A Crash Course on the Lebesgue Integral and Measure Theory

Steve Cheng

April 27, 2008

Contents

Preface iv
0.1 Philosophy . . . iv
0.2 Special thanks . . . v

Organization of this book vi

1 Motivation for Lebesgue integral 1

2 Basic definitions 5
2.1 Measurable spaces . . . 5
2.2 Positive measures . . . 7
2.3 Measurable functions . . . 11
2.4 Definition of the integral . . . 15
2.5 Exercises . . . 21

3 Useful results on integration 25
3.1 Convergence theorems . . . 25
3.2 Interchange of summation and integral . . . 29
3.3 Mass density functions . . . 30
3.4 Change of variables . . . 32
3.5 Integrals with parameter . . . 34
3.6 Exercises . . . 36

4 Some commonly encountered spaces 38
4.1 Space of integrable functions . . . 38
4.2 Probability spaces . . . 42
4.3 Exercises . . . 42

5 Construction of measures 43
5.1 Outer measures . . . 43
5.2 Defining a measure by extension . . . 46
5.3 Probability measure for infinite coin tosses . . . 48
5.4 Lebesgue measure on $\mathbb{R}^n$ . . . 51
5.5 Existence of non-measurable sets . . . 54
5.6 Completeness of measures . . . 54

5.7 Exercises 56

6 Integration on product spaces **59**
6.1 Product measurable spaces 59
6.2 Cross-sectional areas 61
6.3 Iterated integrals 63
6.4 Exercises 65

7 Integration over $\mathbb{R}^n$ **68**
7.1 Riemann integrability implies Lebesgue integrability 68
7.2 Change of variables in $\mathbb{R}^n$ 70
7.3 Integration on manifolds 72
7.4 Stieltjes integrals 74
7.5 Exercises 77

8 Decomposition of measures **84**
8.1 Signed measures 84
8.2 Radon-Nikodým decomposition 88
8.3 Exercises 91

9 Approximation of Borel sets and functions **93**
9.1 Regularity of Borel measures 93
9.2 $\mathbf{C}_0^\infty$ functions are dense in $\mathbf{L}^p(\mathbb{R}^n)$ 95

10 Relation of integral with derivative **98**
10.1 Differentiation with respect to volume 98
10.2 Fundamental theorem of calculus 100

11 Spaces of Borel measures **101**
11.1 Integration as a linear functional 101
11.2 Riesz representation theorem 101
11.3 Fourier analysis of measures 101
11.4 Applications to probability theory 101
11.5 Exercises 101

12 Miscellaneous topics **102**
12.1 Dual space of $\mathbf{L}^p$ 102
12.2 Egorov's Theorem 102
12.3 Integrals taking values in Banach spaces 103

A Results from other areas of mathematics **107**
A.1 Extended real number system 107
A.2 Linear algebra 108
A.3 Multivariate calculus 108
A.4 Point-set topology 109
A.5 Functional analysis 109

A.6 Fourier analysis 109

B Bibliography **110**

C List of theorems **113**

D List of figures **115**

E Copying and distribution rights to this document **117**

Preface

WORK IN PROGRESS

This booklet is an exposition on the Lebesgue integral. I originally started it as a set of notes consolidating what I had learned on on Lebesgue integration theory, and published them in case somebody else may find them useful.

I welcome any comments or inquiries on this document. You can reach me by e-mail at ⟨steve@gold-saucer.org⟩.

0.1 Philosophy

Since there are already countless books on measure theory and integration written by professional mathematicians, that teach the same things on the basic level, you may be wondering why you should be reading this particular one, considering that it is so blatantly informally written.

Ease of reading. Actually, I believe the informality to be quite appropriate, and integral — pun intended — to this work. For me, this booklet is also an experiment to write an engaging, easy-to-digest mathematical work that people would want to read in my spare time. I remember, once in my second year of university, after my professor off-handed mentioned Lebesgue integration as a "nicer theory" than the Riemann integration we had been learning, I dashed off to the library eager to learn more. The books I had found there, however, were all fixated on the stiff, abstract theory — which, unsurprisingly, was impenetrable for a wide-eyed second-year student flipping through books in his spare time. I still wonder if other budding mathematics students experience the same disappointment that I did. If so, I would like this book to be a partial remedy.

Motivational. I also find that the presentation in many of the mathematics books I encounter could be better, or at least, they are not to my taste. Many are written with hardly any motivating examples or applications. For example, it is evident that the concepts introduced in linear functional analysis have something to do with problems arising in mathematical physics, but "pure" mathematical works on the subject too often tend to hide these origins and applications. Perhaps, they may be obvious to the learned reader, but not always for the student who is only starting his exploration of the diverse areas of mathematics.

Rigor. On the other hand, in this work I do not want to go to the other extreme, which is the tendency for some applied mathematics books to be unabashedly unrigorous. Or worse, pure deception: they present arguments with unstated assumptions, and impress on students, by naked authority, that everything they present is perfectly correct. Needless to say, I do not hold such books in very high regard, and probably you may not either. This work will be rigorous and precise, and of course, you can always confirm with the references.

0.2 Special thanks

Organization of this book

The material presented in this book is selected and organized differently than many other mathematics books.

Level of abstraction. The emphasis is on a healthy dose of abstraction which is backed by useful applications. Not so abstract as to get us bogged down by details, but abstract enough so that results can be stated elegantly, concisely and effectively.

Style of exposition. We favor a style of writing, for both the main text and the proofs, that is more wordy than terse and formal. Thus some material may appear to take up a lot of pages, but is actually quick to read through.

Order of presentation. To make the material easier to digest, the theory is sometimes not developed in a strict logical order. Some results will be stated and used, before going back and formally proving them in later sections. It seems to work well, I think.

Order for reading. I almost never read any non-fiction books linearly from front to back, and mathematics books are no exception. Thus I have tried writing to facilitate skipping around sections and reading "on the go".

Exercises. At the end of each chapter, there are exercises that I have culled from standard results and my own experience. As I am a practitioner, not a researcher, I am afraid I will not be able to provide many of the pathological exercises found in the more hard-core books. Nevertheless, I hope my exercises turn out to be interesting.

Prerequisites. The minimal prerequisites for understanding this work is a thorough understanding of elementary calculus, though at various points of the text we will definitely need the concepts and results from topology, linear algebra, multivariate calculus, and functional analysis.

If you have not encountered these topics before, then you probably should endeavour to learn them. Understandably, that will take you time; so this text contains an appendix listing the results that we will need — it may serve as a guide and motivation towards further study.

Level of coverage. My aim is to provide a reasonably complete treatment of the material, yet leave various topics ready for individual exploration.

Perhaps you may be unsatisfied with my approach. Actually, even if I could just arouse your interest in measure theory or any other mathematical topic through this work, then I think I have done something useful.

Chapter 1

Motivation for Lebesgue integral

The Lebesgue integral, introduced by Henri Lebesgue in his 1902 dissertation, “Intégrale, longueur, aire”, is a generalization of the Riemann integral usually studied in elementary calculus.

If you have followed the rigorous definition of the Riemann integral in $\mathbb{R}$ or $\mathbb{R}^n$, you may be wondering why do we need to study yet another integral. After all, why should we even care to integrate nasty things like the Dirichlet function:

$$D(x) = \begin{cases} 1\,, & x \in \mathbb{Q} \\ 0\,, & x \in \mathbb{R} \setminus \mathbb{Q}\,. \end{cases}$$

There are several convincing arguments for studying Lebesgue integration theory.

Closure of operations. Recall the first times you had ever heard of the concepts of negative numbers, irrational numbers, and complex numbers. At the time, you most likely had thought, privately if not announced publically, that there is no such use for “numbers less than nothing”, “numbers with an infinite amount of decimal places” (that we approximate in calculation with a finite amount anyway), and “imaginary numbers” whose square is negative. Yet negative numbers, irrational numbers, and complex numbers find their use because these kinds of numbers are *closed* under certain operations — namely, the operations of subtraction, taking least upper bounds, and polynomial roots. Thus equations involving more “ordinary” quantities can be posed and solved, without putting the special cases or artificial restrictions that would be necessary had the number system not been extended appropriately.

The Lebesgue integral has the similar relation to the Riemann integral. For instance, the D function above can be considered to be the infinite sum

$$D(x) = \sum_{n=1}^{\infty} E_n(x)\,, \quad E_n(x) = \begin{cases} 1\,, & x = x_n \\ 0\,, & x \neq x_n\,, \end{cases}$$

where $\{x_n\}_{n=1}^{\infty}$ is any listing of the members of $\mathbb{Q}$. The function D is not Riemann-integrable, yet each function E_n is. So, in general, the limit of Riemann-integrable functions is not necessarily Riemann-integrable; the class of Riemann-integrable functions is not closed under taking limits.

This failure of the closure causes all sorts of problems, not the least of which is deciding when exactly the following interchange of limiting operations, is valid:

$$\lim_{n\to\infty} \int f_n(x)\, dx = \int \lim_{n\to\infty} f_n(x)\, dx\,.$$

In elementary calculus, the interchange of the limit and integral, over a closed and bounded interval $[a, b]$, is proven to be valid when the sequence of functions $\{f_n\}$ is uniformly convergent. However, in many calculations that require exchanging the limit and integral, uniform convergence is often too onerous a condition to require, or we are integrating on unbounded intervals. In this case, the calculus theorem is not of much help — but, in fact, one of the important, and much applied, theorems by Henri Lebesgue, called the *dominated convergence theorem*, gives practical conditions for which the interchange is valid.

It is true that, if a function is Riemann-integrable, then it is Lebesgue-integrable, and so theorems about the Lebesgue integral could in principle be rephrased as results for the Riemann integral, with some restrictions on the functions to be integrated. Yet judging from the fact that calculus books almost never do actually attempt to prove the dominated convergence theorem, and that the theorem was originally discovered through measure theory methods, it seems fair to say that a proof using elementary Riemann-integral methods may be close to impossible.

Abstraction is more efficient. There is also an argument for preferring the Lebesgue integral because it is more abstract. As with many abstractions in mathematics, there is an up-front investment cost to be paid, but the pay-off in effectiveness is enormous. Cumbersome manipulations of Riemann integrals can be replaced by concise arguments involving Lebesgue integrals.

The *dominated convergence theorem* mentioned above is one example of the power of Lebesgue integrals; here we illustrate another one. Consider evaluating the Gaussian probability integral:

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx\,.$$

The derivation usually goes like this:

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx\right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\, dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy \\
&= \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy \\
&= \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\, d\theta \qquad \text{(using polar coordinates)} \\
&= 2\pi \left[-e^{-\frac{1}{2}r^2}\right]_{r=0}^{r=\infty} \\
&= 2\pi\,.
\end{aligned}$$

So
$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}\,.$$

The above computation looks to be easy, and although it can be justified using the Riemann integral alone, the explanation is quite clumsy. Here are some questions to ask: Why should $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}$ be the same as $\int_{\mathbb{R}^2}$? Is the polar coordinate transformation valid over the unbounded domain $\mathbb{R}^2$? Note that while most calculus books do systematically develop the theory of the Riemann integral over *bounded* sets of integration, the Riemann integral over *unbounded* sets is usually treated in an ad-hoc manner, and it is not always clear when the theorems proven for bounded sets of integration apply to unbounded sets also.

On the other hand, the Lebesgue integral makes no distinction between bounded and unbounded sets in integration, and the full power of the standard theorems apply equally to both cases. As you will see, the proper abstract development of the Lebesgue integral also *simplifies* the *proofs* of the theorems regarding interchanging the order of integration and differential coordinate transformation.

New applications. Finally, with abstraction comes new areas to which the integral can be applied. There are many of these, but here we will briefly touch on a few.

The Lebesgue theory is not restricted to integrating functions over lengths, areas or volumes of $\mathbb{R}^n$. We can specify beforehand a *measure* μ on some space X, and Lebesgue theory provides the tools to define, and possibly compute, the integral
$$\int_X f\,d\mu\,,$$
that represents, loosely, a limit of integral sums
$$\sum_i f(x_i)\,\mu(A_i)\,, \qquad \{A_i\} \text{ is a partition of } X\text{, and } x_i \text{ is a point in } A_i.$$
When integrating over $\mathbb{R}^n$ with the *Lebesgue measure*, the quantity $\mu(A_i)$ is simply the area or volume of the set A_i, but in general, other measure functions $A \mapsto \mu(A)$ could be used as well.

By choosing different measures μ, we can define the Lebesgue integral over curves and surfaces in $\mathbb{R}^n$, thereby uniting the disparate definitions of the line, area, and volume integrals in multivariate calculus.

In functional analysis, Lebesgue integrals are used to represent certain linear functionals. For example, the continuous dual space of $C[a,b]$, the space of continuous functions on the interval $[a,b]$ is in one-to-one correspondence with a certain family of measures on the space $[a,b]$. This fact was actually proven in 1909 by Frigyes Riesz, using the Riemann-Stieltjes integral, a generalization of the Riemann integral — but the result is more elegantly interpreted using the Lebesgue theory.

And finally, the Lebesgue integral is indispensable in the rigorous study of probability theory. A probability measure behaves analogously to an area measure, and in fact a probability measure *is* a measure in the Lebesgue sense.

Having given a small survey of the topic of Lebesgue integration, we now proceed to formulate the fundamental definitions.

Chapter 2

Basic definitions

This chapter describes the Lebesgue integral and the basic machinery associated with it.

2.1 Measurable spaces

Before we can define the protagonist of our story, the Lebesgue integral, we will need to describe its setting, that of measurable spaces and measures.

We are given some function μ of sets which returns the area or volume — formally called the *measure* — of the given set. That is, we assume at the beginning that such a function μ has already been defined for us. This axiomatic approach has the obvious advantage that the theory can be applied to many other measures besides volume in $\mathbb{R}^n$. We have the benefit of hindsight, of course. Lebesgue himself did not work so abstractly, and it was Maurice Fréchet who first pointed out the generalizations of the concrete methods of Lebesgue and his contemporaries that are now standard.

A general domain that can be taken for the function μ is a σ-algebra, defined below.

2.1.1 Definition (Measurable sets)
Let X be any non-empty set. A σ**-algebra**[†] of subsets of X is a family $\mathcal{M}$ of subsets of X, with the properties:

① $\mathcal{M}$ is non-empty.

② *Closure under complement*: If $E \in \mathcal{M}$, then $X \setminus E \in \mathcal{M}$.

③ *Closure under countable union*: If $\{E_n\}_{n=1}^{\infty}$ is a sequence of sets in $\mathcal{M}$, then their union $\bigcup_{n=1}^{\infty} E_n$ is in $\mathcal{M}$.

[†]Read as "sigma algebra". The Greek letter σ in this ridiculous name stands for the German word *summe*, meaning union. In this context it specifically refers to the *countable* union of sets, in contrast to mere finite union.

The pair $(X, \mathcal{M})$ is called a **measurable space**, and the sets in $\mathcal{M}$ are called the **measurable sets**.

2.1.2 REMARK (CLOSURE UNDER COUNTABLE INTERSECTIONS). The axioms always imply that $X, \emptyset \in \mathcal{M}$. Also, by De Morgan's laws, $\mathcal{M}$ is closed under countable intersections as well as countable union.

2.1.3 EXAMPLE. Let X be any non-empty set. Then $\mathcal{M} = \{X, \emptyset\}$ is the trivial σ-algebra.

2.1.4 EXAMPLE. Let X be any non-empty set. Then its power set $\mathcal{M} = 2^X$ is a σ-algebra.

Needless to say, we cannot insist that $\mathcal{M}$ is closed under arbitrary unions or intersections, as that would force $\mathcal{M} = 2^X$ if $\mathcal{M}$ contains all the singleton sets. That would be uninteresting. More importantly, we do not always want $\mathcal{M} = 2^X$ because we may be unable to come up with sensible definitions of "area" for some very wild sets $A \in 2^X$.

On the other hand, we want closure under countable set operations, rather than just finite ones, as we will want to be taking countable limits. Indeed, you may recall that the class of *Jordan-measurable sets* (those that have an area definable by a Riemann integral) is *not* closed under countable union, and that causes all sorts of trouble when trying to prove limiting theorems for Riemann integrals.

To get non-trivial σ-algebras to work with we need the following very unconstructive(!) construction.

If we have a family of σ-algebras on X, then the intersection of all the σ-algebras from this family is also a σ-algebra on X. If all of the σ-algebras from the family contains some fixed $\mathcal{G} \subseteq 2^X$, then the intersection of all the σ-algebras from the family, of course, is a σ-algebra containing $\mathcal{G}$.

Now if we are given $\mathcal{G}$, and we take *all* the σ-algebras on X that contain $\mathcal{G}$, and intersect all of them, we get the smallest σ-algebra that contains $\mathcal{G}$.

2.1.5 DEFINITION (GENERATED σ-ALGEBRA)
For any given family of sets $\mathcal{G} \subseteq 2^X$, the **$\sigma$-algebra generated by** $\mathcal{G}$ is the the smallest σ-algebra containing $\mathcal{G}$; it is denoted by $\sigma(\mathcal{G})$.

The following is an often-used σ-algebra.

2.1.6 DEFINITION (BOREL σ-ALGEBRA)
If X is a space equipped with topology $\mathscr{T}(X)$, the **Borel σ-algebra** is the σ-algebra $\mathscr{B}(X) = \sigma(\mathscr{T}(X))$ generated by all open subsets of X.

When *topological spaces are involved, we will always take the σ-algebra to be the Borel σ-algebra* unless stated otherwise.

$\mathscr{B}(X)$, being generated by the open sets, then contains all open sets, all closed sets, and countable unions and intersections of open sets and closed sets — and then countable unions and intersections of those, and so on, moving deeper and deeper into the hierarchy.

In general, there is no more explicit description of what sets are in $\mathscr{B}(X)$ or $\sigma(\mathcal{G})$, save for a formal construction using the principle of transfinite induction from set theory. Fortunately, in analysis it is rarely necessary to know the set-theoretic description of $\sigma(\mathcal{G})$: any sets that come up can be proven to be measurable by expressing them countable unions, intersections and complements of sets that we already know are measurable. So the language of σ-algebras can be simply viewed as a concise, abstract description of unendingly taking countable set operations.

The Borel σ-algebra is generally *not all* of 2^X — this fact is shown in Theorem 5.5.1, which you can read now if it interests you.

2.2 Positive measures

2.2.1 Definition (Positive measure)

Let $(X, \mathcal{M})$ be a measurable space. (Thus $\mathcal{M}$ is a σ-algebra as in Definition 2.1.1.) A **(positive) measure** on this space is a *non-negative* set function $\mu \colon \mathcal{M} \to [0, \infty]$ such that

① $\mu(\emptyset) = 0$.

② *Countable additivity*: For any sequence of *mutually disjoint* sets $E_n \in \mathcal{M}$,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)\,.$$

The set $(X, \mathcal{M}, \mu)$ will be called a **measure space**.

For convenience, we will often refer to measurable spaces and measure spaces only by naming the domain; the σ-algebra and measure will be implicit, as in: "let X be a measure space", etc.

2.2.2 Example (Counting measure). Let X be an arbitrary set, and $\mathcal{M}$ be a σ-algebra on X. Define $\mu \colon \mathcal{M} \to [0, \infty]$ as

$$\mu(A) = \begin{cases} \text{the number of elements in } A\,, & \text{if } A \text{ is a finite set}\,, \\ \infty\,, & \text{if } A \text{ is an infinite set}\,. \end{cases}$$

This is called the **counting measure**.

We will be able to model the infinite series $\sum_{n=1}^{\infty} a_n$ in Lebesgue integration theory by using $X = \mathbb{N}$ and the counting measure.

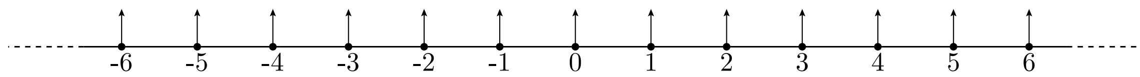

Figure 2.1: Schematic of the counting measure on the integers.

2.2.3 EXAMPLE (LEBESGUE MEASURE). $X = \mathbb{R}^n$, and $\mathcal{M} = \mathscr{B}(\mathbb{R}^n)$. There is the **Lebesgue measure** λ which assigns to the rectangle its usual n-dimensional volume:

$$\lambda\big([a_1, b_1] \times \cdots \times [a_n, b_n]\big) = (b_1 - a_1) \cdots (b_n - a_n)\,.$$

This measure λ should also assign the correct volumes to the usual geometric figures, as well as for all the other sets in $\mathcal{M}$. Intuitively, defining the volume of the rectangle only should suffice to uniquely determine the volume of the other sets, as the volume of any other measurable set can be approximated by the volume of many small rectangles. Indeed, we will later show this intuition to be true.

In fact, many theorems using Lebesgue measure and integrals with respect to Lebesgue measure really depend only on the definition of the volume of the rectangle. So for now we skip the fine details of properly showing the *existence* of Lebesgue measure, coming back to it later.

2.2.4 EXAMPLE (DIRAC MEASURE). Fix $x \in X$, and set $\mathcal{M} = 2^X$. The **Dirac measure** or **point mass** is the measure $\delta_x \colon \mathcal{M} \to [0, \infty]$ with

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

The Dirac measure, considered by itself, appears to be trivial, but it appears as a basic building block for other measures, as we will see later.

The name "Dirac" is attached to this measure as it is the realization in measure theory of the (in)famous "Dirac delta function". The layman's definition of the Dirac delta "function", δ_x, states that it should satisfy, for all continuous functions $f \colon \mathbb{R} \to \mathbb{R}$,

$$\int_{-\infty}^{\infty} f(y)\,\delta_x(y)\,dy = f(x)\,, \quad x \in \mathbb{R}\,.$$

Our *measure* δ_x will behave similarly: since the mass of the measure is concentrated at the point x, our eventual definition of the Lebesgue integral with respect to the measure δ_x will satisfy:

$$\int_{-\infty}^{\infty} f(y)\,d\delta_x(y) = f(x)\,, \quad x \in \mathbb{R}\,.$$

Before we begin the prove more theorems, we remark that we will be manipulating quantities from the *extended real number system* (see section A.1) that includes ∞. We should warn here that the additive cancellation rule *will not work* with ∞. The danger should be sufficiently illustrated in the proofs of the following theorems.

2.2.5 Theorem (Additivity and monotonicity of positive measures)
Let μ be a measure on a measurable space $(X, \mathcal{M})$. It has the following basic properties:

① It is *finitely additive*.

② *Monotonicity*: If $A, B \in \mathcal{M}$, and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

③ If $A, B \in \mathcal{M}$, with $A \subseteq B$ and $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

Proof Property ① is obvious. For property ②, applying property ① to the disjoint sets $B \setminus A$ and A, we have

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

and $\mu(B \setminus A) \geq 0$. For property ③, subtract $\mu(A)$ from both sides.

Note that $B \setminus A = B \cap (X \setminus A)$ belongs to $\mathcal{M}$ by the properties of a σ-algebra. ■

The preceding theorem, as well as the next ones, are quite intuitive and you should have no trouble remembering them.

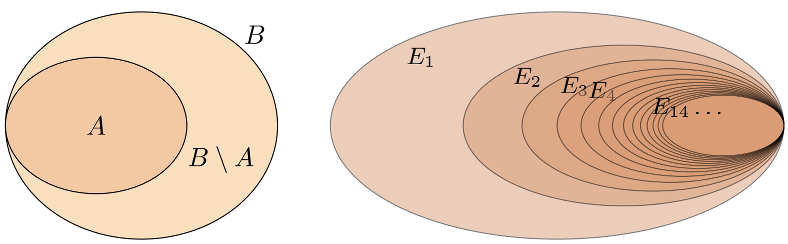

Figure 2.2: The sets and measures involved in Theorem 2.2.5 and Theorem 2.2.7.

2.2.6 Theorem (Countable subadditivity)
For any sequence of sets $E_1, E_2, \ldots \in \mathcal{M}$, not necessarily disjoint,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n)\,.$$

Proof The union $\bigcup_n E_n$ can be "disjointified" — that is, rewritten as a disjoint union — so that we can apply ordinary countable additivity:

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} E_n \setminus (E_1 \cup E_2 \cup \cdots \cup E_{n-1})\right) \\ &= \sum_{n=1}^{\infty} \mu\big(E_n \setminus (E_1 \cup E_2 \cup \cdots \cup E_{n-1})\big) \leq \sum_{n=1}^{\infty} \mu(E_n)\,.\end{aligned}$$

The final inequality comes from the monotonicity property of Theorem 2.2.5. ■

2.2.7 THEOREM (CONTINUITY FROM BELOW AND ABOVE)
Let $(X, \mathcal{M}, \mu)$ be a measure space.

① *Continuity from below*: Let $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ be subsets in $\mathcal{M}$ with union E. The sets E_n are said to **increase** to E, and henceforth we will abbreviate this by $E_n \nearrow E$. Then

$$\mu(E) = \mu\left(\bigcup_{n=0}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n)\,.$$

② *Continuity from above*: Let $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots$ be subsets in $\mathcal{M}$ with intersection E. The sets E_n are said to be **decrease** to E, and this will be abbreviated $E_n \searrow E$. *Assuming that at least one* E_k has $\mu(E_k) < \infty$, then

$$\mu(E) = \mu\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n)\,.$$

Proof Since the sets E_k are increasing, they can be written as the disjoint unions:

$$\begin{aligned} E_k &= E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots \cup (E_k \setminus E_{k-1})\,, \quad E_0 = \varnothing\,, \\ E &= E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots\,, \end{aligned}$$

so that

$$\mu(E) = \sum_{k=1}^{\infty} \mu(E_k \setminus E_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(E_k \setminus E_{k-1}) = \lim_{n\to\infty} \mu(E_n)\,.$$

This proves property ①.

For property ②, we may as well assume $\mu(E_1) < \infty$, for if $\mu(E_k) < \infty$, we can just discard the sets before E_k, and this affects nothing. Observe that $(E_1 \setminus E_n) \nearrow (E_1 \setminus E)$, and $\mu(E) \le \mu(E_n) \le \mu(E_1) < \infty$. so we may apply property ①:

$$\mu(E_1) - \mu(E) = \mu(E_1 \setminus E) = \lim_{n\to\infty} \mu(E_1 \setminus E_n) = \mu(E_1) - \lim_{n\to\infty} \mu(E_n)\,,$$

and cancel the term $\mu(E_1)$ on both sides. ■

Though the general notation employed in measure theory unifies the cases for both finite and infinite measures, it is hardly surprising, as with Theorem 2.2.7, that some results only apply for the finite case. We emphasize this with a definition:

2.2.8 DEFINITION (FINITE MEASURE)

☞ A measurable set E has (μ-)**finite measure** if $\mu(E) < \infty$.

☞ A measure space (X, μ) has **finite measure** if $\mu(X) < \infty$.

2.3 Measurable functions

To do integration theory, we of course need functions to integrate. Given that not all sets may be measurable, it should be expected that perhaps not all functions can be integrated either. The following definition identifies those functions that are candidates for integration.

2.3.1 DEFINITION (MEASURABLE FUNCTIONS)
Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces (Definition 2.1.1). A map $f\colon X \to Y$ is **measurable** if

☞ for all $B \in \mathcal{B}$, the set $f^{-1}(B) = \{f \in \mathcal{B}\}$ is in $\mathcal{A}$.

This definition is not difficult to motivate in at least purely formal terms. It is completely analogous to the definition of continuity between topological spaces. For measure theory, the relevant structures on the domain and range space, are given by σ-algebras in place of topologies of open sets.

2.3.2 EXAMPLE (CONSTANT FUNCTIONS). A constant map f is always measurable, for $f^{-1}(B)$ is either $\emptyset$ or X.

2.3.3 THEOREM (COMPOSITION OF MEASURABLE FUNCTIONS)
The composition of two measurable functions is measurable.

Proof This is immediate from the definition. ∎

Recall that σ-algebra are often defined by generating them from certain elementary sets (Definition 2.1.5). In the course of proving that a particular function $f\colon X \to Y$ is measurable, it seems plausible that it should be atomatically measurable as soon as we verify that the pre-images $f^{-1}(B)$ are measurable for only elementary sets B. In fact, since the generated σ-algebra is defined non-constructively, we can think of no other way of establishing the measurability of $f^{-1}(B)$ directly.

The next theorem, Theorem 2.3.4, confirms this fact. Its technique of proof appears magical at first, but really it is forced upon us from the unconstructive definition of the generated σ-algebra.

2.3.4 THEOREM (MEASURABILITY FROM GENERATED σ-ALGEBRA)
Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, and suppose $\mathcal{G}$ generates the σ-algebra $\mathcal{B}$. A function $f\colon X \to Y$ is measurable if and only if

☞ for every set V in the generator set $\mathcal{G}$, its pre-image $f^{-1}(V)$ is in $\mathcal{A}$.

Proof The "only if" part is just the definition of measurability.

For the "if" direction, define

$$\mathcal{H} = \{V \in \mathcal{B} \colon f^{-1}(V) \in \mathcal{A}\}.$$

It is easily verified that $\mathcal{H}$ is a σ-algebra, since the operation of taking the inverse image commutes with the set operations of union, intersection and complement.

