Then $\mathcal{Z}$ is a sigma-ideal (**Exercise 168**).

Remark: Example (182b) is prototypical. If $\mathcal{Z} \subset \mathcal{X}$ is any sigma-ideal, then there is a measure μ so that $\mathcal{Z}$ is the ideal of sets of μ-measure zero. To see this, let $\mathcal{Y} = \{\mathbf{X} \setminus \mathbf{Z} \ ; \ \mathbf{Z} \in \mathcal{Z}\}$. Then $\mathcal{W} = \mathcal{Z} \sqcup \mathcal{Y}$ is a sigma-algebra (**Exercise 169**). Define the measure μ as follows: for any $\mathbf{W} = \mathbf{Z} \sqcup \mathbf{Y}$ in $\mathcal{W}$,

$$\mu[\mathbf{W}] = \begin{cases} 1 & \text{if } \ \mathbf{Y} \neq \emptyset \\ 0 & \text{if } \ \mathbf{Y} = \emptyset \end{cases}$$

Then μ is a measure (**Exercise 170**), and $\mathcal{Z}$ is the sigma-ideal of sets of μ-measure zero.

6.2 Measure Algebras

Prerequisites: §1.1,§1.3,§6.1

A measure algebra is an abstraction of a measure space; it is what remains if you begin with a measure space $(\mathbf{X}, \mathcal{X}, \mu)$, and 'remove' the base space $\mathbf{X}$. To make sense of this, we need a way to get rid of the *points* in a space, but still leave the structure of *subsets* behind....

6.2(a) Algebraic Structure

A *sigma boolean algebra* is a set, $\mathcal{X}$, (whose elements can be imagined as the subsets of some 'imaginary' space,) equipped with an algebraic structure that mimics the set-theoretic operations of intersection, union, and complementation....

Definition 183 Sigma Boolean Algebra

A **sigma boolean algebra** *is a set $\mathcal{X}$ equipped with three operators: $\bigvee$ ('***join***'), $\bigwedge$ ('***meet***'), and $\neg$ ('***complementation***'), and two distinguished elements: an* **identity element**, $\mathbf{1}$, *and a* **null element**, $\mathbf{0}$.

$\bigvee$ and $\bigwedge$ are both defined for countable collections of elements; in other words, for any collection $\mathbf{X}_1, \mathbf{X}_2, \ldots$ in $\mathcal{X}$, we can define $\bigwedge_{n=1}^{\infty} \mathbf{X}_n$ and $\bigvee_{n=1}^{\infty} \mathbf{X}_n$. These operators satisfy the following axioms:

1. **Identity:** $\mathbf{1} \vee \mathbf{X} = \mathbf{1}$, $\quad \mathbf{0} \wedge \mathbf{X} = \mathbf{0}$, $\quad \mathbf{0} \vee \mathbf{X} = \mathbf{X}$, *and* $\mathbf{1} \wedge \mathbf{X} = \mathbf{X}$.
2. **Associativity:**
   (a) $\left(\bigwedge_{n=1}^{\infty} \mathbf{X}_{1;n}\right) \vee \left(\bigwedge_{n=1}^{\infty} \mathbf{X}_{2;n}\right) = \bigwedge_{n,m=1}^{\infty} (\mathbf{X}_{1;n} \vee \mathbf{X}_{2;m})$.
   (b) $\left(\bigvee_{n=1}^{\infty} \mathbf{X}_{1;n}\right) \wedge \left(\bigvee_{n=1}^{\infty} \mathbf{X}_{2;n}\right) = \bigvee_{n,m=1}^{\infty} (\mathbf{X}_{1;n} \wedge \mathbf{X}_{2;m})$.
3. **de Morgan's Laws:**
   (a) $\neg\left(\bigwedge_{n=1}^{\infty} \mathbf{X}_n\right) = \bigvee_{n=1}^{\infty} \neg\mathbf{X}_n$.
   (b) $\neg\left(\bigvee_{n=1}^{\infty} \mathbf{X}_n\right) = \bigwedge_{n=1}^{\infty} \neg\mathbf{X}_n$.
4. **Cancellation:** $\mathbf{X} \vee \neg\mathbf{X} = \mathbf{1}$ *and* $\mathbf{X} \wedge \neg\mathbf{X} = \mathbf{0}$.