By hypothesis, $\mathcal{G} \subseteq \mathcal{H}$. Therefore, $\sigma(\mathcal{G}) \subseteq \sigma(\mathcal{H})$. But $\mathcal{B} = \sigma(\mathcal{G})$ by the definition of $\mathcal{G}$, and $\mathcal{H} = \sigma(\mathcal{H})$ since $\mathcal{H}$ is a σ-algebra. Unraveling the notation, this means $f^{-1}(V) \in \mathcal{A}$ for every $V \in \mathcal{B}$. ∎

2.3.5 COROLLARY (CONTINUITY IMPLIES MEASURABILITY)
All continuous functions between topological spaces are measurable, with respect to the Borel σ-algebras (Definition 2.1.6) for the domain and range.

2.3.6 REMARK (BOREL MEASURABILITY). If $f\colon X \to Y$ is a function between topological spaces each equipped with their Borel σ-algebras, some authors emphasize by saying that f is **Borel-measurable** rather than just *measurable*.

If $X = \mathbb{R}^n$, there is a weaker notion of measurability, involving a σ-algebra $\mathcal{L}$ on $\mathbb{R}^n$ that is strictly larger than $\mathscr{B}(\mathbb{R}^n)$; a function $f\colon \mathbb{R}^n \to Y$ is called **Lebesgue-measurable** if it is measurable with respect to $\mathcal{L}$ and $\mathscr{B}(Y)$.

We mostly will not care for this technical distinction (which will be later explained in section 5.6), for $\mathscr{B}(\mathbb{R}^n)$ already contains the measurable sets we need in practice. We only raise this issue now to stave off confusion when consulting other texts.

We are mainly interested in integrating functions taking values in $\mathbb{R}$ or $\overline{\mathbb{R}}$. We can specialize the criterion for measurability for this case.

2.3.7 THEOREM (MEASURABILITY OF REAL-VALUED FUNCTIONS)
Let $(X, \mathcal{M})$ be a measurable space. A map $f\colon X \to \overline{\mathbb{R}}$ is measurable if and only if

☞ for all $c \in \mathbb{R}$, the set $\{f > c\} = f^{-1}((c, +\infty])$ is in $\mathcal{M}$.

Proof Let $\mathcal{G}$ be the set of all open intervals (a, b), for $a, b \in \mathbb{R}$, along with $\{-\infty\}, \{+\infty\}$. Let $\mathcal{H}$ be the set of all intervals $(c, +\infty]$ for $c \in \mathbb{R}$.

Evidently $\mathcal{H}$ generates $\mathcal{G}$:

$$\begin{aligned}
(a, b) &= [-\infty, b) \cap (a, +\infty]\,, \\
[-\infty, b) &= \bigcup_{n=1}^{\infty} [-\infty, b - \tfrac{1}{n}] = \bigcup_{n=1}^{\infty} \overline{\mathbb{R}} \setminus (b - \tfrac{1}{n}, +\infty]\,. \\
\{+\infty\} &= \bigcap_{n=1}^{\infty} (n, +\infty]\,. \\
\{-\infty\} &= \overline{\mathbb{R}} \setminus \bigcup_{n=1}^{\infty} (-n, +\infty]\,.
\end{aligned}$$

Because every open set in $\overline{\mathbb{R}}$ can be written as a countable union of bounded open intervals (a, b), the family $\mathcal{G}$ generates the open sets, and hence the Borel σ-algebra on $\overline{\mathbb{R}}$. Then $\mathcal{H}$ must generate the Borel σ-algebra too.

Applying Theorem 2.3.4 to the generator $\mathcal{H}$ gives the result. ∎

The "countable union with $1/n$" trick used in the proof (which is really the Archimedean property of real numbers) is widely applicable.

2.3.8 REMARK. We can replace $(c, +\infty]$, in the statement of the theorem, by $[c, +\infty]$, $[-\infty, c]$, etc. and there is no essential difference.

The next theorem is especially useful; it addresses one of the problems that we encounter with Riemann integrals already discussed in chapter 1.

2.3.9 THEOREM (MEASURABILITY OF LIMIT FUNCTIONS)
Let $f_1, f_2, \ldots$ be a sequence of measurable $\overline{\mathbb{R}}$-valued functions. Then the functions

$$\sup_n f_n, \ \inf_n f_n, \ \limsup_n f_n, \ \liminf_n f_n$$

(the limits are taken pointwise) are all measurable.

Proof If $g(x) = \sup_n f_n(x)$, then $\{g > c\} = \bigcup_n \{f_n > c\}$, and we apply Theorem 2.3.7. Likewise, if $g(x) = \inf_n f_n(x)$, then $\{g < c\} = \bigcup_n \{f_n < c\}$.

The other two limit functions can be expressed as in terms of supremums and infimums over a countable set, so they are measurable also. ∎

2.3.10 THEOREM (MEASURABILITY OF POSITIVE AND NEGATIVE PART)
If $f\colon X \to \overline{\mathbb{R}}$ is a measurable function, then so are:

$$f^+(x) = \max\{+f(x), 0\} \quad (\textbf{positive part} \text{ of } f)$$
$$f^-(x) = \max\{-f(x), 0\} \quad (\textbf{negative part} \text{ of } f)\,.$$

Conversely, if f^+ and f^- are both measurable, then so is f.

Proof Since $\{-f < c\} = \{f > -c\}$, we see that $-f$ is measurable. Theorem 2.3.9 and Example 2.3.2 then show that f^+ and f^- are both measurable.

For the converse, we express f as the difference $f^+ - f^-$; the measurability of f then follows from the first half of Theorem 2.3.11 below. ∎

We will need to do arithmetic in integration theory, so naturally we must have theorems concerning the measurability of arithmetic operations.

2.3.11 THEOREM (MEASURABILITY OF SUM AND PRODUCT)
If $f, g\colon X \to \mathbb{R}$ are measurable functions, then so are the functions $f + g$, fg, and f/g (provided $g \neq 0$).

Proof Consider the countable union

$$\{f + g < c\} = \bigcup_{r \in \mathbb{Q}} \{f < c - r\} \cap \{g < r\} .$$

The set equality is justified as follows: Clearly $f(x) < c - r$ and $g(x) < r$ together imply $f(x) + g(x) < c$. Conversely, if we set $g(x) = t$, then $f(x) < c - t$, and we can increase t slightly to a rational number r such that $f(x) < c - r$, and $g(x) < t < r$.

The sets $\{f < c - r\}$ and $\{g < r\}$ are measurable, so $\{f + g < c\}$ is too, and by Theorem 2.3.7 the function $f + g$ is measurable. Similarly, $f - g$ is measurable.

To prove measurability of the product fg, we first decompose it into positive and negative parts as described by Theorem 2.3.10:

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

Thus we see that it is enough to prove that fg is measurable when f and g are assumed to be both non-negative. Then just as with the sum, this expression as a countable union:

$$\{fg < c\} = \bigcup_{r \in \mathbb{Q}} \{f < c/r\} \cap \{g < r\} ,$$

shows that fg is measurable.

Finally, for the function $1/g$ with $g > 0$,

$$\{1/g < c\} = \begin{cases} \{g > 1/c\} , & c > 0 , \\ \emptyset , & c \leq 0 . \end{cases}$$

For g that may take negative values, decompose $1/g$ as $1/g^+ - 1/g^-$. ∎

2.3.12 Remark (Re-defining measurable functions at singular points). You probably have already noticed there may be difficulty in defining what the arithmetic operations mean when the operands are infinite, or when dividing by zero. The usual way to deal with these problems is to simply redefine the functions whenever they are infinite to be some fixed value. In particular, if the function g is obtained by changing the original measurable function f on a *measurable* set A to be a constant c, we have:

$$\{g \in B\} = \big(\{g \in B\} \cap A\big) \cup \big(\{g \in B\} \cap A^{\mathrm{c}}\big) ,$$

$$\{g \in B\} \cap A = \begin{cases} A , & c \in B \\ \emptyset , & c \notin B , \end{cases}$$

$$\{g \in B\} \cap A^{\mathrm{c}} = \{f \in B\} \cap A^{\mathrm{c}} ,$$

so the resultant function g is also measurable. Very conveniently, any sets like $A = \{f = +\infty\}$ or $\{f = 0\}$ are automatically measurable. Thus the gaps in Theorem 2.3.11 with respect to infinite values can be repaired with this device.

2.4 Definition of the integral

The idea behind Riemann integration is to try to measure the sums of area of the rectangles "below a graph" of a function and then take some sort of limit. The Lebesgue integral uses a similar approach: we perform integration on the "simple" functions first:

2.4.1 DEFINITION (SIMPLE FUNCTION)
A function is **simple** if its range is a finite set.

2.4.2 DEFINITION (INDICATOR FUNCTION)
For any set $S \subseteq X$, its **indicator function**[†] is a function $\mathbb{I}(S)\colon X \to \{0,1\}$ defined by

$$\mathbb{I}(S)(x) = \mathbb{I}(x \in S) = \begin{cases} 1\,, & x \in S\,, \\ 0\,, & x \notin S\,. \end{cases}$$

The indicator function will play an integral role in our definition of the integral.

2.4.3 REMARK (MEASURABILITY OF INDICATOR). The indicator function $\mathbb{I}(S)$ is a measurable function (Definition 2.3.1) if and only if the set S is measurable (Definition 2.1.1).

For brevity, we will tacitly assume simple functions and indicator functions to be measurable unless otherwise stated.

2.4.4 REMARK (REPRESENTATION OF SIMPLE FUNCTIONS). An $\overline{\mathbb{R}}$-valued simple function φ always has a representation as a finite sum:

$$\varphi = \sum_{k=1}^{n} a_k\, \mathbb{I}(E_k)\,,$$

where a_k are the distinct values of φ, and $E_k = \varphi^{-1}(\{a_k\})$.

Conversely, any expression of the above form, where the values a_k need not be distinct, and the sets E_k are not necessarily $\varphi^{-1}(\{a_k\})$, also defines a simple function. For the purposes of integration, however, we will require that E_k be measurable, and that they partition X.

[†] The indicator function is often called the *characteristic function* (of a set) and denoted by χ_S rather than $\mathbb{I}(S)$. I am not a fan of this notation, as the glyph χ looks too much like x, and the subscript becomes a little annoying to read if the set S is written out as a formula like $\{f \in E\}$. However, the notation adopted here is perfectly standard in probability theory. Also, the word "characteristic" is ambiguous and conflicts with the terminology for a different concept from probability theory.

2.4.5 Definition (Integral of simple function)
Let (X, μ) be a measure space. The **Lebesgue integral**, over X, of a measurable simple function $\varphi \colon X \to [0, \infty]$ is defined as

$$\int_X \varphi \, d\mu = \int_X \sum_{k=1}^n a_k \, \mathbb{I}(E_k) \, d\mu = \sum_{k=1}^n a_k \, \mu(E_k) \, .$$

We restrict φ to be non-negative, to avoid having to deal with $\infty - \infty$ on the right-hand side.

Needless to say, the quantity on the right represents the sum of the areas below the graph of φ.

2.4.6 Remark (Multiplication of zero and infinity). There is a convention, concerning ∞ for the extended real number system, that will be in force whenever we deal with integrals.

$$0 \cdot \pm\infty = 0 \quad \text{always.}$$

In particular, this rule affects the interpretation of the right-hand side in the defining equation of Definition 2.4.5. It is not hard to divine the reason behind this rule:

When integrating functions, we often want to ignore "isolated" singularities, for example, the one at the origin for the integral $\int_0^1 dx/\sqrt{x}$. The point 0 is supposed to have Lebesgue-"measure zero", so even though the integrand is ∞ there, the area contribution at that point should still be $0 = 0 \cdot \infty$. Hence the rule.

Similarly, if c is a constant then we should have $\int_A c \, dx = c \cdot \lambda(A)$, where $\lambda(A)$ is the Lebesgue measure of the set A. Obviously if $c = 0$ then the integral is always zero, even if $\lambda(A) = \infty$, e.g. for $A = \mathbb{R}$. To make this work we must have again $0 = 0 \cdot \infty$.

2.4.7 Remark (Integral is well-defined). It had better be the case that the value of the integral does not depend on the particular representation of φ. If $\varphi = \sum_i a_i \mathbb{I}(A_i) = \sum_j b_j \mathbb{I}(B_j)$, where A_i and B_j partition X (so $A_i \cap B_j$ partition X), then

$$\sum_i a_i \, \mu(A_i) = \sum_j \sum_i a_i \, \mu(A_i \cap B_j) = \sum_j \sum_i b_j \, \mu(A_i \cap B_j) = \sum_j b_j \, \mu(B_j) \, .$$

The second equality follows because the value of φ is $a_i = b_j$ on $A_i \cap B_j$, so $a_i = b_j$ whenever $A_i \cap B_j \neq \emptyset$. So fortunately the integral is well-defined.

Using the same algebraic manipulations just now, you can prove the *monotonicity of the integral for simple functions*: if $\varphi \leq \psi$, then $\int_X \varphi \, d\mu \leq \int_X \psi \, d\mu$.

Our definition of the integral for simple functions also satisfies the linearity property expected for any kind of integral:

2.4.8 Lemma (Linearity of integral for simple functions)
The Lebesgue integral for simple functions is linear.

Proof That $\int_X c\varphi\, d\mu = c\int_X \varphi\, d\mu$, for any constant $c \geq 0$, is clear. And if $\varphi = \sum_i a_i\, \mathbb{I}(A_i)$, $\psi = \sum_j b_j\, \mathbb{I}(B_j)$, we have

$$\begin{aligned}
\int_X \varphi\, d\mu + \int_X \psi\, d\mu &= \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) \\
&= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_j \sum_i b_j \mu(B_j \cap A_i) \\
&= \sum_i \sum_j (a_i + b_j)\, \mu(A_i \cap B_j) \\
&= \int_X (\varphi + \psi)\, d\mu\,.
\end{aligned}$$

■

Next, we integrate non-simple measurable functions like this:

2.4.9 DEFINITION (INTEGRAL OF NON-NEGATIVE FUNCTION)
Let $f\colon X \to [0,\infty]$ be measurable and *non-negative*. Define the **Lebesgue integral** of f over X as

$$\int_X f\, d\mu = \sup \left\{ \int_X \varphi\, d\mu \;\middle|\; \varphi \text{ simple}, 0 \leq \varphi \leq f \right\}.$$

($\int_X \varphi\, d\mu$ is defined in Definition 2.4.5.)

Intuitively, the simple functions φ in the definition are supposed to approximate f, as close as we like from below, and the integral of f is the limit of the integrals of these approximations. Theorem 2.4.10 below provides a particular kind of such approximations.

2.4.10 THEOREM (APPROXIMATION BY SIMPLE FUNCTIONS)
Let $f\colon X \to [0,\infty]$ be measurable. There are simple functions $\varphi_n\colon X \to [0,\infty)$ such that $\varphi_n \nearrow f$, meaning φ_n are increasing pointwise and converging pointwise to f. Moreover, on any set where f is bounded, the φ_n converge to f uniformly.

Proof For any $0 \leq y \leq n$, let $\Psi_n(y)$ be y rounded down to the nearest multiple of 2^{-n}, and whenever $y > n$, truncate $\Psi_n(y)$ to n. Explicitly,

$$\Psi_n = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n}\, \mathbb{I}\Big(\Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big)\Big) + n\, \mathbb{I}([n,\infty])\,.$$

Set φ_n to be the measurable simple functions $\Psi_n \circ f$. Then $0 \leq f(x) - \varphi_n(x) < 2^{-n}$ whenever $n > f(x)$, and $\varphi_n(x) = n$ whenever $f(x) = \infty$. This implies the desired convergence properties, and clearly φ_n are increasing pointwise. ■

And in case you were worrying about whether this new definition of the integral agrees with the old one in the case of the non-negative simple functions, well, it does. Use monotonicity to prove this.

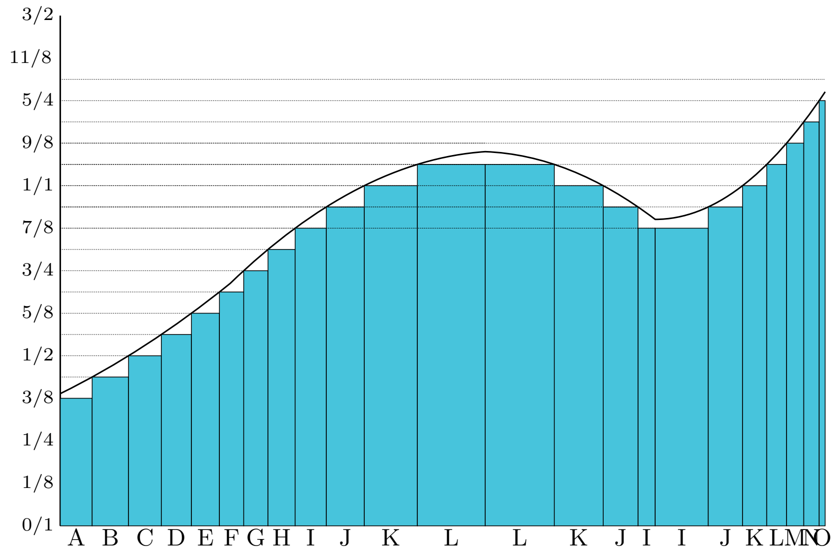

Figure 2.3: In Theorem 2.4.10, we first partition the range $[0, \infty]$. Taking f^{-1} induces a partition on the domain of f, and a subordinate simple function. The approximation to f by that simple function improves as we partition the range more finely. In effect, Lebesgue integration is done by partitioning the *range* and taking limits.

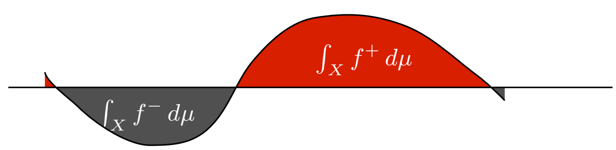

Figure 2.4: $\int_X f \, d\mu$ is the "algebraic area" of the regions bounded by f lying above and below zero.

2.4.11 DEFINITION (INTEGRAL OF MEASURABLE FUNCTION)
For a measurable function $f \colon X \to \overline{\mathbb{R}}$, *not necessarily non-negative*, its **Lebesgue integral** is defined in terms of the integrals of its positive and negative part:

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

provided that the two integrals on the right (from Definition 2.4.9), are not both ∞.

2.4.12 REMARK (NOTATIONS FOR THE INTEGRAL). If we want to display the argument of the integrand function, alternate notations for the integral include:

$$\int_{x \in X} f(x) \, d\mu \,, \quad \int_X f(x) \, d\mu(x) \,, \quad \int_X f(x) \, \mu(dx) \,.$$

For brevity we will even omit certain parts of the integral that are implied by context:

$$\int f\,, \quad \int_X f\,, \quad \int f\,d\mu\,, \quad \int_{x\in X} f(x)\,,$$

although these abbreviations are not to be recommended for general usage.

2.4.13 EXAMPLE (INTEGRAL OVER LEBESGUE MEASURE). If λ is the Lebesgue measure (Example 2.2.3) on $\mathbb{R}^n$, the abstract integral $\int_X f\,d\lambda$ that we have just developed specializes to an extension of the Riemann integral over $\mathbb{R}^n$. This integral[†], as you might expect, is often denoted just as:

$$\int_X f(x)\,dx \quad (X \subseteq \mathbb{R}^n) \quad \text{or} \quad \int_a^b f(x)\,dx \quad (-\infty \leq a \leq b \leq \infty)\,,$$

without mention of the measure λ.

2.4.14 REMARK (INTEGRATING OVER SUBSETS). Often we will want to integrate over subsets of X also. This can be accomplished in two ways. Let A be a measurable subset of X. Either we simply consider integrating over the measure space restricted to subsets of A, or we define

$$\int_A f\,d\mu = \int_X f\,\mathbb{I}(A)\,d\mu\,.$$

If φ is non-negative simple, a simple working out of the two definitions of the integral over A shows that they are equivalent. To prove this for the case of arbitrary measurable functions, we will need the tools of the next section.

2.4.15 THEOREM (MONOTONICITY OF INTEGRAL)
For a measurable function $f\colon X \to \overline{\mathbb{R}}$ on a measure space X, whose integral is defined:

① *Comparison*: If $g\colon X \to \overline{\mathbb{R}}$ is another measurable function whose integral is defined, with $f \leq g$, then $\int_X f \leq \int_X g$.

② *Monotonicity*: If $f \geq 0$, and $A \subseteq B \subseteq X$ are measurable sets, then $\int_A f \leq \int_B f$.

③ *Generalized triangle inequality*: $\left|\int_X f\right| \leq \int_X |f|$.

(“Generalized triangle inequality” alludes to the integral being a generalized form of summation.)

[†] There is no consensus in the literature on whether the term “Lebesgue integral” means the abstract integral developed in this section, or does it only mean the integral applied to Lebesgue measure on $\mathbb{R}^n$. In this book, we will take the former interpretation of the term, and explicitly state a function is to be integrated with Lebesgue measure if the situation is ambiguous.

Proof Assume first that $0 \le f \le g$. Because we have the inclusion

$$\{\varphi \text{ simple}, 0 \le \varphi \le f\} \subseteq \{\varphi \text{ simple}, 0 \le \varphi \le g\},$$

from the definition of the integral of a non-negative function in terms of the supremum, we must have $\int_X f \le \int_X g$.

Now assume only $f \le g$. Then $f^+ \le g^+$ and $f^- \ge g^-$, so:

$$\int_X f = \int_X f^+ - \int_X f^- \le \int_X g^+ - \int_X g^- = \int_X g\,.$$

This proves property ①.

Property ② follows by applying property ① to the two functions $f\mathbb{I}(A) \le f\mathbb{I}(B)$.

Property ③ follows by observing that $-|f| \le f \le |f|$, and from property ① again:

$$-\int_X |f| = \int_X -|f| \le \int_X f \le \int_X |f|\,.$$

■

Naturally, the integral for non-simple functions is linear, but we still have a small hurdle to pass before we are in the position to prove that fact. For now, we make one more convenient definition related to integrals that will be used throughout this book.

2.4.16 DEFINITION (MEASURE ZERO)

☞ A (μ-)measurable set is said to have (μ-)**measure zero** if $\mu(E) = 0$.

☞ A particular property is said to hold **almost everywhere** if the set of points for which the property fails to hold is a set of measure zero.

For example: "a function vanishes almost everywhere"; "$f = g$ almost everywhere".

Typical examples of a measure-zero set are the singleton points in $\mathbb{R}^n$, and lines and curves in $\mathbb{R}^n, n \ge 2$. By countable additivity, any countable set in $\mathbb{R}^n$ has measure zero also.

Clearly, if you integrate anything on a set of measure zero, you get zero. Changing a function on a set of measure zero does not affect the value of its integral.

Assuming that linearity of the integral has been proved, we can demonstrate the following intuitive result.

2.4.17 THEOREM (VANISHING INTEGRALS)

A *non-negative* measurable function $f \colon X \to [0, \infty]$ vanishes almost everywhere if and only if $\int_X f = 0$.

Proof Let $A = \{f = 0\}$, and $\mu(A^c) = 0$. Then

$$\int_X f = \int_X f \cdot (\mathbb{I}(A) + \mathbb{I}(A^c)) = \int_X f\mathbb{I}(A) + \int_X f\mathbb{I}(A^c) = \int_A f + \int_{A^c} f = 0 + 0\,.$$

Conversely, if $\int_X f = 0$, consider $\{f > 0\} = \bigcup_n \{f > \frac{1}{n}\}$. We have

$$\mu\{f > \tfrac{1}{n}\} = \int_{\{f>\frac{1}{n}\}} 1 = n \int_{\{f>\frac{1}{n}\}} \frac{1}{n} \leq n \int_{\{f>\frac{1}{n}\}} f \leq n \int_X f = 0$$

for all n. Hence $\mu\{f > 0\} = 0$. ■

2.5 Exercises

2.1 (Inclusion-exclusion formula) Let (X, μ) be a finite measure space. For any finite number of measurable sets $E_1, \ldots, E_n \subseteq X$,

$$\mu\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{\emptyset \neq S \subseteq \{1,\ldots,n\}} (-1)^{|S|-1} \, \mu\left(\bigcap_{k \in S} E_k\right).$$

For example, for the case $n = 2$:

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

(No deep measure theory is necessary for this problem; it is just a combinatorial argument.)

2.2 (Limit inferior and superior) For any sequence of subsets $E_1, E_2, \ldots$ of X, its **limit superior** and **limit inferior** are defined by:

$$\limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} E_k\,, \quad \liminf_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} E_k\,.$$

The interpretation is that $\limsup_n E_n$ contains those elements of X that occur "infinitely often" in the sets E_n, and $\liminf_n E_n$ contains those elements that occur in all except finitely many of the sets E_n.

Let (X, μ) be a finite measure space, and E_n be measurable. Then the limit superior and limit inferior satisfy these inequalities:

$$\mu(\liminf_n E_n) \leq \liminf_n \mu(E_n) \leq \limsup_n \mu(E_n) \leq \mu(\limsup_n E_n)\,.$$

2.3 (Relation between limit superior of sets and functions) Let f_n be any sequence of $\overline{\mathbb{R}}$-valued functions. Then for any constant $c \in \mathbb{R}$,

$$\left\{\limsup_n f_n > c\right\} \subseteq \limsup_n \{f_n > c\} \subseteq \left\{\limsup_n f_n \geq c\right\}.$$

(The inequality in the middle need not be strict, but the inequalities on the left and right cannot be improved.)

These relations can be useful for showing pointwise convergence of functions, particularly in probability theory. If we want to show that f_n converges to a function f almost everywhere, it is equivalent to show that the set

$$\limsup_n \{|f_n - f| > \varepsilon\}$$

has measure zero for every $\varepsilon > 0$.

2.4 (Continuity of measures) If a finitely additive "measure" function is continuous from below (see Theorem 2.2.7), then it is countably additive. Analogously, provided the space has finite measure, then continuity from above implies countable additivity.

2.5 (Measurability of monotone functions) Show that a monotonically increasing or decreasing function $f\colon \mathbb{R} \to \mathbb{R}$ is measurable.

A curious observation: the limiting sums for the Lebesgue integral of an increasing function $f\colon [a,b] \to \mathbb{R}$ are actually Riemann sums. Not coincidentally, such f is always Riemann-integrable.

2.6 (Measurability of limit domain) Let $f_n\colon X \to \overline{\mathbb{R}}$ be a sequence of measurable functions on a measurable space X. The set of $x \in X$ where $\lim_{n\to\infty} f_n(x)$ exists is measurable.

2.7 (Measurability of limit function) If $f_n\colon X \to \overline{\mathbb{R}}$ is a sequence of measurable functions, then their pointwise limit $\lim_{n\to\infty} f_n$ is also measurable. (If the limit does not exist at a point, define it to be some arbitrary constant.)

2.8 (Measurability of pasted continuous functions) Let $(x,y) \mapsto \theta(x,y)$ be the mapping from a point in $\mathbb{R}^2$ to its polar angle $0 \leq \theta < 2\pi$. Also set $\theta(0,0) = 0$. Defined in this way, θ is not continuous on the non-negative x-axis; nevertheless, it is still a measurable mapping.

More generally, of course, any function manufactured from pasting together continuous functions on measurable sets is going to be measurable.

2.9 (Measurability of uncountable supremum) If $\{f_\alpha\}$ is an *uncountable* family of $\overline{\mathbb{R}}$-valued measurable functions, $\sup_\alpha f_\alpha$ is not necessarily measurable.

However, if f_α are continuous, then uncountability poses no obstacle: $\sup_\alpha f_\alpha$ *is* measurable.

2.10 (Symmetric set difference) If E and F are any subsets within some universe, define their **symmetric set difference** as $E \Delta F = (E^c \cap F) \cup (E \cap F^c)$.

If E and F are measurable for some measure μ, and $\mu(E\Delta F) = 0$, then $\mu(E) = \mu(F)$.

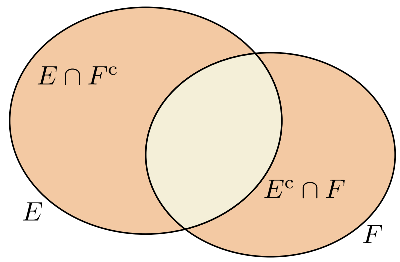

Figure 2.5: Symmetric set difference

2.11 (Metric space on measurable sets) Let $(X, \mathcal{M}, \mu)$ be a measure space. For any $E, F \in \mathcal{M}$, define $d(E, F) = \mu(E \Delta F)$. Then d makes $\mathcal{M}$ into a metric space, provided we "mod out" $\mathcal{M}$ by the equivalence relation $d(E, F) = 0$.

2.12 (Grid approximation to Lebesgue measure) Let $\beta > 1$ be some scaling factor. For any positive integer k, we can draw a rectangular grid $\mathcal{G}(\beta^{-k})$ on the space $\mathbb{R}^n$ consisting of cubes all with side lengths β^{-k} (Like on a piece of graph paper; we may have $\beta = 10$, for example.)

If E is any subset of $\mathbb{R}^n$, its *inner approximation of mesh size* β^{-k} consists of those cubes in $\mathcal{G}(\beta^{-k})$ that are contained in E. Similarly, E has an *outer approximation of mesh size* β^{-k} consisting of those cubes in $\mathcal{G}(\beta^{-k})$ that meet E.

① For any open set $U \subseteq \mathbb{R}^n$, the inner grid approximations to U increase to U (as $k \to \infty$).

② The volume (Lebesgue measure) of U is the limit of the volumes of the inner approximations to U.

③ Similarly, for any compact set $K \subseteq \mathbb{R}^n$, the outer grid approximations to K decrease to K.

④ The volume of K is the limit of the volumes of its outer approximations.

⑤ There are measurable sets E whose volumes are not equal to the limiting volumes of their inner or outer approximations by rectangular grids.

For example, if the topological boundary of E does not have measure zero, the limiting volumes of the inner approximations must disagree with the limiting volumes of the outer approximations. Note that this is precisely the case when E is not Jordan-measurable ($\mathbb{1}(E)$ is not Riemann-integrable; see also the results of section 7.1).

2.13 (Definition of integral by uniform approximations) Figure 2.3 strongly suggests that the integrals of the approximating functions φ_n from Theorem 2.4.10 should approach the integral of f.

This intuition can be turned around to give an alternate definition of $\int f$ for *bounded* functions f on a *finite measure space*. If φ_n are simple functions approaching f *uniformly*, define $\int f = \lim_{n\to\infty} \int \varphi_n$.

Show that this is well-defined: two uniformly convergent sequences approaching the same function have the same limiting integrals. Also prove that linearity and monotonicity holds for this definition of the integral.

2.14 (Bounded convergence theorem) Let $f_n \colon X \to \overline{\mathbb{R}}$ be measurable functions converging to f pointwise everywhere. Assume that f_n are uniformly bounded, and X is a finite measure space. Show that $\lim_{n\to\infty} \int_X |f_n - f| = 0$.

This is a special case of the *dominated convergence theorem* we will present in the next chapter. However, solving this special case provides good practice in deploying measure-theoretic arguments. Hint: Exercise 2.2 and Exercise 2.3.

Chapter 3

Useful results on integration

Here are collected useful results and applications of the Lebesgue integral. Though we will not yet be able to construct all of the foundations for integration theory, by the end of this chapter, you can still begin fruitfully using Lebesgue's integral in place of Riemann's.

3.1 Convergence theorems

This section presents the amazing three convergence theorems available for the Lebesgue integral that make it so much better than the Riemann integral. Behold!

Both the statement and proof of the first theorem below can be motivated from Definition 2.4.9 of $\int_X f \, d\mu$ for non-negative functions f. That definition says $\int_X f \, d\mu$ is the least upper bound of all $\int_X \varphi \, d\mu$ for non-negative simple functions φ *lying below* f. It seems plausible that φ here can in fact be replaced by *any* non-negative measurable function *lying below* f.

3.1.1 Theorem (Monotone convergence theorem)
Let (X, μ) be a measure space. Let $f_n \colon X \to [0, \infty]$ be *non-negative* measurable functions *increasing* pointwise to f. Then

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

Proof f is measurable because it is an increasing limit of measurable functions. Since f_n is an increasing sequence of functions bounded by f, their integrals is an increasing sequence of numbers bounded by $\int_X f \, d\mu$; thus the following limit exists:

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

Next we show the inequality in the other direction.

Take any $0 < t < 1$. Given a fixed simple function $0 \leq \varphi \leq f$, let

$$A_n = \{ f_n \geq t\varphi \} \, .$$

The sets A_n are obviously increasing.

We show that $X = \bigcup_n A_n$. If for a particular $x \in X$, we have $\varphi(x) = 0$, then $x \in A_n$ for all n. Otherwise, $\varphi(x) > 0$, so $f(x) \geq \varphi(x) > t\varphi(x)$, and there is going to be some n for which $f_n(x) \geq t\varphi(x)$, i.e. $x \in A_n$.

For all μ-measurable sets $E \subseteq X$, define

$$\nu(E) = \int_E t\varphi \, d\mu = t \sum_{k=1}^{m} y_k \mu(E \cap E_k)\,, \quad \varphi = \sum_{k=1}^{m} y_k \mathbb{I}(E_k)\,, \quad y_k \geq 0\,.$$

It is not hard to see that ν is a measure. Then

$$\begin{aligned}
\int_X t\varphi \, d\mu = \nu(X) = \nu\Big(\bigcup_{n=1}^{\infty} A_n\Big) &= \lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \int_{A_n} t\varphi \, d\mu \\
&\leq \lim_{n\to\infty} \int_{A_n} f_n \, d\mu\,, \quad \text{since on } A_n \text{ we have } t\varphi \leq f_n \\
&\leq \lim_{n\to\infty} \int_X f_n \, d\mu\,.
\end{aligned}$$

$$\begin{aligned}
t \int_X \varphi \, d\mu &\leq \lim_{n\to\infty} \int_X f_n \, d\mu\,, \quad \text{and take the limit } t \nearrow 1. \\
\int_X \varphi \, d\mu &\leq \lim_{n\to\infty} \int_X f_n \, d\mu\,, \quad \text{and take supremum over all } 0 \leq \varphi \leq f.
\end{aligned}$$ ■

Using the monotone convergence theorem, we can finally prove the linearity of the Lebesgue integral for non-simple functions, which must have been nagging you for a while:

3.1.2 Theorem (Linearity of integral)

If $f, g \colon X \to \overline{\mathbb{R}}$ are measurable functions whose integrals exist, then

① $\int (f+g) = \int f + \int g$ (provided that the right-hand side is not $\infty - \infty$).

② $\int cf = c \int f$ for any constant $c \in \mathbb{R}$.

Proof Assume first that $f, g \geq 0$. By the approximation theorem (Theorem 2.4.10), we can find non-negative simple functions $\varphi_n \nearrow f$, and $\psi_n \nearrow g$. Then $\varphi_n + \psi_n \nearrow f + g$, and so

$$\int f + g = \lim_{n\to\infty} \int \varphi_n + \psi_n = \lim_{n\to\infty} \int \varphi_n + \int \psi_n = \int f + \int g\,.$$

(The second equality follows because we already know the integral is linear for simple functions (Lemma 2.4.8). For the first and third equality we apply the monotone convergence theorem.)

For f, g not necessarily ≥ 0, set $h^+ - h^- = h = f + g = (f^+ - f^-) + (g^+ - g^-)$. This can be rearranged to avoid the negative signs: $h^+ + f^- + g^- = h^- + f^+ + g^+$. (The last equation is valid even if some of the terms are infinite.) Then

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+\,.$$

Regrouping the terms gives $\int h = \int f + \int g$, proving part ①. (This can be done without fear of infinities since $h^\pm \leq f^\pm + g^\pm$, so if $\int f^\pm$ and $\int g^\pm$ are both finite, then so is $\int h^\pm$.)

Part ② can be proven similarly. For the case $f, c \geq 0$, we use approximation by simple functions. For f not necessarily ≥ 0, but still $c \geq 0$,

$$\int cf = \int (cf)^+ - \int (cf)^- = \int cf^+ - \int cf^- = c \int f^+ - c \int f^- = c \int f\,.$$

If $c \leq 0$, write $cf = (-c)(-f)$ to reduce to the preceding case. ■

There are many more applications like this of the monotone convergence theorem. We postpone those for now, in favor of quickly proving the remaining two convergence theorems.

3.1.3 Theorem (Fatou's lemma)

Let $f_n \colon X \to [0, \infty]$ be measurable. Then

$$\int \liminf_{n \to \infty} f_n \leq \liminf_{n \to \infty} \int f_n\,.$$

Proof Set $g_n = \inf_{k \geq n} f_k$, so that $g_n \leq f_n$, and $g_n \nearrow \liminf_n f_n$. Then

$$\int \liminf_n f_n = \int \lim_n g_n = \lim_n \int g_n = \liminf_n \int g_n \leq \liminf_n \int f_n$$

by the monotone convergence theorem. ■

The following definition formulates a crucial hypothesis of the famous *dominated convergence theorem*, probably the most used of the three convergence theorems in applications.

3.1.4 Definition (Lebesgue integrability)

A function $f \colon X \to \overline{\mathbb{R}}$ is called **integrable** if it is measurable and $\int_X |f| < \infty$.

Observe that

$$|f| = f^+ + f^-\,, \quad |f| \geq f^\pm\,,$$

so f is integrable if and only if f^+ and f^- are both integrable. It is also helpful to know, that $\int |f| < \infty$ must imply $|f| < \infty$ almost everywhere.

3.1.5 Theorem (Lebesgue's dominated convergence theorem)

Let (X, μ) be a measure space. Let $f_n \colon X \to \overline{\mathbb{R}}$ be a sequence of measurable functions converging pointwise to f. Moreover, suppose that there is an *integrable* function g such that $|f_n| \leq g$, for all n. Then f_n and f are also integrable, and

$$\lim_{n \to \infty} \int_X |f_n - f|\, d\mu = 0\,.$$

(The function g is said to *dominate* over f_n.)

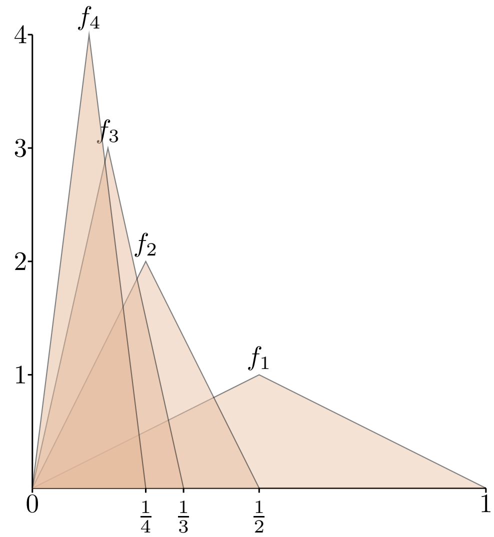

Figure 3.1: A mnemonic for remembering which way the inequality goes in Fatou's lemma. Consider these "witches' hat" functions f_n. As $n \to \infty$, the area under the hats stays constant even though $f_n(x) \to 0$ for each $x \in [0,1]$.

Proof Obviously f_n and f are integrable. Also, $2g - |f_n - f|$ is measurable and non-negative. By Fatou's lemma (Theorem 3.1.3),

$$\int \liminf_n (2g - |f_n - f|) \le \liminf_n \int (2g - |f_n - f|)\,.$$

Since f_n converges to f, the left-hand quantity is just $\int 2g$. The right-hand quantity is:

$$\begin{aligned}\liminf_n \left(\int 2g - \int |f_n - f| \right) &= \int 2g + \liminf_n \left(- \int |f_n - f| \right) \\ &= \int 2g - \limsup_n \int |f_n - f|\,.\end{aligned}$$

Since $\int 2g$ is finite, it may be cancelled from both sides. Then we obtain

$$\limsup_n \int |f_n - f| \le 0\,, \quad \text{that is,} \lim_{n\to\infty} \int |f_n - f| = 0\,.$$

■

3.1.6 Remark (Almost everywhere). We only need to require that f_n converge to f pointwise almost everywhere, or that $|f_n|$ is bounded above by g almost everywhere. (However, if f_n only converges to f almost everywhere, then the theorem would not automatically say that f is measurable.)

3.1.7 REMARK (INTERCHANGE OF LIMIT AND INTEGRAL). By the *generalized triangle inequality* (Theorem 2.4.15), we can also conclude that

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

which is usually how the *dominated convergence theorem* is applied.

The dominated convergence theorem can be vaguely described by saying, given that the graphs of the functions f_n are enveloped by $\pm g$, if $f_n \to f$, then the integrals of f_n are forced to converge to that of f because the area under the graphs of f_n have no opportunity to "escape" out to infinity.

For the functions in fig. 3.1, the functions f_n *do* have an opportunity to escape to infinity, and indeed the limit of integrals does not equal the integral of the limit.

The dominated convergence theorem is a manifestation of the seemingly ubiquitous principle in analysis, that if we can show something is *bounded* or *absolutely convergent*, then we get nice regularity properties.

3.1.8 REMARK (CONTINUOUS LIMITS FOR DOMINATED CONVERGENCE). The theorem also holds for continuous limits of functions, not just countable limits. That is, if we have a continuous sequence of functions, say f_t, $0 \leq t < 1$, we can also say

$$\lim_{t\to 1} \int_X |f_t - f| \, d\mu = 0 \,;$$

for given any sequence $\{a_n\}$ convergent to 1, we can apply the theorem to f_{a_n}. Since this can be done for *any* sequence convergent to 1, the continuous limit is established.

In the next sections, we will take the convergence theorems obtained here to prove useful results.

3.2 Interchange of summation and integral

3.2.1 THEOREM (BEPPO-LEVI)

Let $f_n \colon X \to [0, \infty]$ be *non-negative* measurable functions. Then

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

Proof Let $g_N = \sum_{n=1}^{N} f_n$, and $g = \sum_{n=1}^{\infty} f_n$. By monotone convergence:

$$\int g = \int \lim_{N\to\infty} g_N = \lim_{N\to\infty} \int g_N = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \sum_{n=1}^{\infty} \int f_n \,.$$

■

3.2.2 THEOREM (GENERALIZATION OF BEPPO-LEVI)
Let $f_n \colon X \to \overline{\mathbb{R}}$ be measurable functions, with[†] $\int \sum_n |f_n| = \sum_n \int |f_n|$ being finite. Then

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

Proof Let $g_N = \sum_{n=1}^{N} f_n$ and $h = \sum_{n=1}^{\infty} |f_n|$. Then $|g_N| \leq |h|$. Since $\int |h| < \infty$ by hypothesis, $|h| < \infty$ almost everywhere, so g_N is (absolutely) convergent almost everywhere. By the dominated convergence theorem,

$$\sum_{n=1}^{\infty} \int f_n = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int g_N = \int \limsup_{N\to\infty} g_N = \int \sum_{n=1}^{\infty} f_n \,. \qquad \blacksquare$$

3.2.3 EXAMPLE (DOUBLY-INDEXED SUMS). Here is a perhaps unexpected application. Suppose we have a countable set of real numbers $a_{n,m}$, for $n, m \in \mathbb{N}$. Let μ be the counting measure (Example 2.2.2) on $\mathbb{N}$. Then

$$\int_{m\in\mathbb{N}} a_{n,m} \, d\mu = \sum_{m=1}^{\infty} a_{n,m} \,.$$

Moreover, Theorem 3.2.2 says that we can sum either along n first or m first and get the same results,

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{n,m} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{n,m} \,,$$

as long as the double sum is *absolutely convergent*. (Or the numbers a_{mn} are all *non-negative*, so the order of summation should not possibly matter.) Of course, this fact can also be proven in an entirely elementary way.

3.3 Mass density functions

3.3.1 THEOREM (MEASURES GENERATED BY MASS DENSITIES)
Let $g \colon X \to [0, \infty]$ be measurable in the measure space $(X, \mathcal{A}, \mu)$. Let

$$\nu(E) = \int_E g \, d\mu \,, \quad E \in \mathcal{A} \,.$$

Then ν is a measure on $(X, \mathcal{A})$, and for any measurable $f \colon X \to \overline{\mathbb{R}}$,

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

[†] The first Beppo-Levi theorem, Theorem 3.2.1, shows that always $\int \sum_{n=1}^{\infty} |f_n| = \sum_{n=1}^{\infty} \int |f_n|$, whether it is finite or not.

often written as[†] $d\nu = g\,d\mu$.

Proof We prove ν is a measure. $\nu(\emptyset) = 0$ is trivial. For countable additivity, let $\{E_n\}$ be measurable with disjoint union E, so that $\mathbb{I}(E) = \sum_{n=1}^{\infty} \mathbb{I}(E_n)$, and

$$\begin{aligned}\nu(E) &= \int_E g\,d\mu = \int_X g\,\mathbb{I}(E)\,d\mu \\ &= \int_X \sum_{n=1}^{\infty} g\,\mathbb{I}(E_n)\,d\mu = \sum_{n=1}^{\infty} \int_X g\,\mathbb{I}(E_n)\,d\mu = \sum_{n=1}^{\infty} \nu(E_n)\,.\end{aligned}$$

Next, if $f = \mathbb{I}(E)$ for some $E \in \mathcal{A}$, then

$$\int_X f\,d\nu = \int_X \mathbb{I}(E)\,d\nu = \nu(E) = \int_X \mathbb{I}(E) g\,d\mu = \int_X fg\,d\mu\,.$$

By linearity, we see that $\int f\,d\nu = \int fg\,d\mu$ whenever f is non-negative simple. For general non-negative f, we use a sequence of simple approximations $\varphi_n \nearrow f$, so $\varphi_n g \nearrow fg$. Then by monotone convergence,

$$\int_X f\,d\nu = \lim_{n\to\infty} \int_X \varphi_n\,d\nu = \lim_{n\to\infty} \int_X \varphi_n g\,d\mu = \int_X \lim_{n\to\infty} \varphi_n g\,d\mu = \int_X fg\,d\mu\,.$$

Finally, for f not necessarily non-negative, we apply the above to its positive and negative parts, then subtract them. ■

3.3.2 REMARK (PROOFS BY APPROXIMATION WITH SIMPLE FUNCTIONS). The proof of Theorem 3.3.1 illustrates the standard procedure to proving certain facts about integrals. We first reduce to the case of simple functions and non-negative functions, and then take limits with either the monotone convergence theorem or dominated convergence theorem.

This standard procedure will be used over and over again. (It will get quite monotonous if we had to write it in detail every time we use it, so we will abbreviate the process if the circumstances permit.)

Also, we should note that if f is only measurable but not integrable, then the integrals of f^+ or f^- might be infinite. If both are infinite, the integral of f is not defined, although the equation of the theorem might still be interpreted as saying that the left-hand and right-hand sides are undefined at the same time. For this reason, and for the sake of the clarity of our exposition, we will not bother to modify the hypotheses of the theorem to state that f must be integrable.

Problem cases like this also occur for some of the other theorems we present, and there I will also not make too much of a fuss about these problems, trusting that you understand what happens when certain integrals are undefined.

[†] The function g is sometimes called the density function of ν with respect to μ; we will have much more to say about density functions in a later section.

3.4 Change of variables

3.4.1 THEOREM (CHANGE OF VARIABLES)
Let X, Y be measure spaces, and $g\colon X \to Y$, $f\colon Y \to \overline{\mathbb{R}}$ be measurable. Then

$$\int_X (f \circ g)\, d\mu = \int_Y f\, d\nu\,, \quad \nu(B) = \mu(g^{-1}(B)) \text{ for measurable } B \subseteq Y.$$

ν is called a **transport measure** or **pullback measure**.

Proof First suppose $f = \mathbb{I}(B)$. Let $A = g^{-1}(B) \subseteq X$. Then $f \circ g = \mathbb{I}(A)$, and we have

$$\int_Y f\, d\nu = \int_Y \mathbb{I}(B)\, d\nu = \nu(B) = \mu(g^{-1}(B)) = \mu(A) = \int_X (f \circ g)\, d\mu\,.$$

Since both sides of this equation are linear in f, the equation holds whenever f is simple. Applying the "standard procedure" mentioned above (Remark 3.3.2), the equation is then proved for all measurable f. ∎

3.4.2 REMARK (CHANGE OF VARIABLES IN REVERSE). The change of variables theorem can also be applied "in reverse". Suppose we want to compute $\int_Y f\, d\nu$, where ν is already given to us. Further assume that g is bijective and its inverse is measurable. Then we can define $\mu(A) = \nu(g(A))$, and it follows that $\int_Y f\, d\nu = \int_X (f \circ g)\, d\mu$.

3.4.3 REMARK (ALTERNATE NOTATION FOR CHANGE OF VARIABLES). This alternate notation may help in remembering the change-of-variables formula. To transform

$$\int_Y f(y)\, \nu(dy) \leftrightarrow \int_X f(g(x))\, \mu(dx)\,,$$

substitute

$$y = g(x)\,, \quad \nu(dy) = \mu(dx)\,, \quad dx = g^{-1}(dy)\,.$$

If g^{-1} is also a measurable function, then we may also write $dy = g(dx)$.

Our theorem, especially when stated in the reverse form, is clearly related to the usual "change of variables" theorem in calculus. If $g\colon X \to Y$ is a bijection between open subsets of $\mathbb{R}^n$, and both it and its inverse are continuously differentiable (that is, g is a **diffeomorphism**), and λ is the Lebesgue measure in $\mathbb{R}^n$, then, as we shall prove rigorously in Lemma 7.2.1,

$$\lambda(g(A)) = \int_A |\det \mathrm{D}g|\, d\lambda\,.$$ ❀

Then we can obtain:

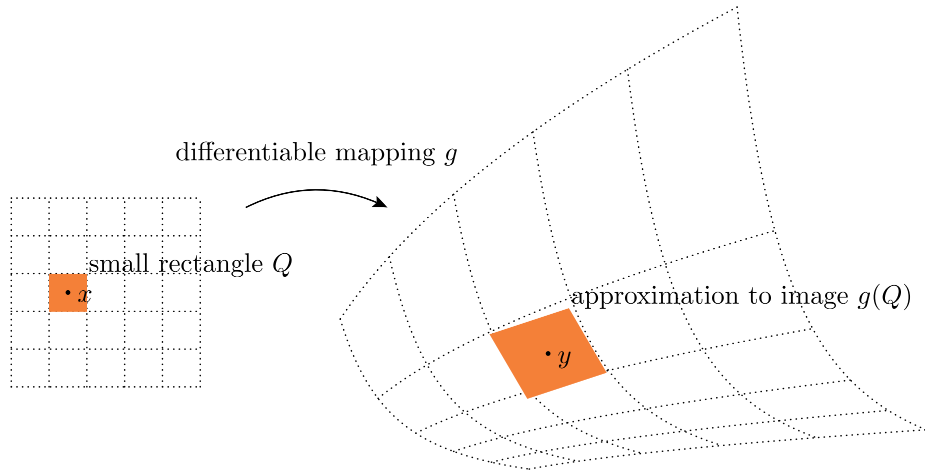

Figure 3.2: Explanation of the change-of-variables formula $dy = |\det \mathrm{D}g(x)|\, dx$. When an image $g(Q)$ is approximated by the linear application $y + \mathrm{D}g(x)(Q - x)$, its volume is scaled by a factor of $|\det \mathrm{D}g(x)|$.

3.4.4 Theorem (Differential change of variables in $\mathbb{R}^n$)

Let $g\colon X \to Y$ be a diffeomorphism of open sets in $\mathbb{R}^n$. If $A \subseteq X$ is measurable, and $f\colon Y \to \overline{\mathbb{R}}$ is measurable, then

$$\int_{g(A)} f(y)\, dy = \int_A f(g(x))\, g(dx) = \int_A f(g(x)) \cdot |\det \mathrm{D}g(x)|\, dx\,.$$

(Substitute $y = g(x)$ and $dy = g(dx) = |\det \mathrm{D}g(x)|\, dx$.)

Proof Take $\nu = \lambda$ and $\mu = \nu \circ g$ as our abstract change of variables, appealing to Theorem 3.4.1 and Remark 3.4.2. Then

$$\int_Y f\, d\lambda = \int_X (f \circ g)\, d(\lambda \circ g)\,.$$

We have $d(\lambda \circ g) = |\det \mathrm{D}g|\, d\lambda$ from equation ❀, so by Theorem 3.3.1,

$$\int_X (f \circ g)\, d(\lambda \circ g) = \int_X (f \circ g) \cdot |\det \mathrm{D}g|\, d\lambda\,.$$

If we replace f by $f\,\mathbb{I}(g(A))$, then

$$\int_{g(A)} f\, d\lambda = \int_A (f \circ g) \cdot |\det \mathrm{D}g|\, d\lambda\,.$$

■

3.5 Integrals with parameter

The next theorem is the Lebesgue version of a well-known result about the Riemann integral.

3.5.1 THEOREM (FIRST FUNDAMENTAL THEOREM OF CALCULUS)
Let $X \subseteq \mathbb{R}$ be an interval, and $f \colon X \to \mathbb{R}$ be integrable with Lebesgue measure on $\mathbb{R}$. Then the function

$$F(x) = \int_a^x f(t)\,dt$$

is continuous. Furthermore, if f is continuous at x, then $F'(x) = f(x)$.

Proof To prove continuity, we compute:

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\,dt = \int_X f(t) \cdot \mathbb{I}(t \in [x, x+h])\,dt\,.$$

In this proof, the indicator $\mathbb{I}(t \in [x, x+h])$ should be interpreted as $-\mathbb{I}(t \in [x+h, x])$ whenever $h < 0$, so that $\int_x^{x+h} = -\int_{x+h}^x$ as in calculus.

Since $|f \cdot \mathbb{I}([x, x+h])| \leq |f|$, we can apply the dominated convergence theorem, together with Remark 3.1.8, to conclude:

$$\begin{aligned}
\lim_{h \to 0} F(x+h) - F(x) &= \lim_{h \to 0} \int_X f(t) \cdot \mathbb{I}(t \in [x, x+h])\,dt \\
&= \int_X \lim_{h \to 0} f(t) \cdot \mathbb{I}(t \in [x, x+h])\,dt \\
&= \int_X f(t) \cdot \mathbb{I}(t \in \{x\})\,dt = 0\,.
\end{aligned}$$

The proof of differentiability is the same as for the Riemann integral:

$$\begin{aligned}
\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \left| \frac{\int_x^{x+h} (f(t) - f(x))\,dt}{h} \right| \\
&\leq \frac{\int_{[x,x+h]} |f(t) - f(x)|\,dt}{|h|} \\
&\leq \frac{\sup_{t \in [x,x+h]} |f(t) - f(x)| \cdot |h|}{|h|}\,,
\end{aligned}$$

which goes to zero as h does. ■

You may be wondering, as a calculus student does, whether the hypothesis of continuity of the integrand may be weakened in Theorem 3.5.1. The answer is affirmative; in chapter 10 we shall prove a great generalization due to Lebesgue himself. For now, this theorem is all we need to compute integrals over the real line analytically in the usual way.

The following theorems are often not found in calculus texts even though they are quite important for applied calculations.

3.5.2 THEOREM (CONTINUOUS DEPENDENCE ON INTEGRAL PARAMETER)
Let X be a measure space, T be any metric space (e.g. $\mathbb{R}^n$), and $f\colon X \times T \to \overline{\mathbb{R}}$, with $f(\cdot, t)$ being measurable for each $t \in T$. Then

$$F(t) = \int_{x \in X} f(x, t)$$

is continuous at $t_0 \in T$, provided the following conditions are met:

① For each $x \in X$, $f(x, \cdot)$ is continuous at t_0.

② There is an integrable function g such that $|f(x,t)| \leq g(x)$ for all $t \in T$.

Proof The hypotheses have been formulated so we can immediately apply dominated convergence:

$$\lim_{t \to t_0} \int_{x \in X} f(x,t) = \int_{x \in X} \lim_{t \to t_0} f(x,t) = \int_{x \in X} f(x, t_0)\,.$$

■

3.5.3 THEOREM (DIFFERENTIATION UNDER THE INTEGRAL SIGN)
Let X be a measure space, T be an open interval of $\mathbb{R}^n$, and $f\colon X \times T \to \mathbb{R}$, with $f(\cdot, t)$ being measurable for each $t \in T$. Then

$$F(t) = \int_{x \in X} f(x, t)\,,$$

is differentiable with the derivative:

$$F'(t) = \frac{d}{dt} \int_{x \in X} f(x,t) = \int_{x \in X} \frac{\partial}{\partial t} f(x,t)\,,$$

provided the following conditions are satisfied:

① For each $x \in X$, $\frac{\partial}{\partial t} f(x,t)$ exists for all $t \in T$.

② There is an integrable function g such that $\left|\frac{\partial}{\partial t} f(x,t)\right| \leq g(x)$ for all $t \in T$.

Proof This theorem is often proven by using iterated integrals and switching the order of integration, but that method is theoretically troublesome because it requires more stringent hypotheses. It is easier, and better, to prove it directly from the definition of the derivative.

$$\begin{aligned}\lim_{h \to 0} \frac{F(t+h) - F(t)}{h} &= \lim_{h \to 0} \int_{x \in X} \frac{f(x, t+h) - f(x,t)}{h} \\ &= \int_{x \in X} \lim_{h \to 0} \frac{f(x, t+h) - f(x,t)}{h} = \int_{x \in X} \frac{\partial}{\partial t} f(x,t)\,.\end{aligned}$$

Note that $\left|\frac{f(x,t+h) - f(x,t)}{h}\right|$ can be bounded by $g(x)$ using the mean value theorem of calculus.

■

3.5.4 REMARK. It is easy to see that we may generalize Theorem 3.5.3 to T being any open set in $\mathbb{R}^n$, taking partial derivatives.

3.5.5 EXAMPLE. The inveterate Γ function, defined on the positive reals by

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt\,,$$

is continuous, and differentiable with the obvious formula for the derivative.

3.6 Exercises

3.1 (Hypotheses for monotone convergence) Can we weaken the requirement, in the monotone convergence theorem and Fatou's lemma, that $f_n \geq 0$?

If f_n are measurable functions decreasing to f, is it true that $\lim_n \int f_n = \int f$? If not, what additional hypotheses are needed?

3.2 (Increasing classes of measures) Let μ_n be an increasing sequence of measures be defined on a common measurable space. (That is, $\mu_n(E) \leq \mu_{n+1}(E)$ for all measurable E.) Then $\mu = \sup\limits_n \mu_n$ is also a measure.