Example 184:

(a) For any set $\mathbf{X}$, the *power set* of $\mathbf{X}$ is a sigma-boolean algebra. In this case,

$$\mathbf{A} \vee \mathbf{B} = \mathbf{A} \cup \mathbf{B}; \qquad \mathbf{A} \wedge \mathbf{B} = \mathbf{A} \cap \mathbf{B}; \qquad \neg\mathbf{A} = \mathbf{A}^{\complement}; \qquad \mathbf{1} = \mathbf{X}, \quad \text{and} \quad \mathbf{0} = \emptyset.$$

(b) Any *sigma-algebra* is a sigma-boolean algebra.

Example 185: Sigma algebra, mod zero

Let $(\mathbf{X}, \mathcal{X})$ be a measurable space, and let $\mathcal{Z} \subset \mathcal{X}$ be a sigma-ideal. We define a new sigma-Boolean algebra, $\mathcal{X}/\mathcal{Z}$, as follows: First. define an equivalence relation $\sim$ on $\mathcal{X}$, so that, for all $\mathbf{A}, \mathbf{B} \in \mathcal{X}$,

$$\Big(\mathbf{A} \sim \mathbf{B} \Big) \iff \Big(\mathbf{A} \triangle \mathbf{B} = \mathbf{Z}, \text{ for some } \mathbf{Z} \in \mathcal{Z} \Big)$$

Thus, $\mathbf{A}$ and $\mathbf{B}$ are "equivalent" if they differ only by a null set. Let $\widetilde{\mathbf{A}} \in \mathcal{X}/\mathcal{Z}$ denote the equivalence class of $\mathbf{A} \in \mathcal{X}$, and define:

$$\mathbf{1} \; = \; \widetilde{\mathbf{X}}; \qquad \mathbf{0} \; = \; \widetilde{\emptyset}; \qquad \neg\widetilde{\mathbf{A}} \; := \; \widetilde{\mathbf{X} - \mathbf{A}}$$

$$\bigwedge_{n=1}^{\infty} \widetilde{\mathbf{A}}_n \; := \; \widetilde{\bigcap_{n=1}^{\infty} \mathbf{A}_n}, \qquad \text{and} \qquad \bigvee_{n=1}^{\infty} \widetilde{\mathbf{A}}_n \; := \; \widetilde{\bigcup_{n=1}^{\infty} \mathbf{A}_n}$$

<u>**Exercise 171**</u> Show that these operations are well-defined, and that $\mathcal{X}/\mathcal{Z}$ is a sigma-Boolean algebra.

Example 186:

(a) Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a probability space, and $\mathcal{Z}$ the sigma-ideal of sets of measure zero. Then elements of $\mathcal{X}/\mathcal{Z}$ are equivalence classes of sets which differ by a subset of measure zero. The element $\mathbf{0}$ is the class of all sets of *measure zero*, and the element $\mathbf{1}$ is the class of all sets of *measure one*.

(b) Let $\mathbf{X}$ be a topological space with Borel sigma-algebra $\mathcal{B}$, and let $\mathcal{M}$ be the sigma-ideal of meager sets. The elements of $\mathcal{B}/\mathcal{M}$ are equivalence classes of sets which differ only by a meager subset. The element $\mathbf{0}$ is the class of all meager subsets, and the element $\mathbf{1}$ is class of all **comeager** (or **residual**) sets.

Definition 187 Measure algebra

A measure algebra is a pair $(\mathcal{X}, \mu)$, where $\mathcal{X}$ is a **sigma Boolean algebra**, *and $\mu : \mathcal{X} \longrightarrow [0, \infty]$ is a* **countably additive** *function: if $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are* **disjoint** *elements in $\mathcal{X}$ (ie. $\mathbf{X}_k \wedge \mathbf{X}_j = \mathbf{0}$, for all $k \neq j$), then* $\mu\left[\bigvee_{n=1}^{\infty} \mathbf{X}_n\right] = \sum_{n=1}^{\infty} \mu[\mathbf{X}_n]$.

Example 188:

Let $(\mathbf{X}, \mathcal{X}, \mu)$ be a measure space, and let $\mathcal{Z} \subset \mathcal{X}$ be the ideal of sets of measure zero. Let $\widetilde{\mathcal{X}} = \mathcal{X}/\mathcal{Z}$, and define $\widetilde{\mu} : \mathcal{Y} \longrightarrow [0, \infty]$ by: $\widetilde{\mu}\left[\widetilde{\mathbf{A}}\right] = \mu[\mathbf{A}]$, where $\widetilde{\mathbf{A}} \in \widetilde{\mathcal{X}}$ denotes the equivalence class of $\mathbf{A} \in \mathcal{X}$.

Exercise 172 Show that $\widetilde{\mu}$ is well-defined and countably additive

When speaking of the measure algebra associated with a particular measure space, it is common to tacitly mod out by sets of measure zero; hence "$(\mathcal{X}, \mu)$" often denotes the measure algebra $(\widetilde{\mathcal{X}}, \widetilde{\mu})$.