3.3 (Jensen's inequality) Let (X, μ) be a finite measure space, and $g\colon I \to \mathbb{R}$ be a (non-strict) convex function on an open interval $I \subseteq \mathbb{R}$. If $f\colon X \to I$ is any integrable function, then

$$g\Big(\frac{1}{\mu(X)} \int_X f\, d\mu\Big) \leq \frac{1}{\mu(X)} \int_X (g \circ f)\, d\mu\,.$$

The inequality is reversed if g is concave instead of convex.

3.4 (Limit of averages of real function) If $f\colon [0,\infty) \to \mathbb{R}$ be a measurable function, and $\lim\limits_{x\to\infty} f(x) = a$, then

$$\lim_{x\to\infty} \frac{1}{x} \int_0^x f(t)\, dt = a\,.$$

3.5 (Continuous functions equal to zero a.e.) If $f\colon \mathbb{R}^n \to \mathbb{R}$ is continuous and equal to zero almost everywhere, then f is in fact equal to zero everywhere.

3.6 (Weak convergence to Dirac delta) Let $g\colon \mathbb{R}^n \to [0,\infty]$ be an integrable function with $\int_{\mathbb{R}^n} g(x)\, dx = 1$. Show that for any bounded function $f\colon \mathbb{R}^n \to \mathbb{R}$ continuous at 0,

$$\lim_{\varepsilon \searrow 0} \int_{\mathbb{R}^n} \frac{1}{\varepsilon^n} f(x)\, g\left(\frac{x}{\varepsilon}\right)\, dx = f(0) = \int_{\mathbb{R}^n} f(x)\, d\delta(x)\,.$$

Thus, as $\varepsilon \searrow 0$, the measures $\nu_\varepsilon(A) = \int_A \varepsilon^{-n} g(x/\varepsilon)\, dx$ converge, in some sense, to δ, the *Dirac measure* around 0 described in Example 2.2.4.

3.7 (Dominated convergence with lower and upper bounds) Given the intuitive interpretation of the dominated convergence theorem given in the text, it seems plausible that the following is a valid generalization. Prove it.

Let $L_n \leq f_n \leq U_n$ be three sequences of $\overline{\mathbb{R}}$-valued measurable functions that converge to L, f, U respectively. Assume L and U are integrable, and $\int L_n \to \int L$, $\int U_n \to \int U$. Then f is integrable and $\int f_n \to \int f$.

3.8 (Weak second fundamental theorem of calculus) Prove the following.

Suppose $F\colon [a,b] \to \mathbb{R}$ is differentiable with *bounded derivative*. Then F' is integrable on $[a,b]$ and

$$\int_a^b F'(x)\,dx = F(b) - F(a)\,.$$

This result manages to avoid requiring F' to be continuous.

However, the hypothesis that F' be bounded cannot be completely removed, though it can be weakened. There are functions F that oscillate so quickly up and down that $|F'|$ is not integrable. The typical counterexample found in calculus textbooks is the function illustrated in Figure ??.

Chapter 4

Some commonly encountered spaces

4.1 Space of integrable functions

This section is a short introduction to spaces of integrable functions, and the properties of convergence in these spaces.

4.1.1 DEFINITION ($\mathbf{L}^p$ SPACE)
Let (X, μ) be a measure space, and let $1 \leq p < \infty$. The space $\mathbf{L}^p(\mu)$ consists of all measurable functions $f \colon X \to \overline{\mathbb{R}}$ such that

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

For such f, define

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p} .$$

(If $f \colon X \to \overline{\mathbb{R}}$ is measurable but $|f|^p$ is not integrable, set $\|f\|_p = \infty$.)

As suggested by the notation, $\|\cdot\|_p$ should be a norm on the vector space $\mathbf{L}^p$. In the formal sense it is not, because $\|f\|_p = 0$ only implies that f is zero *almost everywhere*, not that it is zero identically. This minor annoyance can be resolved logically by redefining $\mathbf{L}^p$ as the *equivalence classes* of functions that equal each other except on sets of measure zero. Though most people still prefer to think of $\mathbf{L}^p$ as consisting of actual functions.

Only the verification of the triangle inequality for $\|\cdot\|_p$ presents any difficulties — they will be solved by the theorems below.

4.1.2 DEFINITION (CONJUGATE EXPONENTS)
Two numbers $1 < p, q < \infty$ are called **conjugate exponents** if $\frac{1}{p} + \frac{1}{q} = 1$.

4.1.3 THEOREM (HÖLDER'S INEQUALITY)
For $\overline{\mathbb{R}}$-valued measurable functions f and g, and conjugate exponents p and q:

$$\left|\int fg\right| \leq \int |f||g| \leq \| f\|_p \|g\|_q \,.$$

Proof The first inequality is trivial. For the second inequality, since it only involves absolute values of functions, for the rest of the proof we may assume that f, g are non-negative.

If $\|f\|_p = 0$, then $|f|^p = 0$ almost everywhere, and so $f = 0$ and $fg = 0$ almost everywhere too. Thus the inequality is valid in this case. Likewise, when $\|g\|_q = 0$.

If $\|f\|_p$ or $\|g\|_q$ is infinite, the inequality is trivial.

So we now assume these two quantities are both finite and non-zero. Define $F = f/\|f\|_p$, $G = g/\|g\|_q$, so that $\|F\|_p = \|G\|_q = 1$. We must then show that $\int FG \leq 1$.

To do this, we employ the fact that the natural logarithm function is concave:

$$\frac{1}{p}\log s + \frac{1}{q}\log t \leq \log\left(\frac{s}{p} + \frac{t}{q}\right), \quad 0 \leq s, t \leq \infty\,;$$

or,

$$s^{1/p}\, t^{1/q} \leq \tfrac{s}{p} + \tfrac{t}{q}\,.$$

Substitute $s = F^p$, $t = G^q$, and integrate both sides:

$$\int FG \leq \frac{1}{p}\int F^p + \frac{1}{q}\int G^q = \frac{1}{p}\|F\|_p^p + \frac{1}{q}\|G\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.$$

■

The special case $p = q = 2$ of Hölder's inequality is the ubiquitous Cauchy-Schwarz inequality.

4.1.4 THEOREM (MINKOWSKI'S INEQUALITY)
For $\overline{\mathbb{R}}$-valued measurable functions f and g, and conjugate exponents p and q:

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

Proof The inequality is trivial when $p = 1$ or when $\|f + g\|_p = 0$. Also, since $\|f + g\|_p \leq \||f| + |g|\|_p$, it again suffices to consider only the case when f, g are non-negative.

Since the function $t \mapsto t^p$, for $t \geq 0$ and $p > 1$, is convex, we have

$$\left(\frac{f+g}{2}\right)^p \leq \frac{f^p + g^p}{2}\,.$$

This inequality shows that if $\|f + g\|_p$ is infinite, then one of $\|f\|_p$ or $\|g\|_p$ must also be infinite, so Minkowski's inequality holds true in that case.

We may now assume $\|f+g\|_p$ is finite. We write:

$$\int (f+g)^p = \int f(f+g)^{p-1} + \int g(f+g)^{p-1}.$$

By Hölder's inequality, and noting that $(p-1)q = p$ for conjugate exponents,

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

A similar inequality holds for $\int g(f+g)^{p-1}$. Putting these together:

$$\|f+g\|_p^p \le (\|f\|_p + \|g\|_p)\|f+g\|_p^{p/q}.$$

Dividing by $\|f+g\|_p^{p/q}$ yields the result. ■

4.1.5 REMARK (MINKOWSKI'S INEQUALITY FOR INFINITE SUMS). Minkowski's inequality also holds for infinitely many summands:

$$\left\|\sum_{n=1}^{\infty} |f_n|\right\|_p \le \sum_{n=1}^{\infty} \|f_n\|_p,$$

and is extrapolated from the case of finite sums by taking limits.

Among the $\mathbf{L}^p$ spaces, the ones that prove most useful are undoubtedly $\mathbf{L}^1$ and $\mathbf{L}^2$. The former obviously because there are no messy exponents floating about; the latter because it can be made into a real or complex Hilbert space by the inner product:

$$\langle f, g\rangle = \int_X f\overline{g}\, d\mu.$$

A simple relation amongst the $\mathbf{L}^p$ spaces is given by the following.

4.1.6 THEOREM

Let (X, μ) have finite measure. Then $\mathbf{L}^p \subseteq \mathbf{L}^r$ whenever $1 \le r < p < \infty$. Moreover, the inclusion map from $\mathbf{L}^p$ to $\mathbf{L}^r$ is continuous.

Proof For $f \in \mathbf{L}^p$, Hölder's inequality with conjugate exponents $\frac{p}{r}$ and $s = \frac{p}{p-r}$ states:

$$\|f\|_r^r = \int |f|^r \le \left(\int |f|^{r\cdot\frac{p}{r}}\right)^{r/p} \left(\int 1^s\right)^{1/s} = \|f\|_p^r\, \mu(X)^{1/s},$$

and so

$$\|f\|_r \le \|f\|_p\, \mu(X)^{1/rs} = \|f\|_p\, \mu(X)^{\frac{1}{r}-\frac{1}{p}} < \infty.$$

To show continuity of the inclusion, replace f with $f-g$ where $\|f-g\|_p < \varepsilon$. ■

4.1.7 EXAMPLE. $\int_0^1 x^{-\frac{1}{2}}\,dx = 2 < \infty$, so automatically $\int_0^1 x^{-\frac{1}{4}}\,dx < \infty$. On the other hand, the condition that $\mu(X) < \infty$ is indeed necessary: $\int_1^\infty x^{-2}\,dx < \infty$, but $\int_1^\infty x^{-1}\,dx = \infty$.

4.1.8 THEOREM (DOMINATED CONVERGENCE IN $\mathbf{L}^p$)
Let $f_n \colon X \to \mathbb{R}$ be measurable functions converging (almost everywhere) pointwise to f, and $|f_n| \le g$ for some $g \in \mathbf{L}^p$, $1 \le p < \infty$. Then $f, f_n \in \mathbf{L}^p$, and f_n converges to f in the $\mathbf{L}^p$ norm, meaning:

$$\lim_{n\to\infty} \left(\int |f_n - f|^p \right)^{1/p} = \lim_{n\to\infty} \|f_n - f\|_p = 0.$$

Proof $|f_n - f|^p$ converges to 0 and $|f_n - f|^p \le (2g)^p \in \mathbf{L}^1$. Apply the usual dominated convergence theorem on these functions. ■

In analysis, it is important to know that the spaces of functions we are working are complete; $\mathbf{L}^p$ is no exception. (Contrast with the Riemann integral.)

4.1.9 THEOREM
$\mathbf{L}^p$, for $1 \le p < \infty$, is a complete normed vector space, that is, a Banach space.

Proof It suffices to show that every absolutely convergent series in $\mathbf{L}^p$ is convergent. Let $f_n \in \mathbf{L}^p$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_p < \infty$. By Minkowski's inequality for infinite sums (Remark 4.1.5), $\|\sum_{n=1}^\infty |f_n|\|_p < \infty$. Then $g = \sum_{n=1}^\infty f_n$ converges absolutely almost everywhere, and

$$\Big\| \sum_{n=1}^{N} f_n - g \Big\|_p = \Big\| \sum_{n=N+1}^{\infty} f_n \Big\|_p \le \sum_{n=N+1}^{\infty} \|f_n\|_p \to 0 \qquad \text{as } N \to \infty,$$

meaning that $\sum_{n=1}^\infty f_n$ converges in $\mathbf{L}^p$ norm to g. ■

For completeness (pun intended), we can also define a space called "$\mathbf{L}^\infty$":

4.1.10 DEFINITION
Let X be a measure space, and $f \colon X \to \overline{\mathbb{R}}$ be measurable. A number $M \in [0, \infty]$ is an almost-everywhere upper bound for $|f|$ if $|f| \le M$ almost everywhere. The infimum of all almost-everywhere upper bounds for $|f|$ is the **essential supremum**, denoted by $\|f\|_\infty$ or $\operatorname{esssup}|f|$.

4.1.11 DEFINITION
$\mathbf{L}^\infty$ is the set of all measurable functions f with $\|f\|_\infty < \infty$. Its norm is given by $\|f\|_\infty$.

"$q = \infty$" is considered a conjugate exponent to $p = 1$, and it is trivial to extend Hölder's inequality to apply to the exponents $p = 1, q = \infty$. In some respects, $\mathbf{L}^\infty$ behaves like the other $\mathbf{L}^p$ spaces, and at other times it does not; a few pertinent results are presented in the exercises.

We wrap up our brief tour of $\mathbf{L}^p$ spaces with a discussion of a remarkable property. For any $f \in \mathbf{L}^p$, let us take some simple approximations $\varphi_n \to f$ such that $|\varphi|$ is dominated by $|f|$ (Theorem 2.4.10). Then φ_n converge to f in $\mathbf{L}^p$ (Theorem 4.1.8), and we have proved:

4.1.12 THEOREM

Suppose we have $f \in \mathbf{L}^p(X, \mu)$, $1 \leq p < \infty$. For any $\varepsilon > 0$, there exist simple functions $\varphi \colon X \to \mathbb{R}$ such that

$$\|\varphi - f\|_p = \left(\int_X |\varphi - f|^p \, d\mu\right)^{1/p} < \varepsilon .$$

This may be summarized as: *the simple functions are dense in $\mathbf{L}^p$ (in the topological sense)*.

In this spruced up formulation, we can imagine other generalizations. For example, if $X = \mathbb{R}^n$ and $\mu = \lambda$ is Lebesgue measure, then we may venture a guess that the continuous functions are dense in $\mathbf{L}^p(\mathbb{R}^n, \lambda)$. In other words, $\overline{\mathbb{R}}$-valued functions that are measurable with respect to the Borel σ-finite generated by the topology for $\mathbb{R}^n$ can be approximated as well as we like by functions continuous with respect to that topology.

We can even go farther, and inquire whether approximations by smooth functions are good enough. And yes, this turns out to be true: the set $\mathbf{C}_0^\infty$ of infinitely differentiable functions on $\mathbb{R}^n$ with compact support[1] is dense in $\mathbf{L}^p(\mathbb{R}^n, \lambda)$.

At this time, we are hardly prepared to demonstrate these facts, but they are useful in the same manner that many theorems for Lebesgue integrals are proven by approximating with simple functions. Some practice will be given in the chapter exercises.

4.2 Probability spaces

4.3 Exercises

4.1 The infimum of almost upper bounds in the definition of $\|f\|_\infty$ is attained.

4.2 Fatou's lemma holds for $\mathbf{L}^\infty$: if $f_n \geq 0$, then $\|\liminf_n f_n\|_\infty \leq \liminf_n \|f_n\|_\infty$.

4.3 (Borel's theorem) Let X_n be a sequence of independently identically-distributed random variables with finite mean. Then $X_n/n \to 0$ almost surely.

[1] The support of a function $\psi \colon X \to \mathbb{R}$ is the closure of the set $\{x \in X : \psi(x) \neq 0\}$. "$\psi$ has compact support" means that the support of ψ is compact.

Chapter 5

Construction of measures

We now come to actually construct Lebesgue measure, as promised.

The construction takes a while and gets somewhat technical, and is not made much easier by making it less abstract. However, the pay-off is worth it, as at the end we will be able to construct many other measures besides Lebesgue measure.

5.1 Outer measures

The idea is to extend an existing set function μ, that has been only partially defined, to an "outer measure" μ^*.

The extension is remarkably simple and intuitive. It also "works" with pretty much any non-negative function μ, provided that it satisfies the trivial conditions in the following definition. Only later will we need to impose stronger conditions on μ, such as additivity.

5.1.1 Definition (Induced outer measure)

Let $\mathcal{A}$ be any family of subsets of a space X, and $\mu\colon \mathcal{A} \to [0, \infty]$ be a non-negative set function. Assume $\emptyset \in \mathcal{A}$ and $\mu(\emptyset) = 0$.

The **outer measure induced by** μ is the function:

$$\mu^*(E) = \inf\left\{ \sum_{n=1}^{\infty} \mu(A_n) \mid A_1, A_2, \ldots \in \mathcal{A} \text{ cover } E \right\},$$

defined for all sets $E \subseteq X$.

As our first step, we show that μ^* satisfies the following basic properties, which we generalize into a definition for convenience.

5.1.2 Definition (Outer measure)

A function $\theta\colon 2^X \to [0, \infty]$ is an **outer measure** if:

① $\theta(\emptyset) = 0$.

② *Monotonicity*: $\theta(E) \leq \theta(F)$ when $E \subseteq F \subseteq X$.

③ *Countable subadditivity*: If $E_1, E_2, \ldots \subseteq X$, then $\theta(\bigcup_n E_n) \le \sum_n \theta(E_n)$.

5.1.3 THEOREM (PROPERTIES OF INDUCED OUTER MEASURE)
The induced outer measure μ^* is an outer measure satisfying $\mu^*(A) \le \mu(A)$ for all $A \in \mathcal{A}$.

Proof Properties ① and ② for an outer measure are obviously satisfied for μ^*. That $\mu^*(A) \le \mu(A)$ for $A \in \mathcal{A}$ is also obvious.

Property ③ is proven by straightforward approximation arguments. Let $\varepsilon > 0$. For each E_n, by the definition of μ^*, there are sets $\{A_{n,m}\}_m \subseteq \mathcal{A}$ covering E_n, with

$$\sum_m \mu(A_{n,m}) \le \mu^*(E_n) + \frac{\varepsilon}{2^n}\,.$$

All of the sets $A_{n,m}$ together cover $\bigcup_n E_n$, so we have

$$\mu^*(\bigcup_n E_n) \le \sum_{n,m} \mu(A_{n,m}) = \sum_n \sum_m \mu(A_{n,m}) \le \sum_n \mu^*(E_n) + \varepsilon\,.$$

Since $\varepsilon > 0$ is arbitrary, we have $\mu^*(\bigcup_n E_n) \le \sum_n \mu^*(E_n)$. ∎

Our first goal is to look for situations where $\theta = \mu^*$ is countably additive, so that it becomes a bona fide measure. We suspect that we may have to restrict the domain of θ, and yet this domain has to be "large enough" and be a σ-algebra.

Fortunately, there is an abstract characterization of the "best" domain to take for θ, due to Constantin Carathéodory (it is a decided improvement over Lebesgue's original):

5.1.4 DEFINITION (MEASURABLE SETS OF OUTER MEASURE)
Let $\theta \colon 2^X \to [0, \infty]$ be an outer measure. Then

$$\mathcal{M} = \{B \in 2^X \mid \theta(B \cap E) + \theta(B^c \cap E) = \theta(E) \quad \text{for all } E \subseteq X\}$$

is the family of **measurable sets of the outer measure** θ.

Applying subadditivity of θ, the following definition is equivalent:

$$\mathcal{M} = \{B \in 2^X \mid \theta(B \cap E) + \theta(B^c \cap E) \le \theta(E) \quad \text{for all } E \subseteq X\}\,.$$

The intuitive meaning of the set predicate is that the set B is "sharply separated" from its complement. (Compare with the fact that a bounded set B is Jordan-measurable (i.e. $\mathbb{I}(B)$ is Riemann-integrable) if and only if its topological boundary has measure zero.)

The appellation "measurable set" is justified by the following result about $\mathcal{M}$:

5.1.5 THEOREM (MEASURABLE SETS OF OUTER MEASURE FORM σ-ALGEBRA)
$\mathcal{M}$ is a σ-algebra.

Proof It is immediate from the definition that $\mathcal{M}$ is closed under taking complements, and that $\emptyset \in \mathcal{M}$. We first show $\mathcal{M}$ is closed under finite intersection, and hence under finite union. Let $A, B \in \mathcal{M}$.

$$\begin{aligned}
&\theta\big((A \cap B) \cap E\big) + \theta\big((A \cap B)^{c} \cap E\big) \\
&= \theta(A \cap B \cap E) + \theta\big((A^{c} \cap B \cap E) \cup (A \cap B^{c} \cap E) \cup (A^{c} \cap B^{c} \cap E)\big) \\
&\leq \theta(A \cap B \cap E) + \theta(A^{c} \cap B \cap E) + \theta(A \cap B^{c} \cap E) + \theta(A^{c} \cap B^{c} \cap E) \\
&= \theta(B \cap E) + \theta(B^{c} \cap E)\,, \quad \text{by definition of } A \in \mathcal{M} \\
&= \theta(E)\,, \qquad\qquad\qquad\quad\ \text{by definition of } B \in \mathcal{M}\,.
\end{aligned}$$

Thus $A \cap B \in \mathcal{M}$.

We now have to show that if $B_1, B_2, \ldots \in \mathcal{M}$, then $\bigcup_n B_n \in \mathcal{M}$. We may assume that B_n are disjoint, for otherwise we can disjointify — consider instead $B'_n = B_n \setminus (B_1 \cup \cdots \cup B_{n-1})$, which are all in $\mathcal{M}$ as just shown.

For notational convenience, let:

$$D_N = \bigcup_{n=1}^{N} B_n \in \mathcal{M}\,, \quad D_N \nearrow D_\infty = \bigcup_{n=1}^{\infty} B_n\,.$$

We will need to know that:

$$\theta\Big(\bigcup_{n=1}^{N} B_n \cap E\Big) = \sum_{n=1}^{N} \theta(B_n \cap E)\,, \quad \text{for all } E \subseteq X, \text{ and } N = 1, 2, \ldots. \tag{❀}$$

This is a straightforward induction. The base case $N = 1$ is trivial. For $N > 1$,

$$\begin{aligned}
\theta(D_N \cap E) &= \theta\big(D_{N-1} \cap (D_N \cap E)\big) + \theta\big(D^{c}_{N-1} \cap (D_N \cap E)\big) \\
&= \theta(D_{N-1} \cap E) + \theta(B_N \cap E) \\
&= \sum_{n=1}^{N} \theta(B_n \cap E)\,, \quad \text{from induction hypothesis.}
\end{aligned}$$

Using equation ❀, we now have:

$$\begin{aligned}
\theta(E) &= \theta(D_N \cap E) + \theta(D^{c}_N \cap E) \\
&= \sum_{n=1}^{N} \theta(B_n \cap E) + \theta(D^{c}_N \cap E) \\
&\geq \sum_{n=1}^{N} \theta(B_n \cap E) + \theta(D^{c}_\infty \cap E)\,, \quad \text{from monotonicity.}
\end{aligned}$$

Taking $N \to \infty$, in conjunction with countable subadditivity, we obtain:

$$\theta(E) \geq \sum_{n=1}^{\infty} \theta(B_n \cap E) + \theta(D^{c}_\infty \cap E) \geq \theta(D_\infty \cap E) + \theta(D^{c}_\infty \cap E)\,.$$

But this implies $D_\infty = \bigcup_{n=1}^{\infty} B_n \in \mathcal{M}$. ∎

5.1.6 THEOREM (COUNTABLE ADDITIVITY OF OUTER MEASURE)
θ is countably additive on $\mathcal{M}$. That is, if $B_1, B_2, \ldots \in \mathcal{M}$ are disjoint, then $\theta(\bigcup_n B_n) = \sum_n \theta(B_n)$.

Proof The finite case $\theta(\bigcup_{n=1}^N B_n) = \sum_{n=1}^N \theta(B_n)$ has already been proven — just set $E = X$ in equation ❀.

By monotonicity, $\sum_{n=1}^N \theta(B_n) \leq \theta(\bigcup_{n=1}^\infty B_n)$. Taking $N \to \infty$ gives $\sum_{n=1}^\infty \theta(B_n) \leq \theta(\bigcup_{n=1}^\infty B_n)$. The inequality in the other direction is implied by subadditivity. ■

5.1.7 REMARK. More generally, the argument above shows that if a set function is finitely additive, is monotone, and is countably subadditive, then it is countably additive. We will be using this again later.

We summarize our work in this section:

5.1.8 COROLLARY (CARATHÉODORY'S THEOREM)
If θ is an outer measure (Definition 5.1.2), then the restriction of θ to its measurable sets $\mathcal{M}$ (Definition 5.1.4) yields a positive measure on $\mathcal{M}$.

5.2 Defining a measure by extension

Let μ be a set function that we want to somehow extend to a measure. In the last section, we have successfully extracted (Corollary 5.1.8), from the outer measure μ^*, a genuine measure $\mu^*|_{\mathcal{M}}$ by restricting to a certain family $\mathcal{M}$.

However, we do not yet know the relation between the original function μ and the new μ^*, nor do we know if the sets of interest actually do belong to $\mathcal{M}$.

In order for μ^* to be a sane extension of μ to $\mathcal{M}$, we will need to impose the following conditions, all of which are quite reasonable.

5.2.1 DEFINITION (PRE-MEASURE)
A **pre-measure** is a set function $\mu \colon \mathcal{A} \to [0, \infty]$ on a non-empty family $\mathcal{A}$ of subsets of X, satisfying the following conditions.

① The collection $\mathcal{A}$ should be *closed under finite set operations*. Any such collection $\mathcal{A}$ is termed an **algebra**.

② $\mu(\emptyset) = 0$.

③ We have *finite additivity* of μ on the algebra $\mathcal{A}$: if $B_1, \ldots, B_n \in \mathcal{A}$ are disjoint, then $\mu(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n \mu(B_i)$.

It follows that μ is *monotone* and *finitely subadditive*.

④ Also, if $B_1, B_2, \ldots \in \mathcal{A}$ are disjoint, *and* $\bigcup_{i=1}^{\infty} B_i$ happens to be in $\mathcal{A}$, then we must have *countable additivity* of μ there: $\mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i)$.

(If μ were not countably additive on $\mathcal{A}$, then it could not possibly be extended to a proper measure.)

These conditions are exactly just what we need to make the extension work out:

5.2.2 THEOREM (CARATHÉODORY EXTENSION PROCESS)
Given a pre-measure $\mu\colon \mathcal{A} \to [0,\infty]$, set μ^* to be its induced outer measure. Then every set in $A \in \mathcal{A}$ is measurable for μ^*, with $\mu^*(A) = \mu(A)$.

Proof Fix $A \in \mathcal{A}$. For any $E \subseteq X$ and $\varepsilon > 0$, by definition of μ^* (Definition 5.1.1), we can find $B_1, B_2, \ldots \in \mathcal{A}$ covering E such that $\sum_n \mu(B_n) \leq \mu^*(E) + \varepsilon$. Using the properties of induced outer measures (Theorem 5.1.3), we find

$$\begin{aligned}
\mu^*(A \cap E) + \mu^*(A^c \cap E) &\leq \mu^*\Big(A \cap \bigcup_n B_n\Big) + \mu^*\Big(A^c \cap \bigcup_n B_n\Big) \\
&\leq \sum_n \mu^*(A \cap B_n) + \sum_n \mu^*(A^c \cap B_n) \\
&\leq \sum_n \mu(A \cap B_n) + \sum_n \mu(A^c \cap B_n) \\
&= \sum_n \mu(B_n)\,, \quad \text{from finite additivity of } \mu, \\
&\leq \mu^*(E) + \varepsilon\,.
\end{aligned}$$

Taking $\varepsilon \searrow 0$, we see that A is measurable for μ^* (Definition 5.1.4).

Now we show $\mu^*(A) = \mu(A)$. Consider any cover of A by $B_1, B_2, \ldots \in \mathcal{A}$. By countable subadditivity and monotonicity of μ,

$$\mu(A) = \mu\Big(\bigcup_n A \cap B_n\Big) \leq \sum_n \mu(A \cap B_n) \leq \sum_n \mu(B_n)\,.$$

This implies $\mu(A) \leq \mu^*(A)$; the other inequality $\mu(A) \geq \mu^*(A)$ is automatic. ∎

Thus $\mu^*|_{\mathcal{M}}$ is a measure extending μ onto the σ-algebra $\mathcal{M}$ of measurable sets for μ^*. $\mathcal{M}$ must then also contain the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$, although $\mathcal{M}$ may be strictly larger than $\sigma(\mathcal{A})$.

Our final results for this section concern the uniqueness of this extension.

5.2.3 THEOREM (UNIQUENESS OF EXTENSION FOR FINITE MEASURE)
Let μ be a pre-measure on an algebra $\mathcal{A}$. Assume $\mu(X) < \infty$. If ν is another measure on $\sigma(\mathcal{A})$, that agrees with μ^* on $\mathcal{A}$, then μ^* and ν agree on $\sigma(\mathcal{A})$ as well.

Proof Let $B \in \sigma(\mathcal{A})$; it is measurable for μ^* by Theorem 5.2.2.

$$\mu^*(B) = \inf_{\substack{A_1, A_2, \ldots \in \mathcal{A} \\ B \subseteq \bigcup_n A_n}} \sum_n \mu(A_n) = \inf \sum_n \nu(A_n) \geq \inf \nu\Big(\bigcup_n A_n\Big) \geq \inf \nu(B)\,.$$

So $\mu^*(B) \geq \nu(B)$. Applying this inequality for B replaced by $X \setminus B$, we find

$$\mu^*(X) - \mu^*(B) = \mu^*(X \setminus B) \geq \nu(X \setminus B) = \nu(X) - \nu(B) = \mu^*(X) - \nu(B),$$

so $\mu^*(B) \leq \nu(B)$ after cancellation. ∎

But the hypothesis that μ has finite measure seems troublesome; it does not hold for Lebesgue measure ($X = \mathbb{R}^n$), for instance. An easy fix that works well in practice is:

5.2.4 DEFINITION (σ-FINITE MEASURES)
A measure space (X, μ) is **σ-finite**, if there are measurable sets $X_1, X_2, \ldots \subseteq X$, such that $\bigcup_n X_n = X$ and $\mu(X_n) < \infty$ for all n.

Clearly, we may as well assume that the sets X_n are increasing in this definition.

If μ is only a pre-measure defined on an algebra, then the sets X_n are assumed to be taken from that algebra.

5.2.5 EXAMPLE (LEBESGUE MEASURE IS σ-FINITE). Lebesgue measure is σ-finite. Take X_n to be open balls with radius n, for instance.

The definition of σ-finiteness is made to facilitate taking limits based on results in the finite case, as illustrated by the proof below.

5.2.6 THEOREM (UNIQUENESS OF EXTENSION FOR σ-FINITE MEASURE)
The conclusion of Theorem 5.2.3 also holds when (X, μ) is σ-finite.

Proof Let $X_n \nearrow X$, $\mu(X_n) < \infty$ as in the definition of σ-finiteness. For each $B \in \sigma(\mathcal{A})$, Theorem 5.2.3 says that $\mu^*(B \cap X_n) = \nu(B \cap X_n)$. Taking the limit $n \to \infty$ yields $\mu^*(B) = \nu(B)$. ∎

Finally, we can state the uniqueness result without referring to the induced outer measure. (Exercise 5.16 gives an alternate proof without intermediating through outer measures.)

5.2.7 COROLLARY (UNIQUENESS OF MEASURES)
If two measures μ and ν agree on an algebra $\mathcal{A}$, and the measure space is σ-finite (under μ or ν), then $\mu = \nu$ on the generated σ-algebra $\sigma(\mathcal{A})$.

5.3 Probability measure for infinite coin tosses

Now that we have Carathéodory's extension theorem in hand (Theorem 5.2.2), we can begin constructing some measures. However, there is some grunt work to do to verify that the hypotheses of Carathéodory's extension theorem are satisfied. In particular, proving that a pre-measure μ is *countably additive* on its algebra is not a mere triviality, as we will see.

Though our goal remains to construct Lebesgue measure on $\mathbb{R}^n$, in this section we tackle a simpler problem of constructing the probability measure for infinite coin tosses.

We want to construct a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, with random variables

$$X_n \colon \Omega \to \{0,1\}\,, \quad \text{for } n = 1, 2, \dots,$$

representing either heads or tails from a coin toss, such that each coin toss is fair and independent from each other:

$$\mathrm{P}(X_1 = x_1\,, \dots, X_k = x_k) = \mathrm{P}(X_1 = x_1) \cdots \mathrm{P}(X_k = x_k) = 2^{-k}\,, \quad x_n = 0 \text{ or } 1\,. \quad ❀$$

To the non-probabilists amongst us, this may seem to be a uninteresting detour, but in fact its solution can be used in turn to construct Lebesgue measure, and it will involve a prototype of the argument we will use later to construct Lebesgue measure more directly. (The impatient reader may skip directly to the next section.)

Definition of sample space. Let

$$\Omega = \prod_{n=1}^{\infty} \{0,1\} = \{0,1\}^{\mathbb{N}}$$

be the countably infinite product of the two-point set $\{0,1\}$. The random variables X_n are defined to be simply the coordinate functions on Ω (projections): $X_n(\omega) = \omega_n$.

Definition of algebra. A set (event) $A \subseteq \Omega$ to said to be a **cylinder set** if it is of the form $A = B \times \{0,1\}^{\mathbb{N}}$ where $B \subseteq \{0,1\}^k$ for some k. In words, we are allowed the stipulate the results of a *finite* number k of coin tosses, but the results of all the tosses after the k^{th} one remain unknown (they can be heads or tails).

The algebra $\mathcal{A}$ is defined to consist of all cylinder sets. It is elementary that $\mathcal{A}$ is closed under finite union, intersection, and complement. For example, if A is represented as $B \times \{0,1\}^{\mathbb{N}}$ as above, then $A^{\mathrm{c}} = B^{\mathrm{c}} \times \{0,1\}^{\mathbb{N}}$.

Definition of pre-measure. Also, define $\mathrm{P}(A)$ to be the number of elements in B divided by 2^k, i.e. the probability of any $\omega \in A$ occurring as in equation ❀. Elementary counting shows that $\mathrm{P}(A)$ is well-defined, and is *finitely additive* on $\mathcal{A}$.

Countable additivity of pre-measure. Now we come to the big question: is P countably additive on $\mathcal{A}$? It may come as a pleasant surprise, but in actuality there are no genuine disjoint countable unions in $\mathcal{A}$ — all countable unions collapse to finite unions, so countable additivity is automatic.

We use compactness. Begin by putting the discrete topology on $\{0,1\}$. The singleton points $\{0\}$ and $\{1\}$ are open.

Observe that every cylinder set is some finite union of the intersections:

$$\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}, \quad n = 1, 2, \ldots .$$

which so happen to form the *basis* for the *product topology* on the sample space Ω. Thus every cylinder set is *open in the product topology*. Since $\mathcal{A}$ is closed under complements, every cylinder set is also *closed*.

The component spaces $\{0, 1\}$ are compact as each has only four open sets in total. By Tychonoff's theorem, the sample space Ω under the product topology is also *compact*.

If $E = \bigcup_n A_n \in \mathcal{A}$ for $A_1, A_2, \ldots \in \mathcal{A}$, this means $\{A_n\}$ forms an open cover of E. The set E is compact because it is a closed set of Ω; hence there is a finite sub-cover which gives the finite union.

Now all we have to do is apply Carathédory's extension theorem, to obtain a measure P that satisfies ❀, defined on some σ-algebra $\mathcal{F}$.

Relation to Lebesgue measure. The coin-toss measure can be transported to the Lebesgue measure on $[0, 1]$. There is the most obvious way of corresponding sequences of binary digits ω to numbers:

$$T\omega = 0.\omega_1\omega_2\omega_3 \cdots = \sum_{n=1}^{\infty} \frac{X_n(\omega)}{2^n} .$$

This mapping $T\colon \Omega \to [0, 1]$ can be shown to be measurable. Then for any Borel set $E \subseteq [0, 1]$, we can take its *Lebesgue measure* to be

$$\lambda(E) = \mathrm{P}(\{\omega \in \Omega \mid T\omega \in E\}) = \mathrm{P}(T^{-1}(E)) .$$

That this is Lebesgue measure can be illustrated by example. If $E = [\frac{1}{8}, \frac{3}{4}]$, then those ω that correspond to elements of E begin with one of the following prefixes:

$$\begin{gathered} 0.\,001\,\omega_4\omega_5\omega_6 \cdots \\ 0.\,010\,\omega_4\omega_5\omega_6 \cdots \\ 0.\,011\,\omega_4\omega_5\omega_6 \cdots \\ 0.\,100\,\omega_4\omega_5\omega_6 \cdots \\ 0.\,101\,\omega_4\omega_5\omega_6 \cdots \end{gathered}$$

(The number $\frac{3}{4} = 0.110$ is included in this list because $0.101\overline{111} \cdots = 0.110\overline{000} \cdots$.) But the probability of these ω occurring is $5/2^3 = \frac{5}{8} = \frac{3}{4} - \frac{1}{8} = \lambda(E)$, just as we expect.

The details of generalizing this example to all Borel sets E are left to you (Exercise 5.5).

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

In this section we construct the n-dimensional volume measure on $\mathbb{R}^n$. As hinted before, the idea is to define volume for rectangles and then extend with Carathéodory's theorem (Theorem 5.2.2).

Our setting will be the collection $\mathcal{R}$ of rectangles $I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$, where I_k is any open, half-open or closed, bounded or unbounded, interval in $\mathbb{R}$. The following definition is merely an abstracted version of the formal facts we need about these rectangles.

5.4.1 DEFINITION (SEMI-ALGEBRA)

Let X be any set. A **semi-algebra** is any $\mathcal{R} \subseteq 2^X$ with the following properties:

① The empty set is in $\mathcal{R}$.

② *Closure under finite intersection*: The intersection of any two sets in $\mathcal{R}$ is in $\mathcal{R}$.

③ The complement of any set in $\mathcal{R}$ is expressible as a *finite disjoint* union of other sets in $\mathcal{R}$.

It is graphically evident that the collection of all rectangles is indeed a semi-algebra. Translating this to an algebraic proof is easy:

5.4.2 THEOREM (SEMI-ALGEBRA OF RECTANGLES)

The set of rectangles form a semi-algebra.

Proof Property ① for a semi-algebra is trivial. For property ②, observe:

$$A = I_1 \times \cdots \times I_n\,, \quad B = J_1 \times \cdots \times J_n \implies A \cap B = (I_1 \cap J_1) \times \cdots \times (I_n \cap J_n)\,.$$

Property ③ follows by induction on dimension, expressing

$$(A \times B)^{\mathrm{c}} = (A^{\mathrm{c}} \times B) \cup (A \times B^{\mathrm{c}}) \cup (A^{\mathrm{c}} \times B^{\mathrm{c}})$$

as a finite disjoint union at each step. ∎

From a semi-algebra, we can automatically construct an algebra $\mathcal{A}$:

5.4.3 THEOREM

The family $\mathcal{A}$ of all finite disjoint unions of elements of a semi-algebra $\mathcal{R}$ is an algebra.

Proof We check the properties for an algebra. Let $A, B \in \mathcal{A}$ be finite disjoint unions of $R_i \in \mathcal{R}$ and $S_j \in \mathcal{R}$ respectively.

① The empty set is trivially in $\mathcal{A}$.

② $A \cap B = \left(\bigcup_i R_i\right) \cap \left(\bigcup_j S_j\right) = \bigcup_{i,j} R_i \cap S_j$, and the final union remains disjoint — so $\mathcal{A}$ is closed under finite intersection.

③ $A^c = \left(\bigcup_i R_i\right)^c = \bigcap_i R_i^c$, and each R_i^c is a finite disjoint union of rectangles by definition of a semi-algebra. The outer finite intersection belongs to $\mathcal{A}$ by step ②, so $\mathcal{A}$ is closed under taking complements.

④ Closure under finite intersection and complement implies closure under finite union. ■

By the way, the σ-algebra generated by $\mathcal{A}$ will contain the Borel σ-algebra: every open set $U \in \mathbb{R}^n$ obviously can be written as a union of open rectangles, and in fact a *countable* union of open rectangles. For, given an arbitrary collection of rectangles covering $U \subseteq \mathbb{R}^n$, there always exists a *countable* subcover (Theorem A.4.1).

Having specified the domain, we construct Lebesgue measure.

Volume of rectangle. The volume of $A = I_1 \times \cdots \times I_n$ is naturally defined as $\lambda(A) = \lambda(I_1) \cdot \cdots \cdot \lambda(I_n)$, with $\lambda(I_k)$ being the length of the interval I_k. The usual rules about multiplying zeroes and infinities will be in force (Remark 2.4.6).

Additivity on $\mathcal{R}$. Suppose that the rectangle A has been partitioned into a finite number of smaller disjoint rectangles B_k. Then $\sum_k \lambda(B_k)$, as we have defined it, should equal $\lambda(A)$.

For example, in the case of one dimension, a partition always looks like:

$$[a_0, a_m] = [a_0, a_1) \cup [a_1, a_2) \cup \cdots \cup [a_{m-1}, a_m] \,,$$

so the sum of lengths telescopes:

$$\lambda[a_0, a_m] = (a_1 - a_0) + (a_2 - a_1) + \cdots + (a_m - a_{m-1}) = a_m - a_0 \,.$$

The intervals here may be changed so that the left-hand side becomes open instead of the right, etc., without affecting the result.

For higher dimensions, we resort to an inductive argument. If we write the sets $A = A^1 \times \cdots \times A^n$ and $B_k = B_k^1 \times \cdots \times B_k^n$ as products of intervals, then

$$\begin{aligned} \mathbb{I}(x \in A) &= \sum_k \mathbb{I}(x \in B_k) \\ \mathbb{I}(x^1 \in A^1) \cdots \mathbb{I}(x^n \in A^n) &= \sum_k \mathbb{I}(x^1 \in B_k^1) \cdots \mathbb{I}(x^n \in B_k^n) \,. \end{aligned}$$

Let us fix the variables $x^2, \dots, x^n$; then both sides are simple functions of the variable x^1. We can *integrate* both sides with respect to x^1. Although λ in one dimension is not yet known to be countably additive, taking integrals of simple functions is okay because the relevant properties depend only on finite additivity — see Remark 2.4.7 and Lemma 2.4.8.

$$\begin{aligned} \int_{x^1 \in \mathbb{R}} \mathbb{I}(x^1 \in A^1) \cdots \mathbb{I}(x^n \in A^n) &= \sum_k \int_{x^1 \in \mathbb{R}} \mathbb{I}(x^1 \in B_k^1) \cdots \mathbb{I}(x^n \in B_k^n) \\ \lambda(A^1)\, \mathbb{I}(x^2 \in A^2) \cdots \mathbb{I}(x^n \in A^n) &= \sum_k \lambda(B_k^1)\, \mathbb{I}(x^2 \in B_k^1) \cdots \mathbb{I}(x^n \in B_k^n) \,. \end{aligned}$$

Repeating this $n-1$ more times thus shows $\lambda(A) = \sum_k \lambda(B_k)$.

Volume on $\mathcal{A}$. The volume of a disjoint union of rectangles is obviously defined as the sum of the volumes of the component rectangles. This volume is well-defined: if we have different decompositions, $\bigcup_i R_i = \bigcup_j S_j$, then taking the common refinement $R_i \cap S_j$, we have $\sum_i \lambda(R_i) = \sum_i \sum_j \lambda(R_i \cap S_j)$ by additivity on each individual rectangle R_i. But $\sum_j \lambda(S_j)$ equals this double sum also.

Countable additivity. It is clear that λ is finitely additive on $\mathcal{A}$; we now prove countable additivity.

According to Remark 5.1.7, it suffices to prove countable subadditivity in place of plain additivity. Thus we suppose that $C \in \mathcal{A}$, and $C \subseteq \bigcup_{k=1}^{\infty} B_k$ for $B_k \in \mathcal{A}$.

If we assume C is compact, and B_k are open, then we can use compactness to obtain a finite sub-cover: $C \subseteq \bigcup_{k=1}^{m} B_k$ for some finite m. Then countable subadditivity follows from finite subadditivity:

$$\lambda(C) \leq \sum_{k=1}^{m} \lambda(B_k) \leq \sum_{k=1}^{\infty} \lambda(B_k)\,.$$

Suppose next that C is a bounded set in $\mathbb{R}^n$. Then its closure $\overline{C}$ is compact, and it evidently belongs to $\mathcal{A}$ too. The sets B_k may not be open, and they may not cover $\overline{C}$, but each set B_k can be slightly expanded and made open, in such a way that the volumes $\lambda(B_k)$ increase by at most $\varepsilon/2^k$, and the new B_k now cover $\overline{C}$. The argument in the previous paragraph applies:

$$\lambda(C) \leq \lambda(\overline{C}) \leq \sum_{k=1}^{\infty} \lambda(B_k) + \varepsilon\,, \text{ then take } \varepsilon \searrow 0.$$

Finally, consider a set C that is unbounded. But countable subadditivity applies to the bounded set $C \cap [-a,a]^n \in \mathcal{A}$:

$$\lambda(C \cap [-a,a]^n) \leq \sum_{k=1}^{\infty} \lambda(B_k)\,.$$

Direct computation shows that $\lim\limits_{a\to\infty} \lambda(C \cap [-a,a]^n) = \lambda(C)$ for $C \in \mathcal{A}$.

Applying Carathéodory's theorem we conclude:

5.4.4 THEOREM

Lebesgue measure in $\mathbb{R}^n$ exists, and it is uniquely determined by the assignment of volumes of rectangles. It is translation-invariant.

Invariance under translations is readily verified by looking at formula for the induced outer measure (Definition 5.1.1); or it can be proven axiomatically.

It is intuitive that Lebesgue measure should also be invariant under other rigid motions such as rotations and reflections. As our construction of Lebesgue measure is heavily coordinate-dependent (it "hard-codes" the standard basis of $\mathbb{R}^n$), invariance under rigid motion is hardly obvious, but it will be a consequence of Lemma 7.2.1.

5.5 Existence of non-measurable sets

Are there any sets that are not Borel, or that cannot be assigned any volume?

The following theorem gives a classic example, and should also serve to convince you why our strenuous efforts are necessary.

5.5.1 THEOREM (VITALI'S NON-MEASURABLE SET)
There exists a set in $[0,1]$ that is not Lebesgue-measurable. In other words, Lebesgue measure cannot be defined consistently for *all* subsets of $[0,1]$.

Proof The key premise in this proof is translation-invariance. In particular, given any measurable $H \subseteq [0,1]$, define its "shift with wrap-around":

$$H \oplus x = \{h + x : h \in H, h + x \leq 1\} \cup \{h + x - 1 : h \in H, h + x > 1\} .$$

Then $\lambda(H \oplus x) = \lambda(H)$.

Let $x \sim y$ for two real numbers x and y if $x - y$ is rational. The interval $[0,1]$ is partitioned by this equivalence relation. Compose a set $H \subset [0,1]$ by picking exactly one element from each equivalence class, and also say $0 \notin H$. Then $(0,1]$ equals the disjoint union of all $H \oplus r$, for $r \in [0,1) \cap \mathbb{Q}$. Countable additivity leads to:

$$1 = \lambda((0,1]) = \sum_{r \in [0,1) \cap \mathbb{Q}} \lambda(H \oplus r) = \sum_{r \in [0,1) \cap \mathbb{Q}} \lambda(H) ,$$

a contradiction, because the sum on the right can only be 0 or ∞. Hence H cannot be measurable. ■

It has become obligatory to alert that the set H in the proof of Vitali's theorem comes from the *axiom of choice*. Robert Solovay has shown in 1970, essentially, that if the axiom of choice is not available, then every set of real numbers is Lebesgue-measurable. Since not having the axiom of choice would severely cripple the field of analysis anyway, I prefer not to make a big fuss[†] about it.

The famous *Hausdorff paradox* and the *Banach-Tarski paradox* are related to Vitali's theorem; they all use the axiom of choice to dig up some pretty strange sets in $\mathbb{R}^3$.

5.6 Completeness of measures

Here we settle some technical points that were glossed over before.

Up to this point, whenever we discussed Lebesgue measure λ on $\mathbb{R}^n$, we have assumed that the domain of λ is $\mathscr{B}(\mathbb{R}^n)$. Certainly λ can be defined on $\mathscr{B}(\mathbb{R}^n)$, but as we had remarked, the σ-algebra $\mathcal{M}$ that comes out of the Carathéodory extension process may be bigger than $\mathscr{B}(\mathbb{R}^n)$.

[†] For example, even standard calculus texts blithely assert the equivalence of continuity for real functions and sequential continuity, without noting the proof requires the axiom of choice, or certain weaker forms of it.

The mystery is resolved by this observation: if θ is an outer measure (Definition 5.1.2), and $\theta(B) = 0$, then $\theta(A) = 0$ for any subset $A \subseteq B$. If, furthermore, $B \in \mathcal{M}$, then we see in Definition 5.1.4 that $A \in \mathcal{M}$ automatically for *any* $A \subseteq B$. We give a name to this phenomenon:

5.6.1 Definition (Completeness of measure)
A measure μ is **complete** if given any set of B of μ-measure zero, *every* subset of B is μ-measurable (and necessarily must have measure zero).

Thus, λ defined on the σ-algebra $\mathcal{M}$ is *complete*.

On the other hand, it seems a bit tad unlikely that λ restricted to $\mathscr{B}(\mathbb{R}^n)$ is complete, seeing that the definition of $\mathscr{B}(\mathbb{R}^n)$ does not even involve measure functions. The proof will be left to you.

Knowing that a measure is complete is sometimes convenient. If a sequence of measurable functions f_n converges to some arbitrary function f pointwise *almost everywhere*, it does not follow that f is measurable unless the measure space is complete. (See also Remark 3.1.6.)

5.6.2 Theorem
Suppose the measure space X is complete.

① If $f\colon X \to Y$ is a measurable function and $f = g$ almost everywhere, then g is measurable.

② If $f_n\colon X \to \overline{\mathbb{R}}$ are measurable functions that converge to another function g almost everywhere, then g is measurable.

Proof Let $f(x) = g(x)$ for $x \in A$, where A^c has measure zero. For any measurable $E \subseteq Y$,

$$g^{-1}(E) = \big(g^{-1}(E) \cap A\big) \cup B = \big(f^{-1}(E) \cap A\big) \cup B\,, \quad B = g^{-1}(E) \cap A^c\,,$$

and B is measurable since $B \subseteq A^c$ under a complete measure space. Thus $g^{-1}(E)$ is measurable, and this proves part ①. For part ②, set $f = \limsup_n f_n$. ∎

Nevertheless, if a measure space is not complete, for integration purposes we ought to be able to assume g (in Theorem 5.6.2) is measurable anyway, since the "bad" sets "really" have measure zero anyway. Formally, it can be done by this procedure:

5.6.3 Theorem
Let $(X, \mathcal{M}, \mu)$ be a measure space. Define

$$\overline{\mathcal{M}} = \{A \cup N \mid A \in \mathcal{M} \text{ and } N \subseteq B \text{ for some } B \in \mathcal{M} \text{ with } \mu(B) = 0\}\,,$$

and define $\overline{\mu}(A \cup N) = \mu(A)$ for any $E \in \overline{\mathcal{M}}$ represented as $E = A \cup N$ as above.

Then $(X, \overline{\mathcal{M}}, \overline{\mu})$ is a complete measure space, called the **completion** of $(X, \mathcal{M}, \mu)$; and $\overline{\mu}$ is the unique extension of μ onto $\overline{\mathcal{M}}$.

5.6.4 THEOREM
Let $(X, \mathcal{M}, \mu)$ be a measure space, and $(X, \overline{\mathcal{M}}, \overline{\mu})$ be its completion. If $f\colon X \to \mathbb{R}$ is $\overline{\mathcal{M}}$-measurable, then there is a $\mathcal{M}$-measurable function $g\colon X \to \mathbb{R}$ that equals f almost everywhere.

Another way to get at the completion of a measure is to plug it in the Carathédory extension procedure:

5.6.5 THEOREM
Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. Then the measure space $(X, \mathcal{M}, \mu^*|_{\mathcal{M}})$ obtained from the Carathéodory extension is exactly the completion of $(X, \mathcal{A}, \mu)$.

The Lebesgue measure λ is usually considered to be defined on $\mathcal{L} = \overline{\mathscr{B}(\mathbb{R}^n)}$, though in practical applications the difference between $\mathcal{L}$ and $\mathscr{B}(\mathbb{R}^n)$ is almost hair-splitting. See also Remark 2.3.6. Sets in $\mathcal{L}$ are called **Lebesgue-measurable**.

5.7 Exercises

5.1 (Characterization of Lebesgue measure zero) Elementary textbooks that talk about measure zero without measure theory all use the following definition. A set $A \in \mathbb{R}^n$ has **(Lebesgue) measure zero** if and only if,

☞ for every $\varepsilon > 0$, the set A can be covered by a countable number of open (or closed) rectangles whose total volume is $< \varepsilon$.

Show how this follows from our definition of Lebesgue measure.

5.2 (Proof of finite additivity for Lebesgue measure) The short proof of the finite additivity of the Lebesgue pre-measure λ given in section 5.4 seems almost miraculous. What is actually happening in that proof?

It is not hard to see intuitively that finite additivity for rectangles basically only involves the fundamental properties of real numbers such as commutativity of addition, additive cancellation, and distributivity of multiplication over addition. But if you had wanted to explicitly write down a summation, that when its terms are rearranged, would show finite additivity, what is the partition on the rectangles you would use?

5.3 (Probabilities of coin tosses) Using the probability measure P appearing in section 5.3, compute the probabilities that:

① all the coin tosses come up heads.

② every other coin toss come up heads.

③ there is eventually a repeating pattern of fixed period in the coin toss sequence.

④ infinitely many heads come up.

⑤ the first head that comes up must be followed by another head.

⑥ the series $\sum_{n=1}^{\infty} \frac{(-1)^{X_n}}{n}$ converges, where X_n are the coin-toss outcomes represented as 0 or 1.

5.4 (Homeomorphism of coin-toss space to unit interval) Verify the claim in section 5.3, that the binary-expansion mapping $T\colon \{0,1\}^{\mathbb{N}} \to [0,1]$ is measurable.

In fact, it is a continuous mapping. If we identify those sequences that are binary expansions of the same number (those that end with $0\overline{111}\cdots$ and the corresponding one ending with $1\overline{000}\cdots$), then the quotient space of $\{0,1\}^{\mathbb{N}}$ is homeomorphic to $[0,1]$ via T.

5.5 (From coin tosses to Lebesgue measure) Show that the measure λ transported from the coin-toss measure (section 5.3) is a realization of the Lebesgue measure on $[0,1]$. Also, extend this λ so that it covers the whole real line.

5.6 (From Lebesgue measure to coin tosses) Do the reverse transformation: construct the coin-toss measure using Lebesgue measure on $[0,1]$ as the raw material — that is, without invoking Carathéodory's extension theorem separately again.

5.7 (Infinite product of Lebesgue measures) Construct a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and random variables $X_n\colon \Omega \to [0,1]$ on that space, such that

$$\mathrm{P}(X_1 \in E_1, \dots, X_n \in E_n) = \mathrm{P}(X_1 \in E_1) \cdots \mathrm{P}(X_n \in E_n) = \lambda(E_1) \cdots \lambda(E_n)$$

for any n, where λ is Lebesgue measure on $[0,1]$. Schematically, $\mathrm{P} = \lambda \otimes \lambda \otimes \cdots$, the analogue of the finite product studied in chapter 6.

Interpretation: the random variables X_n are independent and identically distributed, with the uniform distribution on $[0,1]$. This result can be used, in general, to construct a probability space containing a countable number independent random variables, with *any* distributions.

Hint: Take $\Omega = [0,1]^{\mathbb{N}}$. Alternatively, take $\Omega = \{0,1\}^{\mathbb{N}}$, the coin-toss space.

5.8 (Measurable sets for coin tosses) Identify the σ-algebra $\mathcal{F}$ constructed for the coin-toss measure in section 5.3.

5.9 (Borel's normal number theorem) The binary digit space $\{0,1\}$ used for the coin-toss probability space can obviously be replaced by $\{0,1,\dots,\beta-1\}$ for any digit base β.

A number $x \in [0,1]$ is *normal in base* β if every digit occurs with equal frequency in the base-β expansion of x:

$$\lim_{n\to\infty} \frac{\text{number of the first } n \text{ digits of } x \text{ that equal k}}{n} = \frac{1}{\beta}, \quad k = 0,1,\dots,\beta-1\,.$$

Show that almost every number in $[0,1]$ (under Lebesgue measure) is normal for any base.

5.10 (Shift invariance) Consider the coin-toss measure P on $\Omega = \{0,1\}^{\mathbb{N}}$ from section 5.3. Define the left-shift operator

$$L(\omega_1, \omega_2, \dots) = (\omega_2, \omega_3, \dots)\,.$$

Then $L\colon \Omega \to \Omega$ is measurable, and $\mathrm{P}(L^{-1}E) = \mathrm{P}(E)$ for all events $E \subseteq \Omega$.

5.11 (Translation invariance) Prove translation invariance of Lebesgue measure axiomatically.

5.12 (Cantor set) Show that the Cantor set has Lebesgue-measure zero.

5.13 (Lebesgue-Cantor function)

5.14 (σ-finiteness hypothesis for uniqueness) Is the σ-finiteness hypothesis for the uniqueness of measures necessary?

5.15 (Approximation by algebra) Demonstrate the following without outer-measure technology.

Let μ be a finite pre-measure defined on an algebra $\mathcal{A}$, and λ be some extension of μ to $\sigma(\mathcal{A})$. For every $E \in \sigma(\mathcal{A})$ and $\varepsilon > 0$, there exists $A \in \mathcal{A}$ such that $\lambda(A \Delta E) < \varepsilon$.

Phrased in another way, every set in $\sigma(\mathcal{A})$ can be approximated by the generating sets in $\mathcal{A}$, and the approximation error, measured using λ, can be made as small as we like.

Hint: If I gave you an explicit representation of a set in $E \in \sigma(\mathcal{A})$ as a countable union of sets in $\mathcal{A}$, how would you actually go about finding an approximation to E given some error tolerance ε?

5.16 (Uniqueness of measures) Use Exercise 5.15 to prove uniqueness of the extension of a σ-finite pre-measure defined on an algebra.

Chapter 6

Integration on product spaces

This chapter demonstrates the ever-useful result that integrals over higher dimensions (so-called "multiple integrals") can be evaluated in terms of iterated integrals over lower dimensions. For example,

$$\int_{[0,a]\times[0,b]} f(x,y)\,dx\,dy = \int_0^b \Big[\int_0^a f(x,y)\,dx\Big]dy = \int_0^a \Big[\int_0^b f(x,y)\,dy\Big]dx\,.$$

Compared to the machinations required for the analogous result using Riemann integrals, the Lebesgue version is conceptually quite simple. The hard part is showing all the stuff involved to be measurable.