Suppose $(\mathbf{X}, \mathcal{X}, \mu)$ and $(\mathbf{Y}, \mathcal{Y}, \nu)$ are two measure spaces, whose structures we compare, say, to establish an isomorphism between them. It may be difficult to establish a precise *point-by-point* correspondence between $\mathbf{X}$ and $\mathbf{Y}$. It is often easier to establish a correspondence between elements of $\mathcal{X}$ and $\mathcal{Y}$.

Definition 189 Morphism of Measure Algebras

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two measure algebras. A **morphism** *of measure-algebras is a map $f : \mathcal{X} \longrightarrow \mathcal{Y}$, so that:*

1. *The sigma-boolean operations are preserved: for any $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \ldots \in \mathcal{X}$,*

$$\left(\bigvee_{k=1}^{\infty} \mathbf{U}_k\right) = \bigvee_{k=1}^{\infty} F(\mathbf{U}_k); \qquad F\left(\bigwedge_{k=1}^{\infty} \mathbf{U}_k\right) = \bigwedge_{k=1}^{\infty} F(\mathbf{U}_k); \qquad \textit{and} \quad F(\neg\mathbf{U}) = \neg F(\mathbf{U}).$$

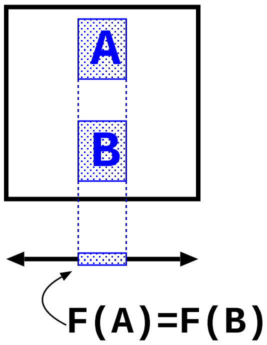


Figure 6.1: $F(\mathbf{A}) = F(\mathbf{B})$; thus, $F(\mathbf{A}) \cap F(\mathbf{B}) \neq \emptyset = F(\mathbf{A} \cap \mathbf{B})$.

2. *The measure is preserved: for any* $\mathbf{U} \in \mathcal{X}$, $\mu_1[\mathbf{U}] = \mu_2[F(\mathbf{U})]$.

Example 190:

Suppose $(\mathbf{X}_1, \mathcal{X}_1, \mu_1)$ and $(\mathbf{X}_2, \mathcal{X}_2, \mu_2)$ are measure spaces, and suppose $f : \mathbf{X}_1 \longrightarrow \mathbf{X}_2$ is a measurable map, mod zero. Then the map

$$f^{-1} : \mathcal{X}_2 \ni \mathbf{U} \mapsto f^{-1}[\mathbf{U}] \in \mathcal{X}_1$$

is a morphism of *sigma-boolean* algebras.

If f is also *measure-preserving*, then f^{-1} is a morphism of *measure algebras* (**Exercise 173**).

Note: In general, the map

$$F : \mathcal{X}_1 \ni \mathbf{U} \mapsto F(\mathbf{U}) \in \mathcal{X}_2$$

is *not* a measure-algebra homomorphism. For example, if f is not injective, then F fails to preserve the boolean operations.

For example, let $F = \mathbf{pr} : \mathbb{I}^2 \longrightarrow \mathbb{I}$ be the projection from the unit square to the unit interval. Figure 6.1 shows two disjoint subsets $\mathbf{A}, \mathbf{B} \subset \mathbb{I}^2$, such that $F(\mathbf{A}) = F(\mathbf{B})$. Thus, $F(\mathbf{A}) \cap F(\mathbf{B}) \neq \emptyset = F(\mathbf{A} \cap \mathbf{B})$.

Theorem 191: *Suppose* $(\mathbf{X}, \mathcal{X}, \mu)$ *and* $(\mathbf{Y}, \mathcal{Y}, \nu)$ *are measure spaces, and* $f : \mathbf{X} \longrightarrow \mathbf{Y}$ *is a measure-preserving map, mod zero.*

1. *The map* $f^{-1} : \mathcal{Y} \longrightarrow \mathcal{X}$ *is always injective.*

2. f^{-1} *is surjective if and only if* f *is almost-injective.*

Proof: **Exercise 174** □

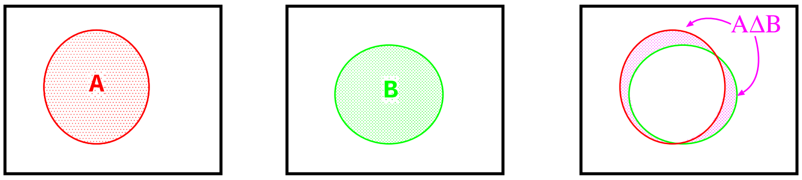


Figure 6.2: $d_\triangle(\mathbf{A},\mathbf{B})$ measures the symmetric difference between $\mathbf{A}$ and $\mathbf{B}$.