6.1 Product measurable spaces

Since a two-variable function is naturally seen as a function over a product space, we begin by studying such spaces.

6.1.1 Definition (Measurable rectangle)
Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces. A **measurable rectangle** in $X \times Y$ is a set of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

The σ-algebra generated by all the measurable rectangles $A \times B$ is the **product σ-algebra** of $\mathcal{A}$ and $\mathcal{B}$, and is denoted by $\mathcal{A} \otimes \mathcal{B}$. This will be the σ-algebra we use for $X \times Y$.

6.1.2 Definition (Cross sections)
Let E be a measurable set from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$. Its **cross sections** in Y and X are:

$$E^x = \{y \in Y : (x,y) \in E\}\,, \quad x \in X\,.$$
$$E^y = \{x \in X : (x,y) \in E\}\,, \quad y \in Y\,.$$

6.1.3 THEOREM (MEASURABILITY OF CROSS SECTION)
If $E \in \mathcal{A} \otimes \mathcal{B}$, then $E^y \in \mathcal{A}$, and $E^x \in \mathcal{B}$.

Proof We prove the theorem for E^y; the proof for E^x is the same.
Consider the collection $\mathcal{M} = \{E \in \mathcal{A} \otimes \mathcal{B} : E^y \in \mathcal{A}\}$. We show that $\mathcal{M}$ is a σ-algebra containing the measurable rectangles; then it must be all of $\mathcal{A} \otimes \mathcal{B}$, and this would establish the theorem.

① Suppose $E = A \times B$ is a measurable rectangle. Then $E^y = A$ when $y \in \mathcal{B}$, otherwise $E^y = \varnothing$. In both cases $E^y \in \mathcal{A}$, so $E \in \mathcal{M}$.

② $\varnothing^y = \{x \in X : (x,y) \in \varnothing\} = \varnothing \in \mathcal{A}$, so $\mathcal{M}$ contains $\varnothing$.

③ If $E_1, E_2, \ldots \in \mathcal{M}$, then $(\bigcup E_n)^y = \bigcup E_n^y \in \mathcal{A}$. So $\mathcal{M}$ is closed under countable union.

④ If $E \in \mathcal{M}$, and $F = E^c$, then $F^y = \{x \in X : (x,y) \notin E\} = X \setminus E^y \in \mathcal{A}$. So $\mathcal{M}$ is closed under complementation. ■

6.1.4 THEOREM (MEASURABILITY OF FUNCTION WITH FIXED VARIABLE)
Let $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ and $(Z, \mathcal{M})$ be measurable spaces.
If $f \colon X \times Y \to Z$ is measurable, then the functions $f_y \colon X \to Z$, $f_x \colon Y \to Z$ obtained by holding one variable fixed are also measurable.

Proof Again we consider only f_y. Let $I_y \colon X \to X \times Y$ be the inclusion mapping $I_y(x) = (x,y)$. Given $E \in \mathcal{A} \otimes \mathcal{B}$, by Theorem 6.1.3 $I_y^{-1}(E) = E^y \in \mathcal{A}$, so I_y is a measurable function. But $f_y = f \circ I_y$. ■

If X and Y are topological spaces, such as $\mathbb{R}^n$ and $\mathbb{R}^m$, then they come with their own product topology. Thus we have two structures that are relevant:

$$\mathscr{B}(X \times Y) \text{ and } \mathscr{B}(X) \otimes \mathscr{B}(Y)\,.$$

In general, it seems that $\mathscr{B}(X \times Y)$ should be the bigger of the two structures. For the rectangles $U \times V$, for open sets $U \subseteq X$ and $V \subseteq Y$, form a basis for the topology of $X \times Y$, but topologies allow arbitrary unions while the σ-algebras only allow countable unions. However, if arbitrary unions always reduce to countable ones, the two structures should be equal.

6.1.5 LEMMA (GENERATING SETS OF PRODUCT σ-ALGEBRA)
Let $X \in \mathcal{A}' \subseteq 2^X$ and $Y \in \mathcal{B}' \subseteq 2^Y$. If $\mathcal{A} = \sigma(\mathcal{A}')$ and $\mathcal{B} = \sigma(\mathcal{B}')$, then $\mathcal{A} \otimes \mathcal{B} = \sigma(\{U \times V \mid U \in \mathcal{A}', V \in \mathcal{B}'\})$.

Proof Clearly,

$$\mathcal{M} = \sigma(\{U \times V \mid U \in \mathcal{A}', V \in \mathcal{B}'\}) \subseteq \sigma(\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}) = \mathcal{A} \otimes \mathcal{B}.$$

Next, consider the family $\{A \subseteq X \mid A \times Y \in \mathcal{M}\}$. It is evidently a σ-algebra that contains $\mathcal{A}'$ and hence $\mathcal{A}$. In other words, $A \times Y \in \mathcal{M}$ for all $A \in \mathcal{A}$.

Switching the roles of X and Y and repeating the argument, we have $X \times B \in \mathcal{M}$ for $B \in \mathcal{B}$. Thus $(A \times Y) \cap (X \times B) = A \times B$ is in $\mathcal{M}$ too, and this shows $\mathcal{M} \supseteq \mathcal{A} \otimes \mathcal{B}$. ∎

6.1.6 THEOREM (PRODUCT BOREL σ-ALGEBRA)
If X and Y are second-countable topological spaces (e.g. $\mathbb{R}^n$), then $\mathscr{B}(X \times Y) = \mathscr{B}(X) \otimes \mathscr{B}(Y)$.

Proof Let $\{U_i\}$ and $\{V_j\}$ be countable bases for the topological spaces X and Y respectively. Then $\{U_i \times V_j\}$ is a countable basis for the product topological space $X \times Y$. Then $\mathscr{B}(X \times Y) \subseteq \mathscr{B}(X) \otimes \mathscr{B}(Y)$ since $U_i \times V_j \in \mathscr{B}(X) \otimes \mathscr{B}(Y)$.

On the other hand, by Lemma 6.1.5,

$$\mathscr{B}(X) \otimes \mathscr{B}(Y) = \sigma(\{U \times V \mid U \subseteq X, V \subseteq Y \text{ open}\}) \subseteq \mathscr{B}(X \times Y).$$ ∎

6.2 Cross-sectional areas

We would like to take the cross-sections defined in Definition 6.1.2, and calculate their areas à la Cavalieri's principle. So we want to have a theorem like Theorem 6.2.3 below.

Its proof requires the following technical tool, which comes equipped with a definition.

6.2.1 DEFINITION (MONOTONE CLASS)
A family $\mathcal{A}$ of subsets of X is a **monotone class** if it is closed under increasing unions and decreasing intersections.

The intersection of any set of monotone classes is a monotone class. The smallest monotone class containing a given set $\mathcal{G}$ is the intersection of all monotone classes containing $\mathcal{G}$. This construction is analogous to the one for σ-algebras, and the result is also said to be the **monotone class generated by** $\mathcal{G}$.

6.2.2 THEOREM (MONOTONE CLASS THEOREM)
If $\mathcal{A}$ be an algebra on X, then the monotone class generated by $\mathcal{A}$ coincides with the σ-algebra generated by $\mathcal{A}$.

Proof Since a σ-algebra is a monotone class, the generated σ-algebra contains the generated monotone class $\mathcal{M}$. So we only need to show $\mathcal{M}$ is a σ-algebra.

We first claim that $\mathcal{M}$ is actually closed under complementation. Let $\mathcal{M}' = \{S \in \mathcal{M} : X \setminus S \in \mathcal{M}\} \subseteq \mathcal{M}$. This is a monotone class, and it contains the algebra $\mathcal{A}$. So $\mathcal{M} = \mathcal{M}'$ as desired.

To prove that $\mathcal{M}$ is closed under countable unions, we only need to prove that it is closed under finite unions, for it is already closed under countable *increasing* unions.

First let $A \in \mathcal{A}$, and $\mathcal{N}(A) = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\} \subseteq \mathcal{M}$. Again this is a monotone class containing the algebra $\mathcal{A}$; thus $\mathcal{N}(A) = \mathcal{M}$.

Finally, let $S \in \mathcal{M}$, with the same definition of $\mathcal{N}(S)$. The last paragraph, rephrased, says that $\mathcal{A} \subseteq \mathcal{N}(S)$. And $\mathcal{N}(S)$ is a monotone class containing $\mathcal{A}$ by the same arguments as the last paragraph. Thus $\mathcal{N}(S) = \mathcal{M}$, proving the claim. ∎

6.2.3 Theorem (Measurability of cross-sectional area function)

Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces (Definition 5.2.4). If $E \in \mathcal{A} \otimes \mathcal{B}$,

① $\nu(E^x)$ is a measurable function of $x \in X$.

② $\mu(E^y)$ is a measurable function of $y \in Y$.

Proof We concentrate on $\mu(E^y)$. Assume first that $\mu(X) < \infty$. Let

$$\mathcal{M} = \{E \in \mathcal{A} \otimes \mathcal{B} : \mu(E_y) \text{ is a measurable function of } y\}\,.$$

$\mathcal{M}$ is equal to $\mathcal{A} \otimes \mathcal{B}$, because:

① If $E = A \times B$ is a measurable rectangle, then $\mu(E^y) = \mu(A)\mathbb{I}(y \in B)$ which is a measurable function of y. If E is a finite disjoint union of measurable rectangles E_n, then $\mu(E^y) = \sum_n \mu(E_n^y)$ which is also measurable.

The measurable rectangles form a semi-algebra, analogous to the rectangles in $\mathbb{R}^n$. Then $\mathcal{M}$ contains the algebra of finite disjoint unions of measurable rectangles (Theorem 5.4.3).

② If E_n are increasing sets in $\mathcal{M}$ (not necessarily measurable rectangles), then $\mu\big((\bigcup_n E_n)^y\big) = \mu(\bigcup_n E_n^y) = \lim_{n\to\infty} \mu(E_n^y)$ is measurable, so $\bigcup_n E_n \in \mathcal{M}$.

Similarly, if E_n are decreasing sets in $\mathcal{M}$, then using limits we see that $\bigcap_n E_n \in \mathcal{M}$. (Here it is crucial that E_n^y have finite measure, for the limiting process to be valid.)

③ These arguments show that $\mathcal{M}$ is a monotone class, containing the algebra of finite unions of measurable rectangles. According to the monotone class theorem (Theorem 6.2.2), $\mathcal{M}$ must therefore be the same as the σ-algebra $\mathcal{A} \otimes \mathcal{B}$.

Thus we know $\mu(E^y)$ is measurable under the assumption that $\mu(X) < \infty$. For the σ-finite case, let $X_n \nearrow X$ with $\mu(X_n) < \infty$, and re-apply the arguments replacing μ by the finite measure $\mu_n(F) = \mu(F \cap X_n)$. Then $\mu(E^y) = \lim_{n\to\infty} \mu_n(E^y)$ is measurable. ∎

6.3 Iterated integrals

So far we have not considered the product measure which ought to assign to a measurable rectangle a measure that is a product of the measures of each of its sides. Such a measure can be obtained from the Carathéodory extension process, but it is just as easy to give an explicit integral formula:

6.3.1 THEOREM (PRODUCT MEASURE AS INTEGRAL OF CROSS-SECTIONAL AREAS)
Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. There exists a unique product measure $\mu \otimes \nu \colon \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$, such that

$$(\mu \otimes \nu)(E) = \int_{x \in X} \underbrace{\int_{y \in Y} \mathbb{I}((x,y) \in E)\, d\nu}_{\nu(E^x)}\, d\mu = \int_{y \in Y} \underbrace{\int_{x \in X} \mathbb{I}((x,y) \in E)\, d\mu}_{\mu(E^y)}\, d\nu\,.$$

Proof Let $\lambda_1(E)$ denote the double integral on the left, and $\lambda_2(E)$ denote the one on the right. These integrals exist by Theorem 6.2.3. λ_1 and λ_2 are countably additive by Beppo-Levi's theorem (Theorem 3.2.1), so they are both measures on $\mathcal{A} \otimes \mathcal{B}$. Moreover, if $E = A \times B$, then expanding the two integrals shows $\lambda_1(E) = \mu(A)\nu(B) = \lambda_2(E)$.

It is clear that $X \times Y$ is σ-finite under either λ_1 or λ_2, so by uniqueness of measures (Corollary 5.2.7), $\lambda_1 = \lambda_2$ on all of $\mathcal{A} \otimes \mathcal{B}$. ∎

There is not much work left for our final results:

6.3.2 THEOREM (FUBINI)
Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. If $f \colon X \times Y \to \overline{\mathbb{R}}$ is $\mu \otimes \nu$-integrable, then

$$\int_{X \times Y} f\, d(\mu \otimes \nu) = \int_{x \in X} \Big[\int_{y \in Y} f(x,y)\, d\nu \Big]\, d\mu = \int_{y \in Y} \Big[\int_{x \in X} f(x,y)\, d\mu \Big]\, d\nu\,.$$

This equation also holds when $f \geq 0$ (if it is merely measurable, not integrable).

Proof The case $f = \mathbb{I}(E)$ is just Theorem 6.3.1. Since all three integrals are additive, they are equal for non-negative simple f, and hence also for all other non-negative f, by approximation and monotone convergence.

For f not necessarily non-negative, let $f = f^+ - f^-$ as usual and apply linearity. Now it may happen that $\int_{y \in Y} f^{\pm}(x,y)\, d\nu$ might be both ∞ for some $x \in X$, and so $\int_{y \in Y} f(x,y)\, d\nu$ would not be defined. However, if f is $\mu \otimes \nu$-integrable, this can happen only on a set of measure zero in X (see the next theorem). The outer integral will make sense provided we ignore this set of measure zero. ∎

6.3.3 THEOREM (TONELLI)
Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces, and $f\colon X \times Y \to \overline{\mathbb{R}}$ be $\mu \otimes \nu$-measurable. Then f is $\mu \otimes \nu$-integrable if and only if

$$\int_{x \in X} \Big[\int_{y \in Y} |f(x,y)| \, d\nu \Big] \, d\mu < \infty$$

(or with X and Y reversed). Briefly, the integral of f can be described as being *absolutely convergent*.

Consequently, if a double integral is absolutely convergent, then it is valid to switch the order of integration.

Proof Immediate from Fubini's theorem applied to the non-negative function $|f|$. ■

6.3.4 EXAMPLE. When f is not non-negative, the hypothesis of absolute convergence is crucial for switching the order of integration. An elementary counterexample is the doubly-indexed sequence $a_{m,n} = (-m)^n / n!$ integrated with the counting measure. We have $\sum_n a_{m,n} = e^{-m}$, so $\sum_m \sum_n a_{m,n} = (1 - e^{-1})^{-1}$, but $\sum_m a_{m,n}$ diverges for all n.

6.3.5 EXAMPLE. This is another counterexample, using Lebesgue measure on $[0,1] \times [0,1]$. Define

$$f(x) = \frac{x^2 - y^2}{(x^2 + y^2)^2}, \quad (x,y) \neq (0,0).$$

We have

$$\frac{\partial}{\partial y} \left(\frac{y}{x^2 + y^2} \right) = \frac{x^2 - y^2}{(x^2 + y^2)^2} = \frac{\partial}{\partial x} \left(\frac{-x}{x^2 + y^2} \right),$$

giving

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \, dy \, dx = \int_0^1 \frac{dx}{1 + x^2} = \frac{\pi}{4}$$
$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \, dx \, dy = -\int_0^1 \frac{dy}{1 + y^2} = -\frac{\pi}{4}.$$

Thus f cannot be integrable. This can be seen directly by using polar coordinates:

$$\int_0^1 \int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dx \, dy \geq \int_0^{\pi/2} |\cos 2\theta| \, d\theta \int_0^1 \frac{1}{r} \, dr = \infty.$$

6.4 Exercises

6.1 (Alternative construction of product measure) Construct the product measure $\mu \otimes \nu$ using Carathéodory's theorem.

Hint: we have already done a similar thing before.

6.2 (σ-finiteness is necessary for Fubini's theorem) Cook up a counterexample to show that Fubini's Theorem does *not* hold when one of the measure spaces in the product measure space is not σ-finite.

6.3 (Infinite product σ-algebra) There is a definition of the product σ-algebra that works with any finite or infinite number of factors. It parallels the construction of the product topology.

If $\{(X_\alpha, \mathcal{M}_\alpha)\}_{\alpha \in \Lambda}$ are any measurable spaces, then

$$\bigotimes_{\alpha \in \Lambda} \mathcal{M}_\alpha = \sigma(\{\pi_\alpha^{-1}(E) \mid \alpha \in \Lambda, E \in \mathcal{M}_\alpha\}),$$

where π_α are the coordinate projections from the product space to X_α. Observe that $\bigotimes_{\alpha \in \Lambda} \mathcal{M}_\alpha$ is the smallest σ-algebra such that the projections π_α are measurable.

Show that this definition coincides with the definition in the text for the case of two factors. Also extend Lemma 6.1.5 and Theorem 6.1.6 to cover the infinite case; you may need to assume Λ is countable.

6.4 (Measurability of sum and product) Give a new proof using product σ-algebras, that if $f, g \colon X \to \mathbb{R}$ are measurable, then so is $f + g$, fg, and f/g. The new proof will be conceptually easier than the one we presented for Theorem 2.3.11, when we had not yet introduced product measure spaces.

6.5 (Measurability of metric) Suppose X is a measurable space, and Y is a *separable* metric space with metric d. Let there be two measurable functions $f, g \colon X \to Y$. Then the map $d(f, g) \colon X \to \mathbb{R}$ is measurable.

6.6 (Functions continuous in each variable) Suppose $f \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a function such that:

① $f(x, y)$ is measurable as x is varied while y is held fixed;

② $f(x, y)$ is continuous as y is varied while x is held fixed.

We can re-construct f on its entire domain as a limit using only countably infinitely many of the functions $x \mapsto f(x, y_n)$. Thus, a function of multiple real variables that is only continuous in each variable separately is still measurable in $\mathscr{B}(\mathbb{R} \times \cdots \times \mathbb{R})$.

6.7 (Associativity of product measure) Prove that $\otimes$ is associative for σ-algebras and σ-finite measures.

6.8 (Lebesgue measure as a product) If λ^n denotes the Lebesgue measure in n dimensions, we might try to summarize the result for product measures as:

$$\lambda^{n+m} = \lambda^n \otimes \lambda^m \,.$$

However, strictly speaking this equation is not correct, since a product measure is almost never complete (why?), while the left-hand side is a complete measure. Show that if we take the completion of the right-hand side, then the equation becomes correct.

This, of course, gives yet another method to construct Lebesgue measure in n dimensions starting from one: $\lambda^n = \lambda^1 \otimes \cdots \otimes \lambda^1$.

6.9 (Sine integral) Compute

$$\lim_{x\to\infty} \int_0^x \frac{\sin t}{t}\, dt$$

by considering the integral $\int_0^\infty \int_0^x e^{-ty} \sin t \, dt \, dy$ and switching the order of integration. However, you will need to be careful about the switch because $\int_0^\infty |\sin(t)/t| \, dt$ diverges.

6.10 (Area under a graph) Prove rigorously this statement from elementary calculus: "The integral $\int_a^b f(x)\, dx$, for any non-negative Riemann-integrable f, is the area under the graph of f."

(You will have to show that the region under a graph is Lebesgue-measurable in $\mathbb{R}^2$ in the first place.)

My friend once suggested that the Lebesgue integral $\int_X f \, d\mu$ could have been more simply *defined* as the measure of the region "under the graph of f", once we have in hand the measure μ (the theory in chapter 5 is entirely separate from integration). Unfortunately this idea falls down on one crucial point. What is the obstacle?

6.11 (β function) Show, using iterated integration, that the Γ function satisfies:

$$\frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)} = \beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt \,, \quad x, y > 0 \,.$$

6.12 (Monotone class theorem for functions) Let X be any set; we consider the set $\mathbb{R}^X$ of functions from X to $\mathbb{R}$, which forms a vector subspace under pointwise addition and scalar multiplication.

For $\mathbb{R}^X$ there are the two "lattice" operations:

$$f \vee g = \max(f, g) \,, \quad \bigvee_\alpha f_\alpha = \sup_\alpha f_\alpha \,,$$
$$f \wedge g = \min(f, g) \,, \quad \bigwedge_\alpha f_\alpha = \inf_\alpha f_\alpha \,.$$

(At first glance, these $\vee$ ("join") and $\wedge$ ("meet") symbols may seem to be extraneous notation, but they draw the connection to the $\cup$ and $\cap$ operations for sets.) A set $\mathcal{A} \subseteq \mathbb{R}^X$ is a *lattice* if it is closed under finite $\vee$ and $\wedge$.

A set $\mathcal{A} \subseteq \mathbb{R}^X$ is a *monotone class* if it is closed under limits of increasing or decreasing functions (provided that the limits are finite everywhere on X). The *monotone class generated by a set* $\mathcal{A} \subseteq \mathbb{R}^X$ is the smallest monotone class containing $\mathcal{A}$.

Then we have this result analogous to the one for sets: If $\mathcal{A} \subseteq \mathbb{R}^X$ is a vector subspace that is also a lattice, then the monotone class generated by $\mathcal{A}$ is a vector subspace closed under countable $\vee$ and $\wedge$.

Chapter 7

Integration over $\mathbb{R}^n$

This chapter treats basic topics on the Lebesgue integral over $\mathbb{R}^n$ not yet taken care of by the abstract theory. They may be read in any order.

7.1 Riemann integrability implies Lebesgue integrability

You have probably already suspected that any function that is Riemann-integrable is also Lebesgue-integrable, and certainly with the same values for the two integrals. We shall prove this formally.

First, we recall one definition of Riemann integrability. (This definition is different from most, but is easily seen to be equivalent; it makes our proofs a good deal simpler.)

Let $f\colon A \to \mathbb{R}$ be a bounded function on a bounded rectangle $A \subseteq \mathbb{R}^m$. Consider $\mathbb{R}$-valued functions that are simple with respect to a rectangular partition of A, otherwise known as step functions. Step functions are obviously both Riemann- and Lebesgue- integrable with the same values for the integral. The **lower and upper Riemann**[†] **integrals** for f are:

$$\mathscr{L}(f) = \sup\left\{\int_A l\,d\lambda : \text{step functions } l \leq f\right\}$$
$$\mathscr{U}(f) = \inf\left\{\int_A u\,d\lambda : \text{step functions } u \geq f\right\}.$$

We always have $\mathscr{L}(f) \leq \mathscr{U}(f)$; we say that f is (properly) **Riemann-integrable** if $\mathscr{L}(f) = \mathscr{U}(f)$, and the Riemann integral of f is defined as $\mathscr{L}(f) = \mathscr{U}(f)$.

Equivalently, f is Riemann-integrable when there exists sequences of lower simple functions $l_n \leq f$, and upper simple functions $u_n \geq f$, such that

$$\lim_{n\to\infty} \int_A l_n = \mathscr{L}(f) = \mathscr{U}(f) = \lim_{n\to\infty} \int_A u_n\,.$$

[†]More properly attributed to Gaston Darboux.

7.1.1 THEOREM (RIEMANN INTEGRABILITY IMPLIES LEBESGUE INTEGRABILITY)
Let $A \subset \mathbb{R}^m$ be a bounded rectangle. If $f\colon A \to \mathbb{R}$ is properly Riemann-integrable, then it is also Lebesgue-integrable (with respect to Lebesgue measure) with the same value for the integral.

Proof Choose a sequence l_n and u_n as above. Let $L = \sup_n l_n$ and $U = \inf_n u_n$. Clearly, these are measurable functions, and we have

$$l_n \leq L \leq f \leq U \leq u_n \,.$$

Lebesgue-integrating and taking limits,

$$\lim_{n\to\infty} \int l_n \leq \int L \leq \int U \leq \lim_{n\to\infty} \int u_n \,.$$

But the limit on the two sides are the same, because the Riemann and Lebesgue integrals for l_n and u_n coincide, so we must have $\int (U - L) = 0$. Then $U = L$ almost everywhere, and U (or L) equals f almost everywhere. Since Lebesgue measure is complete, f is a Lebesgue-measurable function (Theorem 5.6.2).

Finally, the Lebesgue integral $\int f$, which we now know exists, is squeezed in between the two limits on the left and the right, that both equal the Riemann integral of f. ∎

Having obtained a satisfactory answer for proper Riemann integrals, we turn briefly to improper Riemann integrals. Convergent improper integrals can be classed into two types: absolutely convergent and conditionally convergent. A typical example of the latter is:

$$\lim_{x\to\infty} \int_0^x \frac{\sin t}{t}\, dt = \frac{\pi}{2}\,, \quad \text{for which} \quad \lim_{x\to\infty} \int_0^x \left|\frac{\sin t}{t}\right| dt = \infty\,.$$

Since the Lebesgue integral is defined by integrating positive and negative parts separately, it cannot represent conditionally convergent integrals that converge by subtractive cancellation. A little thought shows that it is an intrinsic limitation of abstract measure theory: any integrable $f\colon X \to \overline{\mathbb{R}}$ must satisfy

$$\int_X f = \int_{E_1} f + \int_{E_2} f + \int_{E_3} f + \cdots$$

for any partition $\{E_n\}$ of X; yet we can re-arrange E_n in any way we like, so the series on the right is forced to be absolutely convergent.

Thus, in the context of Lebesgue measure theory, conditionally convergent integrals must still be handled by ad-hoc methods. On the other hand, absolutely convergent integrals are completely subsumed in the Lebesgue theory:

7.1.2 THEOREM (ABSOLUTELY-CONVERGENT IMPROPER RIEMANN INTEGRALS)
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function for which the limit of proper Riemann integrals

$$\lim_{n\to\infty} \int_{A_n} |f(x)|\, dx$$

is convergent, for some sequence of bounded rectangles A_n increasing to $\mathbb{R}^n$. Then the improper Riemann integral of f over $\mathbb{R}^n$, if it exists, equals its Lebesgue integral:

$$\lim_{n\to\infty} \int_{A_n} f(x)\,dx = \int_{\mathbb{R}^n} f(x)\,dx\,.$$

Proof Theorem 7.1.1 says each proper Riemann integral over A_n equals the Lebesgue integral. The monotone convergence theorem, applied to $f_n = |f|\mathbb{I}(A_n)$, shows that the Lebesgue integral $\int_{\mathbb{R}^n} |f(x)|\,dx$ is finite. The result follows from the dominated convergence of $f_n = f\,\mathbb{I}(A_n)$. ■

7.2 Change of variables in $\mathbb{R}^n$

This section will be devoted to completing the proof of the differential change of variables formula, Theorem 3.4.4. As we have noted, it suffices to prove the following.

7.2.1 Lemma (Volume differential)
Let $g\colon X \to Y$ be a diffeomorphism between open sets in $\mathbb{R}^n$. Then for all measurable sets $A \subseteq X$,

$$\lambda(g(A)) = \int_{g(A)} 1 = \int_A |\det \mathrm{D}g|\,. \qquad ❀$$

Proof We first begin with two simple reductions.

① It suffices to prove the lemma locally.

That is, suppose there exists an open cover of X, $\{U_\alpha\}$, so that the equation ❀ holds for measurable A contained inside one of the U_α. Then equation ❀ actually holds for all measurable $A \subseteq X$.

Proof By passing to a countable subcover (Theorem A.4.1), we may assume there are only countably many U_i. Define the disjoint measurable sets $E_i = U_i \setminus (U_1 \cup \cdots \cup U_{i-1})$, which also cover X. And define the two measures:

$$\mu(A) = \lambda(g(A))\,, \quad \nu(A) = \int_A |\det \mathrm{D}g|\,.$$

Now let $A \subseteq X$ be any measurable set. We have $A \cap E_i \subseteq U_i$, so $\mu(A \cap E_i) = \nu(A \cap E_i)$ by hypothesis. Therefore,

$$\mu(A) = \mu\Big(\bigcup_i A \cap E_i\Big) = \sum_i \mu(A \cap E_i) = \sum_i \nu(A \cap E_i) = \nu(A)\,.$$

■

② Suppose equation ❀ holds for two diffeomorphisms g and h, and all measurable sets. Then it holds for the composition diffeomorphism $g \circ h$, and all measurable sets.

Proof For any measurable A,

$$\int_{g(h(A))} 1 = \int_{h(A)} |\det \mathrm{D}g| = \int_A |(\det \mathrm{D}g) \circ h| \cdot |\det \mathrm{D}h| = \int_A |\det \mathrm{D}(g \circ h)| .$$

The second equality follows from Theorem 3.4.4 applied to the diffeomorphism h, which is valid once we know $\lambda(h(B)) = \int_B |\det \mathrm{D}h|$ for all measurable B. ∎

We proceed to prove the lemma by induction, on the dimension n.

Base case $n = 1$. Cover X by a countable set of bounded intervals I_k in $\mathbb{R}$. By reduction ①, it suffices to prove the lemma for measurable sets contained in each of the I_k individually. By the uniqueness of measures (Corollary 5.2.7), it also suffices to show $\mu = \nu$ only for the intervals $[a, b]$, (a, b), etc. But this is just the Fundamental Theorem of Calculus:

$$\int_{g([a,b])} 1 = |g(b) - g(a)| = |\int_a^b g'| = \int_a^b |g'| .$$

For the last equality, remember that g, being a diffeomorphism, must have a derivative that is positive on all of $[a, b]$ or negative on all of $[a, b]$.