6.2(b) Metric Structure

A measure algebra $(\mathcal{X},\mu)$ has a natural metric structure. For any $\mathbf{A},\mathbf{B}\in\mathcal{X}$, we define

$$d_\triangle(\mathbf{A},\mathbf{B}) \quad := \quad \mu\left[\mathbf{A}\triangle\mathbf{B}\right]. \qquad \text{(see Figure 6.2)}$$

Proposition 192: *Let $(\mathcal{X},\mu)$ be a measure algebra, with metric $d_\triangle$,*

1. *$(\mathcal{X},d_\triangle)$ is a complete metric space.*
2. *If μ is a finite measure, then $(\mathcal{X},d_\triangle)$ is bounded.*

Proof: <u>**Exercise 175**</u> Hint: Let $\mathbf{L}^1_1(\mathbf{X},\mathcal{X},\mu)$ denote the class of all functions in $\mathbf{L}^1$ which take only the values zero or one. Define $\Phi:(\mathcal{X},\mu)\ni\mathbf{A}\mapsto\mathbb{1}_\mathbf{A}\in\mathbf{L}^1_1(\mathbf{X},\mathcal{X},\mu)$. Show that:

(1) Φ is a bijection.

(2) Φ is an isometry between the metric $d_\triangle$ and the $\mathbf{L}^1$ -norm. ______ □

The algebraic operations of a measure algebra are continuous with respect to this metric:

Lemma 193: *Let $(\mathcal{X},\mu)$ be a measure algebra, with associated metric $d_\triangle$. Then:*

1. *The map $\Phi:(\mathcal{X},\mu)\times(\mathcal{X},\mu)\ni(\mathbf{A},\mathbf{B})\mapsto\mathbf{A}\vee\mathbf{B}\in(\mathcal{X},\mu)$ is continuous with respect to $d_\triangle$. Furthermore, Φ is a* **contraction:** *for any $\mathbf{A}_1,\mathbf{A}_2,\mathbf{B}_1,\mathbf{B}_2\in\mathcal{X}$, we have:*

$$d_\triangle\left[\mathbf{A}_1\vee\mathbf{B}_1\,,\;\mathbf{A}_2\vee\mathbf{B}_2\right] \quad\le\quad d_\triangle\left[\mathbf{A}_1,\mathbf{A}_2\right]+d_\triangle\left[\mathbf{B}_1,\mathbf{B}_2\right].$$

2. *Similarly, the map $\Psi:(\mathcal{X},\mu)\times(\mathcal{X},\mu)\ni(\mathbf{A},\mathbf{B})\mapsto\mathbf{A}\wedge\mathbf{B}\in(\mathcal{X},\mu)$ is continuous and a contraction with respect to $d_\triangle$. For any $\mathbf{A}_1,\mathbf{A}_2,\mathbf{B}_1,\mathbf{B}_2\in\mathcal{X}$, we have:*

$$d_\triangle\left[\mathbf{A}_1\wedge\mathbf{B}_1\,,\;\mathbf{A}_2\wedge\mathbf{B}_2\right] \quad\le\quad d_\triangle\left[\mathbf{A}_1,\mathbf{A}_2\right]+d_\triangle\left[\mathbf{B}_1,\mathbf{B}_2\right].$$

3. *If μ is* **finite***, then the map $(\mathcal{X}\ni\mathbf{A}\mapsto\neg\mathbf{A}\in\mathcal{X})$ is also continuous with respect to $d_\triangle$.*

Proof: <u>**Exercise 176**</u> ______ □

Theorem 194: *Let* $(\mathbf{X}, \mathcal{X}, \mu)$ *and* $(\mathbf{Y}, \mathcal{Y}, \nu)$ *be measure spaces, and let* $f : \mathbf{X} \longrightarrow \mathbf{Y}$ *be measurable. Consider the induced map* $f^{-1} : \mathcal{Y} \longrightarrow \mathcal{X}$.

1. $\Big($ f^{-1} *is continuous* $\Big)$ $\iff$ $\Big($ $f^*(\mu)$ is absolutely continuous *with respect to* ν. $\Big)$

2. $\Big($ f^{-1} *is an isometry* $\Big)$ $\iff$ $\Big($ f *is* measure-preserving. $\Big)$

   In this case, f^{-1} *isometrically embeds* $\mathcal{Y}$ *as a closed subspace of* $\mathcal{X}$.

Proof: **Exercise 177** □