If the interval is not closed, say (a, b), then we may not be able to apply the Fundamental Theorem directly, but we may still obtain the desired equation by taking limits:

$$\int_{g((a,b))} 1 = \lim_{n\to\infty} \int_{g([a+\frac{1}{n}, b-\frac{1}{n}])} 1 = \lim_{n\to\infty} \int_{[a+\frac{1}{n}, b-\frac{1}{n}]} |g'| = \int_{(a,b)} |g'| .$$

Induction step. According to Lemma A.3.3 in the appendix, the diffeomorphism g can always be factored locally (i.e. on a sufficiently small open set around each point $x \in X$) as $g = h_k \circ \cdots \circ h_2 \circ h_1$, where each h_i is a diffeomorphism and fixes one coordinate of $\mathbb{R}^n$. By reduction ①, it suffices to consider this local case only. By reduction ②, it suffices to prove the lemma for each of the diffeomorphisms h_i.

So suppose g fixes one coordinate. For convenience in notation, assume g fixes the last coordinate: $g(u, v) = (h_v(u), v)$, for $u \in \mathbb{R}^{n-1}, v \in \mathbb{R}$, and h_v are functions on open subsets of $\mathbb{R}^{n-1}$. Clearly h_v are one-to-one, and most importantly, $\det \mathrm{D}h_v(u) = \det \mathrm{D}g(u, v) \neq 0$.

Next, let a measurable set A be given, and consider its projection and cross-section:

$$V = \{v \in \mathbb{R} : (u, v) \in A\} , \quad U_v = \{u \in \mathbb{R}^{n-1} : (u, v) \in A\} .$$

We now apply Fubini's theorem (Theorem 6.3.2) and the induction hypothesis on the diffeomorphisms h_v:

$$\begin{aligned}\int_{g(A)} 1 &= \int_{v\in V}\int_{h_v(U_v)} 1 \\ &= \int_{v\in V}\int_{u\in U_v} |\det \mathrm{D}h_v(u)| \\ &= \int_{v\in V}\int_{u\in U_v} |\det \mathrm{D}g(u,v)| = \int_A |\det \mathrm{D}g|\,. \end{aligned}$$ ■

The proof presented here is an "algebraic" one; the only fact about the determinant that was ever used is that it is a multiplicative function on matrices. This comes as no surprise — the determinant can be defined as the unique matrix function satisfying certain multi-linearity axioms, and those axioms characterize the signed volume function on parallelograms.

Most treatises on measure theory go for a more direct and "geometric" proof of Lemma 7.2.1. They first demonstrate that equation ❀ holds for linear transformations $g = T$ applied to rectangles A, then they show it holds in general by approximation arguments. We elaborate on these arguments in the exercises.

Our approach, adapted from [Munkres], has the advantage of not needing nitty-gritty "ε" estimates, and it handles the special case for a linear transformation $g = T$ in one fell swoop.

7.3 Integration on manifolds

So far, we have worked chiefly only with the measure of n-dimensional volume on $\mathbb{R}^n$. There is also a concept of k-dimensional volume, defined for k-dimensional subsets of $\mathbb{R}^n$.

A **manifold** is a generalization of curves and surfaces to higher dimensions. We shall concentrate on differentiable manifolds embedded in $\mathbb{R}^n$; the theory is elucidated in [Spivak2] or [Munkres]. Here we give a definition of the k-dimensional volume for k-dimensional manifolds that does not require those dreaded "partitions of unity".

7.3.1 DEFINITION (VOLUME OF k-DIMENSIONAL PARALLELOPIPED.)
Define, for any vectors $w_1, \dots, w_k \in \mathbb{R}^n$ (represented under the standard basis or any other orthonormal basis),

$$\begin{aligned}\mathrm{V}(w_1,\dots,w_k) &= \sqrt{\det \begin{bmatrix} w_1 & w_2 & \dots & w_k\end{bmatrix}^{\mathrm{tr}} \begin{bmatrix} w_1 & w_2 & \dots & v_k\end{bmatrix}} \\ &= \sqrt{\det\,[w_i\cdot w_j]_{i,j=1,\dots,k}}\,,\end{aligned}$$

This is the **k-dimensional volume** of a k-dimensional *parallelopiped spanned by* the vectors $w_1, \dots, w_k$ in $\mathbb{R}^n$.

One can easily show that this volume is invariant under orthogonal transformations, and that it agrees with the usual k-dimensional volume (as defined by the Lebesgue measure) when the parallelopiped lies in the subspace $\mathbb{R}^k \times 0 \subseteq \mathbb{R}^n$.

7.3.2 DEFINITION (VOLUME OF k-DIMENSIONAL PARAMETERIZED MANIFOLD)
Suppose a k-dimensional manifold $M \subseteq \mathbb{R}^n$ is covered by a single coordinate chart $\alpha \colon U \to M$, for $U \subseteq \mathbb{R}^k$ open. Let $\mathrm{D}\alpha$ denote the n-by-k matrix

$$\mathrm{D}\alpha = \begin{bmatrix} \dfrac{d\alpha}{dt_1} & \dfrac{d\alpha}{dt_2} & \dots & \dfrac{d\alpha}{dt_k} \end{bmatrix}.$$

The **k-dimensional volume** of any $E \in \mathscr{B}(M)$ is defined as:

$$\nu(E) = \int_{\alpha^{-1}(E)} \mathrm{V}(\mathrm{D}\alpha)\, d\lambda\,.$$

The integrand, of course, is supposed to represent "infinitesimal" elements of surface area (k-dimensional volume), or approximations of the surface area of E by polygons that are "close" to E. As indicated by the quotation marks, these assertions about "surface area" are not rigorous, and we will not belabor to prove them. We take the equation above as our definition of k-dimensional volume. But it should be pointed out that there are better theories of k-dimensional volume available, such as the Hausdorff measure, which are intrinsic to the sets being measured, instead of our computational theory.

7.3.3 DEFINITION (VOLUME OF k-DIMENSIONAL MANIFOLD)
If the manifold M is not covered by a single coordinate chart, but more than one, say $\alpha_i \colon U_i \to M$, $i = 1, 2, \dots$, then partition M with $W_1 = \alpha_1(U_1)$, $W_i = \alpha_i(U_i) \setminus W_{i-1}$, and define

$$\nu(E) = \sum_i \int_{\alpha_i^{-1}(E \cap W_i)} \mathrm{V}(\mathrm{D}\alpha_i)\, d\lambda\,.$$

It is left as an exercise to show that $\nu(E)$ is well-defined: it is independent of the coordinate charts α_i used for M.

Finally, the scalar integral of $f \colon M \to \mathbb{R}$ over M is simply

$$\int_M f\, d\nu\,.$$

And the integral of a differential form ω on an oriented manifold M is

$$\int_{p \in M} \omega\big(p\,; T(p)\big)\, d\nu\,,$$

where $T(p)$ is an orthonormal frame of the tangent space of M at p, oriented according to the given orientation of M.

Again it is not hard to show that the formulae given here are exactly equivalent to the classical ones for evaluating scalar integrals and integrals of differential forms, which are of course needed for actual computations. But there are several advantages to our new definitions. First is that they are elegant: they are mostly coordinate-free, and all the different integrals studied in calculus have been unified to the Lebesgue integral by employing different measures. In turn, this means that the nice properties and convergence theorems we have proven all carry over to integrals on manifolds.

For example, everybody "knows" that on a sphere, any circular arc C has "measure zero", and so may be ignored when integrating over the sphere. To prove this rigorously using our definitions, we only have to remark that $\nu(C) = 0$, since $\lambda(\alpha^{-1}(C)) = 0$ for some coordinate chart α for the sphere.

7.4 Stieltjes integrals

The definition of the Stieltjes integrals and measures is best motivated by probability theory. Suppose we have a $\mathbb{R}$-valued random variable Z with distribution μ — that is, for every Borel set $B \subseteq \mathbb{R}$,

$$\mathrm{P}\{Z \in B\} = \mu(B)\,.$$

Then we can form the the cumulative probability function for the measure μ:

$$F(z) = \mathrm{P}\{Z \leq z\} = \mu(-\infty, z]\,.$$

The numerical function F is a useful summary of the much more complicated set function μ that can be readily used for computation. (For example, the cumulative probability function for the Gaussian distribution is tabulated in almost all statistics textbooks and widely implemented in computer spreadsheets and statistical software.)

It is clear that F determines μ on all Borel sets on $\mathbb{R}$: we have

$$\mu(a, b] = \mu\big((-\infty, b] \setminus (-\infty, a]\big) = F(b) - F(a)\,,$$

and the intervals $(a, b]$ generate $\mathscr{B}(\mathbb{R})$.

The function F satisfies two key properties:

① F is *increasing*, for $F(b) - F(a) = \mu(a, b] \geq 0$ whenever $a \leq b$.

② Since F is increasing, it always has only a countable number of discontinuities, and these discontinuities must all be *jump discontinuities*. At these jumps, F is *right-continuous*, for

$$F(b) - F(a) = \mu(a, b] = \lim_{x \searrow b} \mu(a, x] = \lim_{x \searrow b} F(x) - F(a)\,.$$

(F has a jump at b exactly when μ has a point mass at b: $\mu[b, b] \neq 0$.)

We might ask: is it possible to carry out this procedure in reverse? That is, *given an increasing function F, can we find a measure μ* on $\mathscr{B}(\mathbb{R})$ that assigns, to any interval $(a, b]$, a "length" of $F(b) - F(a)$? This section will show that it is possible.

7.4.1 Definition (Cumulative distribution function)
Let μ be a positive measure on $\mathscr{B}(\mathbb{R})$ that is finite on all bounded subsets of $\mathbb{R}$. A **cumulative distribution function** for μ is any function $F\colon \mathbb{R} \to \mathbb{R}$ such that:

$$F(b) - F(a) = \mu(a, b]\,, \quad a \leq b\,.$$

As demonstrated above, such F is always *increasing* and *right-continuous*.

7.4.2 Remark. If μ is a finite measure, then we may take $F(z) = \mu(-\infty, z]$ as before. Otherwise, we may use

$$F(z) = \begin{cases} \mu(c, z]\,, & z \geq c \\ -\mu(z, c]\,, & z \leq c \end{cases}$$

for any finite centre $c \in \mathbb{R}$. Clearly, the particular choice of c is inconsequential and only changes F by an additive constant.

7.4.3 Theorem
A function $F\colon \mathbb{R} \to \mathbb{R}$ that is increasing and right-continuous is a cumulative distribution function for a unique positive measure μ on $\mathscr{B}(\mathbb{R})$.

The technique for constructing the measure μ is not much different from that for Lebesgue measure on $\mathbb{R}$, which is not surprising, since the cumulative distribution function $F(x) = x$ corresponds exactly to Lebesgue measure.

Definition of pre-measure. We will need to deal with unbounded intervals as well, so extend F with

$$F(-\infty) = \lim_{z \to -\infty} F(z) = \inf_{z \in \mathbb{R}} F(z)\,, \quad F(+\infty) = \lim_{z \to +\infty} F(z) = \sup_{z \in \mathbb{R}} F(z)\,.$$

These limits may be $-\infty$ and $+\infty$ respectively.

Define $\mu(a, b] = F(b) - F(a)$ for $a, b \in \overline{\mathbb{R}}$ and $a < b$. We extend μ in the obvious way to the algebra $\mathcal{A}$ of disjoint unions of intervals of the form $(a, b] \subseteq \mathbb{R}$ (Theorem 5.4.3).

It is straightforward to verify the well-definedness and finite additivity of μ on $\mathcal{A}$ by arguments similar to those of section 5.4.

Countable additivity. Countable additivity of μ on $\mathcal{A}$ requires the typical approximation via finite additivity. It suffices, from Remark 5.1.7, to prove that $\mu(I) \leq \sum_{n=1}^{\infty} \mu(J_n)$ if the sets $J_n \in \mathcal{A}$ cover $I \in \mathcal{A}$.

We first assume that $I = (a, b]$ is a simple bounded interval. And we may as well assume, without loss of generality, that $J_n = (a_n, b_n]$ are simple bounded intervals.

Since F is right-continuous, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$0 \leq F(b) - F(a) < F(b) - F(a + \delta) + \varepsilon\,.$$

Also there exists $\delta_n > 0$ such that

$$0 \leq F(b_n + \delta_n) - F(a_n) < F(b_n) - F(a_n) + \frac{\varepsilon}{2^n}\,.$$

There exists a finite set $n_1, \ldots, n_k$ such that

$$(a + \delta, b] \subseteq \bigcup_{i=1}^{k} (a_{n_i}, b_{n_i} + \delta_{n_i}]\,,$$

since the open sets $(a_n, b_n + \delta_n)$ cover the compact set $[a + \delta, b]$. Taking μ of both sides, we find

$$F(b) - F(a + \delta) \leq \sum_{i=1}^{k} F(b_{n_i} + \delta_{n_i}) - F(a_{n_i}) \leq \sum_{n=1}^{\infty} F(b_n + \delta_n) - F(a_n)\,,$$

$$F(b) - F(a) \leq 2\varepsilon + \sum_{n=1}^{\infty} F(b_n) - F(a_n)\,,$$

and we take $\varepsilon \searrow 0$.

If the bounds on the interval I are infinite, then it suffices to apply the bounded case above and take the limits $a \to -\infty$, $b \to +\infty$ or both. Finally, if $I \in \mathcal{A}$ is a disjoint union of simple intervals $I_1, \ldots, I_k$, then

$$\mu(I) = \sum_{m=1}^{k} \mu(I_m) \leq \sum_{m=1}^{k} \sum_{n=1}^{\infty} \mu(J_n \cap I_m) = \sum_{n=1}^{\infty} \sum_{m=1}^{k} \mu(J_n \cap I_m) = \sum_{n=1}^{\infty} \mu(J_n)\,,$$

since $\{J_n \cap I_m\}_n$ cover each interval I_m.

Carathéodory's theorem (Theorem 5.2.2) then yields an extension of μ to $\mathscr{B}(\mathbb{R})$; it is unique as μ is σ-finite.

The measure μ is called the **Lebesgue-Stieltjes measure** induced from F, while an integral with respect to μ, denoted variously:

$$\int g\, d\mu = \int g\, dF = \int g(x)\, dF(x)\,,$$

is called a **Lebesgue-Stieltjes integral**. The Lebesgue-Stieltjes integral is the generalization of the *Riemann-Stieltjes* integral that is the limits of sums of the form:

$$\sum_i g(\xi_i) \cdot \big(F(x_i) - F(x_{i-1})\big)\,, \quad x_{i-1} < \xi_i \leq x_i\,,$$

first introduced by Thomas Stieltjes in 1894.

Further developments about Lebesgue-Stieltjes integrals may be anticipated if we observe that:

① If F is continuously differentiable, then dF really means what the notation suggests: $dF = F'dx$ and $F(b) - F(a) = \int_a^b dF$.

Is there an analogous formula that applies even when F is not smooth?

② In the formula $F(b) - F(a) = \int_a^b F'(x)\,dx$ obviously there needs no restriction that the anti-derivative F has to be increasing.

The anti-derivative F is decreasing whenever its slope F' is negative. Indeed, we integrate over a domain where F' is negative, the area between the graph of F' and the horizontal axis is counted negatively, and the cumulative area function is decreasing.

Perhaps we can articulate a general situation where certain areas are counted negatively, and construct "measures with sign" corresponding to cumulative distribution functions F that are not necessarily increasing.

In the next few chapters, we will investigate the answers to these questions.

7.5 Exercises

7.1 (Necessary and sufficient conditions for Riemann integrability) Let A be a bounded rectangle in $\mathbb{R}^m$. A bounded function $f\colon A \to \mathbb{R}$ is Riemann-integrable if and only if f is continuous almost everywhere.

This result is usually found with elementary proofs in beginning analysis books, but the "advanced" proof, using the language of measure theory, may be easier to grasp. Use these facts:

① Consider these quantities, the continuous analogue of lim inf and lim sup for discrete sequences:

$$\underline{f}(x) = \lim_{\delta \searrow 0} \inf_{\|y-x\| \le \delta} f(y)\,, \quad \overline{f}(x) = \lim_{\delta \searrow 0} \sup_{\|y-x\| \le \delta} f(y)\,.$$

We have $\underline{f}(x) = \overline{f}(x)$ if and only if f is continuous at x.

② If the lower and upper step approximations l_n and u_n for f are determined by finer and finer partitions with dyadic mesh sizes proportional to 2^{-n}, then $L = \sup_n l_n = \underline{f}$, and $U = \inf_n u_n = \overline{f}$ almost everywhere.

7.2 Is a Riemann-integrable function necessarily Borel-measurable?

7.3 Let μ be a finite measure on a metric space X. A bounded function $f\colon X \to \mathbb{R}$ is μ-Riemann-integrable if for every $\varepsilon > 0$, there exist continuous functions $l \le f$ and $u \ge f$ such that $\int_X (u - l)\,d\mu < \varepsilon$.

It is possible to generalize the Riemann integral to other

7.4 (Volume of parallelopiped) Show directly, without using Lemma 7.2.1, that the volume of

$$P = \{a + t_1 v_1 + \cdots t_n v_n \mid 0 \leq t_i \leq 1\}, \quad \text{for } a, v_1, \ldots, v_n \in \mathbb{R}^n,$$

is $|\det(v_1, \ldots, v_n)|$.

7.5 (Alternate proof of differential change-of-variables) Lemma 7.2.1 can be proven by directly estimating the volume $\lambda(g(Q))$ for small cubes Q, and then appealing to the Radon-Nikodým theorem and related results from the next chapter.

Assume that $g \colon U \to V$ is a continuous differentiable mapping between open sets U, V of $\mathbb{R}^n$, with non-singular derivative. Define the measure $\mu = \lambda \circ g$.

① If $E \subseteq U$ has measure zero, then so does $g(E)$. Thus the Radon-Nikodým derivative $\frac{d\mu}{d\lambda}$ exists.

This part is tricky. Using completeness of the Lebesgue measure, it can be proven in a few lines.

A more constructive proof is also possible, by bounding the size of $g(Q)$ for small cubes Q. You will need σ-compactness to be able to bound the derivative $\mathrm{D}g$.

② Let $h > 0$, and let

$$Q_h = \{x \in \mathbb{R}^n \mid \|x - a\| \leq h/2\}, \quad \|u\| = \max_{j=1,\ldots,n} |u_j|$$

be the cube with centre a and sides of length h. Show that as $h \to 0$,

$$\mu(Q_h) \leq |\det \mathrm{D}g(a)| \big(h + o(h)\big)^n.$$

③ Hence $\frac{d\mu}{d\lambda} \leq |\det \mathrm{D}g|$.

④ Analogously $\frac{d\lambda}{d\mu} \leq |\det \mathrm{D}g|^{-1}$, and so conclude with $\frac{d\mu}{d\lambda} = |\det \mathrm{D}g|$.

The constructive proof of part ① should suggest generalizations of the results to functions g that are merely *locally Lipschitz*. The interested reader may wish to pursue this further; the line of argument in this exercise is suggested in [Guzman].

7.6 (Special case of Sard's theorem) Let $g \colon X \to Y$ be a continuously differentiable function on an open set $U \subseteq \mathbb{R}^n$, and

$$A = \{x \in X \mid \det \mathrm{D}g(x) = 0\}$$

be the set of critical points of g. Then $g(A)$ has Lebesgue-measure zero.

We can prove this theorem using similar techniques to to that of the previous exercise: estimating volumes of the images of cubes. However, this time around we require a more careful argument, as the function g may not be bijective, and the domain A may be large in volume.

① It suffices to show that $\lambda(g(Q \cap A)) = 0$ for any closed cube $Q \subseteq X$, not necessarily small. And then we shall show $\lambda(g(Q \cap A)) = 0$ by establishing $\lambda(g(Q \cap A)) < \varepsilon$ for arbitrary $\varepsilon > 0$.

② Exploit the uniform continuity of $\mathrm{D}g$ on a compact subset to show,

$$g(y) - g(x) = \mathrm{D}g(x)(y - x) + o(\|y - x\|)\,,$$

as $\|y - x\| \to 0$, *uniformly* for $x, y \in Q$.

More precisely, for every $\varepsilon > 0$, there exists $\delta > 0$, such that:

$$x, y \in Q\,,\ \|y - x\| \leq \delta \implies \|g(y) - g(x) - \mathrm{D}g(x)(y - x)\| \leq \varepsilon \|y - x\|\,.$$

③ Divide the cube Q into small sub-cubes whose sides are shorter than δ. If an individual sub-cube S intersects A, then $g(S)$ is bounded by a box whose height along some axis is proportional to $\varepsilon\delta$. Consequently, there exists a constant $C > 0$, depending only on the dimension n, such that

$$\lambda(g(S)) \leq C\varepsilon\delta^n\,.$$

④ Complete the proof.

7.7 (Extended version of change-of-variables) Prove the following intuitively-obvious generalization of Lemma 7.2.1.

Let $g\colon X \to Y$ be a continuously differentiable function on an open set $X \subseteq \mathbb{R}^n$, not necessarily a diffeomorphism. Then for any measurable $E \subseteq X$,

$$\int_E |\det \mathrm{D}g(x)|\, dx = \int_{g(E)} \#g|_E^{-1}(y)\, dy\,,$$

where $\#g|_E^{-1}(y)$ counts the number of pre-images of y in E.

(The set $\{x \in X \mid \det \mathrm{D}g(x) = 0\}$ can be neglected, according to Sard's theorem from the previous exercise.)

7.8 (Yet another proof of differential change-of-variables) Here is a geometrical proof of Lemma 7.2.1, combining the ideas found in Exercise 7.5 and Exercise 7.6, but without recourse to the Radon-Nikodým theorem.

Let $g\colon X \to Y$ be a diffeomorphism of open sets in $\mathbb{R}^n$.

① Given $\varepsilon > 0$, taking S to be a sufficiently small cube with $x \in S$ as its centre, we have

$$\lambda(g(S)) \leq (1 + \varepsilon)|\det \mathrm{D}g(x)|\,\lambda(S)\,.$$

② For a cube Q of any size,

$$\lambda(g(Q)) \leq \int_Q |\det \mathrm{D}g|\, d\lambda\,.$$

③ The previous inequality continues to hold if Q is replaced by any measurable set E.

④ Conclude that $\lambda(g(E)) = \int_E |\det \mathrm{D}g| \, d\lambda$ for all measurable $E \subseteq X$.

This line of argument comes from [Schwartz], which also has a brief survey of some other proofs of the change-of-variables formula.

7.9 (Change-of-variables for Lebesgue-measurable sets) Extend Lemma 7.2.1 to cover the case when the set A is only Lebesgue-measurable.

7.10 (Lower-dimensional objects have measure zero) Any k-dimensional differentiable manifold in $\mathbb{R}^n$, for $n > k$, always has Lebesgue-measure zero in $\mathbb{R}^n$.

7.11 (Pappus's or Guldin's theorem for volumes) The solid of revolution in $\mathbb{R}^3$, generated by rotating a plane figure R in the x^+-z plane, about the z axis, has volume $V = ad$, where a is the area of R, and d is the distance travelled by the centroid of R under the rotation.

The **centroid** of a measurable set $E \subseteq \mathbb{R}^n$, with respect to a (signed) measure ν, is defined as:

$$\vec{m} = \frac{1}{\nu(E)} \int_{\vec{x} \in E} \vec{x} \, d\nu \,,$$

provided that $\nu(E) \neq 0, \infty$.

7.12 (Pappus's or Guldin's theorem for surfaces) The surface of revolution in $\mathbb{R}^3$, generated rotating a curve γ in the x^+-z plane about the z axis, has surface area $A = ld$, where l is the arc-length of γ, and d is the distance travelled by the centroid of γ under the rotation.

7.13 (Gaussian integral) Compute explicitly

$$\int_{\mathbb{R}^n} e^{-a\|x\|^2} \, dx \,, \quad a > 0 \,, \quad \|x\|^2 = \sum_{i=1}^n x_i^2 \,.$$

The calculation in chapter 1 is the special case $a = \frac{1}{2}$, $n = 1$.

7.14 (Expectation of Gaussian integral) Let X be a standard Gaussian random variable; that is, it has the distribution

$$\mathrm{P}(X \leq z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}\zeta^2} \, d\zeta \,.$$

Compute the expectation $\mathrm{E}[\Phi(aX + b)]$ for any constants $a, b \in \mathbb{R}$.

7.15 (Relation between surface area and volume of ball) Regard S^{n-1}, the $(n-1)$-dimensional unit sphere, as a $(n-1)$-dimensional differentiable manifold in $\mathbb{R}^n$. Let E be measurable in S^{n-1}. Let

$$E_R = \{rx \mid x \in E,\, 0 \le r \le R\} \subseteq \mathbb{R}^n\,,$$

be the cone formed from the spherical patch E with radius R. Using the computational definitions in section 7.3, show that the Lebesgue measure of E_R is

$$\lambda(E_R) = \int_0^R r^{n-1}\sigma(E)\,dr\,,$$

where σ is the surface area measure on S^{n-1}.

What is the intuitive meaning behind this equation?

7.16 (Polar coordinates) If $x \in \mathbb{R}^n \setminus \{0\}$, the **generalized polar coordinates** of x are:

$$r = \|x\| \in (0,\infty)\,, \qquad \hat{r} = \frac{x}{\|x\|} \in S^{n-1}\,.$$

(The Euclidean norm is used.)

Let $\phi\colon \mathbb{R}^n \setminus \{0\} \to (0,\infty) \times S^{n-1}$ be this polar coordinate mapping. Show, for any Borel-measurable $E \subseteq \mathbb{R}^n$, that

$$\lambda(\phi^{-1}(E)) = (\rho \otimes \sigma)(E)\,, \quad d\rho = r^{n-1}\,dr$$

where λ is the usual Lebesgue measure on $\mathbb{R}^n$, and σ is the surface area measure on S^{n-1}.

Conclude that for any measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$, non-negative or integrable,

$$\int_{\mathbb{R}^n} f(x)\,dx = \int_0^\infty \int_{\hat{r} \in S^{n-1}} f(r\hat{r})\, r^{n-1}\,d\sigma\,dr\,.$$

7.17 (Formula for surface area and volume of ball) Calculate the surface area of the unit sphare and the volume of the unit ball in $\mathbb{R}^n$, by integrating $e^{-\|x\|^2}$ over $x \in \mathbb{R}^n$ with generalized polar coordinates. (If you get stuck, look at the next exercise for a hint.)

7.18 (Evaluating polynomials over a sphere) Amazingly, the trick used in the previous exercise can be extended to integrate any polynomial over S^{n-1}. It gives the neat formula:

$$\int_{x \in S^{n-1}} |x_1|^{\alpha_1} \cdots |x_n|^{\alpha_n}\,d\sigma = \frac{2\,\Gamma\big(\frac{\alpha_1+1}{2}\big)\cdots\Gamma\big(\frac{\alpha_n+1}{2}\big)}{\Gamma\big(\frac{\alpha_1+\cdots+\alpha_n+n}{2}\big)}\,, \quad \alpha_1,\ldots,\alpha_n \ge 0\,.$$

The Γ function was defined in Example 3.5.5.

A related problem is to evaluate, without relying on anti-derivatives,

$$\int_0^{\pi/2} \cos^k\theta\,\sin^l\theta\,d\theta\,.$$

7.19 (Multi-variate cumulative distribution functions) We can generalize the notion of the Lebesgue-Stieltjes measure and cumulative distribution functions to the setting of $\mathbb{R}^n$. It is important in the field of probability and statistics to describe the distribution of a finite set of random variables that may not be independent.

Define the operator $\Delta_i(a)$ acting on functions of n variables:

$$\Delta_j(a)F(x_1,\dots,x_n) = \\ F(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) - F(x_1,\dots,x_{i-1},a,x_{i+1},\dots,x_n)\,,$$

for $1 \leq i \leq n$ and $a \in \mathbb{R}$. (In words, take a finite difference on the i^{th} variable but leave the other variables fixed.)

Suppose μ is a finite measure on $\mathbb{R}^n$. (We stick to finite measures for simplicity.) Its cumulative distribution function is:

$$F(x_1,\dots,x_n) = \mu\big((-\infty,x_1] \times \cdots \times (-\infty,x_n]\big)\,.$$

Then

$$\mu\big((a_1,b_1] \times \cdots \times (a_n,b_n]\big) = \Delta_1(a_1)\cdots\Delta_n(a_n)F(b_1,\dots,b_n)\,.$$

(What is the geometric meaning of this equation? Note that the operators $\Delta_1(a_1)$, $\dots$, $\Delta_n(a_n)$ commute.)

It follows that the necessary conditions for a function $F\colon \mathbb{R}^n \to \mathbb{R}$ to be a cumulative distribution function for a finite measure μ on $\mathscr{B}(\mathbb{R}^n)$ are:

① F must be increasing in each variable when the other variables are held fixed.

② F must be right-continuous in each variable.

③ If any of the variables tend to $-\infty$, then F tends to 0.

④ $\Delta_1(a_1)\cdots\Delta_n(a_n)F(b_1,\dots,b_n) \geq 0$ for any choice of real numbers $a_1 \leq b_1$, $\dots$, $a_n \leq b_n$.

Show that these conditions are also sufficient for constructing the finite measure μ corresponding to the multi-variate cumulative distribution function F.

7.20 (Riemann-Stieltjes sums) What are the conditions for the Riemann-Stieltjes sums to converge to the Lebesgue-Stieltjes integral?

7.21 (Mapping of random variables to uniform distribution) It is often required to simulate random samples from some specified probability distribution. However, most computer algorithms only generate uniformly distributed (pseudo-)random samples. To get other probability distributions, we can do a mapping to and from the uniform distribution on $[0,1]$, as follows.

Let X be a $\mathbb{R}$-valued random variable, and $F\colon \mathbb{R} \to [0,1]$ be its cumulative probability distribution function. Then $U = F(X)$ has the uniform distribution.

If F is invertible, then conversely, given any U uniformly distributed on $[0,1]$, the random variable $X = F^{-1}(U)$ has the probability distribution determined by F. If F is not invertible, the same procedure works provided we define:

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \mid F(x) \geq u\}, \quad u \in (0,1).$$

7.22 (Constructing independent random variables) For any countable set of probability measures $\mu_n \colon \mathscr{B}(\mathbb{R}) \to [0,1]$, there exists a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ having independent random variables $X_n \colon \Omega \to \mathbb{R}$ whose individual probability laws are given by μ_n respectively.

Hint: Exercise 7.21 and Exercise 5.7.

Chapter 8

Decomposition of measures

In this chapter, we discuss ways to decompose a given measure in terms of some other measures. These decompositions are used a fair amount in probability theory, and also in chapter 10.

8.1 Signed measures

A **signed measure** is nothing more than measure that is not restricted to take non-negative values. We have already seen a need for signed measures as we developed Stieltjes integrals in section 7.4. Mathematically, considering signed measures is useful as they form a vector space closed under addition and scalar multiplication. In terms of physical applications, we can, for example, model an electric charge distribution with signed measures, just as we can model a mass distribution with ordinary positive measures.

8.1.1 DEFINITION (SIGNED MEASURE)
A **signed measure** on a measurable space $(X, \mathcal{M})$ is a function $\nu\colon \mathcal{M} \to \mathbb{R}$ satisfying

① $\nu(\emptyset) = 0$.

② *Countable additivity*: For any sequence of *mutually disjoint* sets $E_n \in \mathcal{M}$,

$$\nu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \nu(E_n)\,.$$

Since the set union is commutative, the infinite sum on the right must converge to the same value no matter how the summands are rearranged. In other words, we must have absolute convergence.

We disallow infinite values for a signed measure so that no ambiguity can arise in adding or subtracting values. Infinite values for signed measures are not needed in most applications anyway.

8.1.2 EXAMPLE (SIGNED MEASURE FROM DENSITIES). If μ is a positive measure, and g is integrable with respect to μ, then

$$\nu(E) = \int_E g\,d\mu = \int_E g^+\,d\mu - \int_E g^-\,d\mu$$

is a signed measure. This generalizes Theorem 3.3.1 which restricts g to be non-negative.

8.1.3 EXAMPLE (DIFFERENCE OF MEASURES). If μ_1 and μ_2 are two positive measures on a measurable space, then their difference $\nu = \mu_1 - \mu_2$ is a signed measure.

Despite the seeming generality of the definition, all signed measures turn out to be instances of these two examples: every signed measure is a difference of two positive measures. The whole purpose of this section is to reduce the study of signed measures to the study of the positive measures that we are familiar with.

The idea behind the canonical decomposition of signed measures is to find parts of the measurable space X where the signed measure takes the same sign throughout, positive or negative. This motivates the following definition.

8.1.4 DEFINITION (NULL, POSITIVE, NEGATIVE SETS)
Let ν be a signed measure on a measurable space X. A measurable set $E \subseteq X$ is:

- **null for** μ if $\mu(F) = 0$ for all $F \subseteq E$ measurable;
- **positive for** μ if $\mu(F) \geq 0$ for all $F \subseteq E$ measurable;
- **negative for** μ if $\mu(F) \leq 0$ for all $F \subseteq E$ measurable.

8.1.5 THEOREM (HAHN DECOMPOSITION)
If ν is a signed measure on X, then X can be partitioned into two measurable sets, one of which is positive for ν and the other is negative for ν.

The partition is unique up to ν-null sets.

Proof Let $\alpha = \sup\{\nu(E) \colon E \subseteq X \text{ measurable}\}$. This quantity is finite according to Lemma 8.1.6 immediately following this theorem.

Take measurable sets E_n such that $\alpha - \nu(E_n) \leq 2^{-n}$. We claim that this set,

$$P = \limsup_n E_n\,, \quad \text{that is,} \quad P = \bigcap_n F_n\,, \text{ where } F_n = \bigcup_{k \geq n} E_k\,,$$

can be taken as a positive set for ν. We will make some estimates to show $\nu(P) = \alpha$; then it follows immediately that P cannot contain any sets of strictly negative measure, and that any set not contained in P cannot have strictly positive measure.

Since the sets F_n are decreasing as $n \to \infty$, we have $\nu(P) = \lim_n \nu(F_n)$. (The proof is the same as for positive measures.) So we estimate $\nu(F_n)$:

$$\begin{aligned}
\nu(F_n) &= \nu(E_n) + \sum_{k=n}^{\infty} \nu(E_{k+1} \setminus E_k) && \text{(disjointify the union of } E_k\text{)} \\
&= \nu(E_n) + \sum_{k=n}^{\infty} \bigl[\nu(E_{k+1}) - \nu(E_{k+1} \cap E_k)\bigr] \\
&\geq \alpha - \frac{1}{2^n} - \sum_{k=n}^{\infty} \frac{1}{2^{k+1}} \quad \to \alpha\,, && \text{as } n \to \infty.
\end{aligned}$$

For the last inequality, we use the fact that $\nu(E_{k+1} \cap E_k) \leq \alpha \leq \nu(E_{k+1}) + 2^{-(k+1)}$.

But this means $\nu(P) \geq \alpha$ (and hence $= \alpha$).

We now show uniqueness. Suppose X has two partitions, $\{P, N\}$ and $\{P', N'\}$, where P, P' are positive for ν, and N, N' are negative for ν. Let E be any measurable set; then $\nu\bigl(E \cap P \cap (X \setminus P')\bigr) = \nu(E \cap P \cap N')$ is zero since it is both ≤ 0 and ≥ 0. Similarly, $\nu(E \cap N \cap P')$ is zero. Thus the difference between P versus P', and the difference between N versus N', are ν-null sets. ∎

8.1.6 Lemma (Finiteness of total variation)

For any signed measure ν on a measurable space X, the set of real numbers

$$\{\nu(E) \mid E \subseteq X \text{ measurable}\}$$

is bounded above.

This lemma is really a manifestation of the fact that the total variation $|\nu|$ (to be defined shortly) is never infinite; it is certainly more easily remembered this way.

Proof Suppose not. We will construct a sequence of disjoint measurable sets E_i, such that $\sum_i \nu(E_i)$ is conditionally (but not absolutely) convergent, leading to a contradiction with countable additivity of ν.

Take any measurable set $E \subseteq X$ such that $\nu(E) \geq |\nu(X)| + 1$. Then both quantities $|\nu(E)|$ and $|\nu(X \setminus E)|$ are at least one:

$$\begin{aligned}
|\nu(X \setminus E)| = |\nu(X) - \nu(E)| \geq |\,\nu(E)| - |\,\nu(X)| \geq \nu(E) - |\nu(X)| \geq 1 \\
|\nu(E)| \geq \nu(E) \geq |\nu(X)| + 1 \geq 1\,.
\end{aligned}$$

By hypothesis, one of E or $X \setminus E$ has subsets of arbitrarily large positive measure — say $X \setminus E$. Set $E_1 = E$, and $|\nu(E_1)| > 1$.

Repeat the above argument with X replaced by E_1, to obtain a measurable set $E_2 \subseteq X \setminus E_1$ (so E_2 is disjoint from E_1) such that $|\nu(E_2)| > 1$. Repeat again and again. Continuing this way, we obtain a sequence of disjoint measurable sets $E_1, E_2, \ldots$ such that $|\nu(E_i)| > 1$. ∎

Not only can we partition the space X into its positive and negative part, we also aim to partition the signed measure itself into two or more pieces that, in some way, "disjoint", "independent" or "perpendicular" from one another. We formalize this heuristic in the next definition.

8.1.7 DEFINITION (MUTUALLY SINGULAR MEASURES)
Two signed measures μ and ν are **(mutually) singular**, denoted by $\mu \perp \nu$, if there is a partition $\{E, F = X \setminus E\}$ of X, such that:

- ☞ μ lives on E, that is, F is null for μ (see Definition 8.1.4);
- ☞ ν lives on F, that is, E is null for ν.

8.1.8 EXAMPLE. Counting measure on $\mathbb{N} \subseteq \mathbb{R}$, and Lebesgue measure on $\mathbb{R}$ are mutually singular measures.

8.1.9 THEOREM (JORDAN DECOMPOSITION)
A signed measure ν can be uniquely written as a difference $\nu = \nu^+ - \nu^-$ of two positive measures that are singular to each other.

Proof Let P be a positive set for ν from the Hahn decomposition, and $N = X \setminus P$ be the negative set. For all measurable sets $E \subseteq X$, set:

$$\nu^+(E) = \nu(E \cap P), \quad \nu^- = -\nu(E \cap N).$$

It is plainly evident that ν^+ and ν^- are positive measures, $\nu^+ \perp \nu^-$, and $\nu = \nu^+ - \nu^-$.

Establishing uniqueness is straightforward. Suppose ν is a difference $\nu = \mu^+ - \mu^-$ of two other positive measures with $\mu^+ \perp \mu^-$. By definition there is some set P' that μ^+ lives on, and μ^- lives on the complement N'. Then for any measurable E,

$$\nu(E \cap P') = \mu^+(E \cap P'), \quad \nu(E \cap N') = -\mu^-(E \cap N'),$$

and this means P' is positive for ν and N' is negative for ν. But the Hahn decomposition is unique, so P, P' and N, N' differ on ν-null sets. Thus $\nu^+(E) = \nu(E \cap P) = \nu(E \cap P') = \mu^+(E \cap P') = \mu^+(E)$, so $\nu^+ = \mu^+$. And likewise $\nu^- = \mu^-$. ∎

8.1.10 DEFINITION (VARIATIONS OF SIGNED MEASURE)
For any signed measure ν, the positive measures ν^+ and ν^- in its Jordan decomposition (Theorem 8.1.9), are called, respectively, the **positive variation** and the **negative variation** of ν.

The positive measure $|\nu| = \nu^+ + \nu^-$ is called the **(total) variation** of ν.

8.1.11 REMARK (INTRINSIC DEFINITION OF VARIATIONS OF SIGNED MEASURE). "Intrinsic" definitions of ν^+, ν^- and $|\nu|$ that do not explicitly depend on Theorem 8.1.9 can be given:

$$\begin{aligned} \nu^+(E) &= \sup \left\{ \nu(F) \mid F \subseteq E \text{ measurable} \right\} . \\ -\nu^-(E) &= \inf \left\{ \nu(F) \mid F \subseteq E \text{ measurable} \right\} . \\ |\nu|(E) &= \sup \left\{ \sum_{i=1}^{n} |\nu(E_i)| \mid \text{the measurable sets } E_1, \dots, E_n \text{ partition } E \right\} . \end{aligned}$$

(In the definition of $|\nu|$, countable partitions can be used in place of finite partitions.)

8.1.12 DEFINITION (INTEGRAL WITH RESPECT TO SIGNED MEASURE)
Let ν be a signed measure on X. A function $f \colon X \to \mathbb{R}$ is **integrable** with respect to ν if f is measurable and $|f|$ is integrable with respect to $|\nu|$. The integral of such f with respect to ν is defined by:

$$\int_X f \, d\nu = \int_X f \, d\nu^+ - \int_X f \, d\nu^- .$$

8.2 Radon-Nikodým decomposition

Recall our trusty example, that for any function g integrable with respect to a positive measure μ,

$$\nu(E) = \int_E g \, d\mu$$

generates a new signed measure ν. This ν has the notable property described in the following definition.

8.2.1 DEFINITION (ABSOLUTE CONTINUITY OF MEASURES)
A positive or signed measure ν on some measurable space is **absolutely continuous** with respect to a positive measure μ on the same measurable space, if

☞ Any measurable set E having μ-measure zero also has ν-measure zero.

ν is also said to be **dominated by** μ, symbolically indicated by $\nu \ll \mu$.

In this section, we show the converse: if ν is dominated by μ, then ν must be generated by integration of some density function for μ.

8.2.2 REMARK. Absolute continuity is the polar opposite of mutual singularity (Definition 8.1.7). If $\nu \ll \mu$ and $\nu \perp \mu$ at the same time, then ν must be identically zero.

8.2.3 THEOREM (POSITIVE FINITE CASE OF RADON-NIKODÝM DECOMPOSITION)
If ν and μ be positive finite measures on a measurable space X, then ν may be decomposed as a sum $\nu = \nu_a + \nu_s$ of two positive measures, with $\nu_a \ll \mu$ and $\nu_s \perp \mu$.

Moreover, ν_a has a density function with respect to μ.

Proof We present this slick argument, using Hilbert spaces, due to John von Neumann. The argument can be motivated by the fact that, Hilbert spaces have a general theory of orthogonal decompositions, that can be readily applied, if we can manage to find an inner product describing the desired orthogonality relation.

Let $\lambda = \nu + \mu$ be the "master" measure. We can consider $\mathbf{L}^2(\lambda)$, which can be made into a real Hilbert space with the inner product $\langle f, g \rangle = \int fg \, d\lambda$. The map $f \mapsto \int f \, d\nu$ is a bounded linear functional on $\mathbf{L}^2(\lambda)$, so by the Riesz representation theorem, there exists a $g \in \mathbf{L}^2(\lambda)$ such that

$$\int_X f \, d\nu = \langle f, g \rangle = \int_X fg \, d\lambda \,.$$

Rearranging terms we have

$$\int_X f \cdot (1 - g) \, d\nu = \int_X fg \, d\mu \,. \tag{❀}$$

From this we can see that $0 \leq g \leq 1$ must hold λ-almost everywhere. For if we take $f = \mathbb{I}(\{g \geq 1\})$, the left side is ≤ 0 while the right side is ≥ 0. Similarly, we have the reverse situation if we take $f = \mathbb{I}(\{g \leq 0\})$.

For simplicity, we modify g so that $0 \leq g \leq 1$ *everywhere*. For measurable E, let

$$\nu_a(E) = \nu(E \cap \{g < 1\}), \quad \nu_s(E) = \nu(E \cap \{g = 1\}) \,.$$

Obviously $\nu = \nu_a + \nu_s$. By taking $f = \mathbb{I}(E \cap \{g < 1\})/(1 - g)$ in equation ❀, we have

$$\int_{E \cap \{g<1\}} \frac{g}{1-g} \, d\mu = \int_{E \cap \{g<1\}} \frac{1}{1-g} \cdot (1 - g) \, d\nu = \nu_a(E) \,.$$

So $\nu_a \ll \mu$, and ν_a has density function $\mathbb{I}(\{g < 1\}) \, g/(1 - g)$.

Moreover, if we take $f = \mathbb{I}(E \cap \{g = 1\})$, from equation ❀ we get

$$0 = \int_{E \cap \{g=1\}} (1 - g) \, d\nu = \int_{E \cap \{g=1\}} g \, d\mu = \mu(E \cap \{g = 1\}) \,,$$

so μ is null on $\{g = 1\}$. And clearly ν_s is null on the complement of $\{g = 1\}$. Therefore $\nu_s \perp \mu$. ■

The proof Theorem 8.2.3 contains the real meat of this section; the rest is mostly book-keeping:

8.2.4 Theorem (Radon-Nikodým decomposition)

Let μ be a σ-finite measure (Definition 5.2.4) on a measurable space X. Any other signed measure or σ-finite positive measure ν on X can be uniquely decomposed as $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mu$ and $\nu_s \perp \mu$. Moreover, ν_a has a density function with respect to μ.

Proof

① We dispose of the σ-finite case first, assuming that ν is positive. Let $\{X_n\}$ be a countable partition of X such that $\nu^n(E) = \nu(E \cap X_n)$ and $\mu^n(E) = \mu(E \cap X_n)$ are finite measures. Apply Theorem 8.2.3 to ν^n and μ^n to obtain:

$$\nu^n = \nu_a^n + \nu_s^n\,, \quad \nu_a^n \ll \mu^n\,, \quad \nu_s^n \perp \mu^n\,.$$
$$\nu = \nu_a + \nu_s\,, \quad \nu_a = \sum_{n=1}^{\infty} \nu_a^n\,, \quad \nu_s = \sum_{n=1}^{\infty} \nu_s^n\,.$$

$\nu_a \ll \mu$ holds, because $\mu(E) = 0$ implies $\mu^n(E) = 0$ for all n, and hence $\nu_a^n(E) = 0$ and $\nu_a(E) = 0$.

Let μ^n and ν_s^n live on disjoint sets A_n and B_n respectively. We may assume that $A_n, B_n \subseteq X_n$. Then μ lives on $\bigcup_n A_n$, and ν_s lives on $\bigcup_n B_n$; these two unions are disjoint, so $\nu_s \perp \mu$.

② Now let ν be any signed measure. Split ν into its positive and negative variations (Definition 8.1.10) and apply the positive version of this theorem to each separately:

$$\nu^+ = \nu_a^+ + \nu_s^+\,, \quad \nu_a^+ \ll \mu\,, \quad \nu_s^+ \perp \mu\,.$$
$$\nu^- = \nu_a^- + \nu_s^-\,, \quad \nu_a^- \ll \mu\,, \quad \nu_s^- \perp \mu\,.$$
$$\nu = \nu^+ - \nu^- = \underbrace{(\nu_a^+ - \nu_a^-)}_{\nu_a} + \underbrace{(\nu_s^+ - \nu_s^-)}_{\nu_s}\,.$$

That $\nu_a \ll \mu$ is clear. Suppose $\nu_s^\pm$ lives on $B^\pm$, and μ lives on its complement $A^\pm$. Thus ν_s lives on $B^+ \cup B^-$, and μ lives on $A^+ \cap A^-$ too. But $X \setminus (B^+ \cup B^-) = A^+ \cap A^-$; therefore $\nu_s \perp \mu$.

③ In all cases, ν_a has a density with respect to μ, because the density of a sum or difference of measures is just the sum or difference of the individual densities.

④ Finally, we show uniqueness. If ν has two decompositions $\nu_a + \nu_s$ and $\nu_a' + \nu_s'$, then $\nu_a - \nu_a' = \nu_s' - \nu_s$. The left-hand side of this equation is $\ll \mu$, while the right-hand side is $\perp \mu$. Then both sides must be identically zero (Remark 8.2.2). (If ν can take infinite values, apply this argument after restricting the measures to each X_n as above.) ∎

8.2.5 Example (Discrete measures). An **atomic measure** or **discrete measure** is one takes takes non-zero values only on a discrete set of points. (If it is finite or σ-finite, that discrete set must be countable.)

Any atomic measure ν is mutually singular to the Lebesgue measure λ, so the Radon-Nikodým decomposition of $\sigma = \nu + \lambda$ with respect to λ will be given by $\sigma_a = \lambda$ and $\sigma_s = \nu$.

8.2.6 EXAMPLE (SINGULARLY CONTINUOUS MEASURES). There are also measures singular to Lebesgue measure yet are not atomic. A surface measure (section 7.3) on some set $M \subseteq \mathbb{R}^2$ is an obvious example of such **singularly continuous** measures. On $\mathbb{R}$, the uniform measure on the Cantor set is singularly continuous.

8.3 Exercises

8.1 (Properties of variations of signed measure) Show the following basic properties for a signed measure ν:

① $\nu^+ = \frac{1}{2}(|\nu| + \nu)$.

② $\nu^- = \frac{1}{2}(|\nu| - \nu)$.

③ $|\nu(E)| \leq |\nu|(E)$.

④ $|\nu + \mu| \leq |\nu| + |\mu|$. ($\mu$ is another signed measure.)

8.2 (Intrinsic definitions of variations of signed measure) Demonstrate the equivalence of the intrinsic definitions (Remark 8.1.11) of the positive, negative, and total variations of a signed measure with the definitions in terms of the Jordan decomposition.

Also try showing directly that the formulas for the intrinsic definitions do yield positive measures.

8.3 (Variation of signed measure expressed with densities) For any signed measure ν and σ-finite positive measure μ,

$$\frac{d\nu^\pm}{d\mu} = \left(\frac{d\nu}{d\mu}\right)^\pm, \quad \frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|;$$

also

$$\nu \ll |\nu| \quad \text{and} \quad \frac{d\nu}{d|\nu|} = \pm 1 \text{ almost everywhere.}$$

8.4 (Chain rule for Radon-Nikodým derivative) Assume σ-finiteness.

① If $\nu \ll \mu \ll \lambda$, then λ-almost everywhere

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

② If $\nu \ll \mu \ll \nu$ (the measures ν and μ are said to be **equivalent**), then

$$\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}.$$

8.5 (Integration with a signed measure) Integration with respect to a signed measure ν can also be defined directly instead of reducing it to integrals with respect to ν^+ and ν^-.

Follow these steps.

① Define $\int \varphi \, d\nu$ for simple functions $\varphi \colon X \to \mathbb{R}$ by simple summation, and prove its basic properties.

② Show:

$$\left| \int \varphi \, d\nu \right| \leq \int |\varphi| \, d|\nu| \,.$$

③ We say that a measurable function $f \colon X \to \overline{\mathbb{R}}$ is *integrable* with respect to a signed measure ν if $\int |f| \, d|\nu| < \infty$.

A function $f \colon X \to \overline{\mathbb{R}}$ is integrable with respect to ν if and only if there exist simple functions φ_n converging to f in $\mathbf{L}^1(|\nu|)$. For any such sequence φ_n converging to f, define:

$$\int f \, d\nu = \lim_{n \to \infty} \int \varphi_n \, d\nu \,.$$

Show the limit is well-defined.

④ Show this version of the dominated convergence theorem holds for the new integral:

If a sequence of measurable functions $f_n \colon X \to \overline{\mathbb{R}}$ has limit f, and $|f_n| \leq g$ for some $g \in \mathbf{L}^1(|\nu|)$, then

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

8.6 (σ-finite is necessary for Radon-Nikodým decomposition) Find counterexamples to the conclusion of the Radon-Nikodým theorem if one of the measures is not σ-finite.

Chapter 9

Approximation of Borel sets and functions

9.1 Regularity of Borel measures

As we have said, we must now approximate arbitrary Borel sets $B \in \mathscr{B}(\mathbb{R}^n)$ by compact sets. (We will also need approximation by open sets.) It turns out that this part of the proof is purely topological, and generalizes to other metric spaces X besides $\mathbb{R}^n$. Henceforth we consider the more general case.

Let d denote the metric for the metric space X.

9.1.1 THEOREM
Let $(X, \mathscr{B}(X), \mu)$ be a finite measure space, and let $B \in \mathscr{B}(X)$. For every $\varepsilon > 0$, there exists a closed set V and an open set U such that $V \subseteq B \subseteq U$ and $\mu(U \setminus V) < \varepsilon$.

Proof Let $\mathcal{M}$ be the set of all $B \in \mathscr{B}(X)$ for which the statement is true. We show that $\mathcal{M}$ is a sigma algebra containing all the open sets in X.

① Let $B \in \mathcal{M}$ with V and U as above. Then V^c open $\supseteq B^c \supseteq U^c$ closed, and $\mu(V^c) - \mu(U^c) < \varepsilon$. This shows $B^c \in \mathcal{M}$.

② Let $B_n \in \mathcal{M}$. Choose V_n and U_n for each B_n such that $\mu(U_n \setminus V_n) < \varepsilon/2^n$. Let $U = \bigcup_n U_n$ which is open, and $V = \bigcup_n V_n$, so that $V \subseteq \bigcup_n B_n \subseteq U$.

Of course V is not necessarily closed, but $W_N = \bigcup_{n=1}^N V_n$ are, and these W_N increase to V. Hence $\mu(V \setminus W_N) \to 0$ as $N \to \infty$, meaning that for large enough N, $\mu(V \setminus W_N) < \varepsilon$.

Next, we have

$$
\begin{aligned}
U \setminus W_N &= (U \setminus V) \uplus (V \setminus W_N) \\
&\subseteq \bigcup\nolimits_n (U_n \setminus V_n) \cup (V \setminus W_N)\,, \\
\mu(U \setminus W_N) &= \mu(U \setminus V) + \mu(V \setminus W_N) \\
&\leq \sum\nolimits_n (U_n \setminus V_n) + \mu(V \setminus W_N) < \varepsilon + \varepsilon\,.
\end{aligned}
$$

■

This shows that $\bigcup_n B_n \in \mathcal{M}$.

③ Let B be open, and $A = B^c$. Also let $d(x, A) = \inf_{y \in A} d(x, y)$ be the distance from $x \in X$ to A. Set $D_n = \{x \in X : d(x, A) \geq 1/n\}$. D_n is closed, because $d(\cdot, A)$ is a continuous function, and $[1/n, \infty]$ is closed.

Clearly $d(x, A) \geq 1/n > 0$ implies $x \in A^c = B$, but since A is closed, the converse is also true: for every $x \in A^c = B$, $d(x, A) > 0$. Obviously the D_n are increasing, so we have just shown that they in fact increase to B. Hence $\mu(B \setminus D_n) < \varepsilon$ for large enough n. Thus $B \in \mathcal{M}$.

The case that μ is not a finite measure is taken care of, as you would expect, by taking limits like we did for sigma-finite measures in Section 5.1. But since compact and open sets are involved, we need stronger hypotheses:

① There exists $\{K_n\} \nearrow X$, with K_n compact and $\mu(K_n) < \infty$.

② There exists $\{X_n\} \nearrow X$, with X_n open and $\mu(X_n) < \infty$.

It is easily seen that these properties are satisfied by $X = \mathbb{R}^n$ and the Lebesgue measure λ, as well as many other "reasonable" measures μ on $\mathscr{B}(\mathbb{R}^n)$. We will discuss this more later.

We assume henceforth that X and μ have the properties just listed.

9.1.2 THEOREM

Let $B \in \mathscr{B}(X)$ with $\mu(B) < \infty$. For every $\varepsilon > 0$, there exists a compact set V and an open set U such that $K \subseteq B \subseteq U$ and $\mu(U \setminus K) < \varepsilon$.

Proof It suffices to show that $\mu(U \setminus B) < \varepsilon$ and $\mu(B \setminus K) < \varepsilon$ separately.

Existence of K. Since $\{B \cap K_n\} \nearrow B$, there exists some n such that $\mu(B) - \mu(B \cap K_n) < \varepsilon/2$.

For this n, define the finite measure $\mu_{K_n}(E) = \mu(E \cap K_n)$, for $E \in \mathscr{B}(X)$. By Theorem 9.1.1, there are sets $V \subseteq B \subseteq U$, V closed, and $\mu_{K_n}(B \setminus V) \leq \mu_{K_n}(U \setminus V) < \varepsilon/2$. Since K_n is compact, it is closed. Then $K = V \cap K_n$ is also closed, and hence compact, because it is contained in the compact set K_n. We have,

$$\begin{aligned}\mu(B \setminus K) &= \mu(B) - \mu(B \cap K_n) + \mu(B \cap K_n) - \mu(K) \\ &= \mu(B) - \mu(B \cap K_n) + \mu_{K_n}(B) - \mu_{K_n}(V) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.\end{aligned}$$

Existence of U. For every n, define the finite measure $\mu_{X_n}(E) = \mu(E \cap X_n)$, for $E \in \mathscr{B}(X)$. By Theorem 9.1.1, there are sets $V_n \subseteq B \subseteq U_n$, U_n open, and $\mu_{X_n}(U_n \setminus B) \leq \mu_{X_n}(U_n \setminus V_n) < \varepsilon/2^n$.

Let $U = \bigcup_n U_n \cap X_n \supseteq B$. We have,

$$\mu(U \setminus B) \leq \textstyle\sum_n \mu(U_n \cap X_n \setminus B) = \sum_n \mu_{X_n}(U_n \setminus B) < \varepsilon. \qquad \blacksquare$$

9.1.3 DEFINITION

A Borel measure μ is:

- **inner regular** on a set E if $\mu(E) = \sup\{\mu(K) \mid \text{compact } K \subseteq B\}$.
- **outer regular** on a set E if $\mu(E) = \inf\{\mu(U) \mid \text{open } U \supseteq B\}$.
- **regular** if it is inner regular and outer regular on all Borel sets.

9.1.4 COROLLARY

Let $(X, \mathscr{B}(X), \mu)$ with the same properties as before. Then μ is regular.

Proof If $\mu(B) < \infty$, then inner and outer regularity on B is immediate from Theorem 9.1.2.

Otherwise, $\mu(B \cap X_n) < \infty$, for every n, and so we know there are compact K_n such that $\mu(B \cap X_n) - \mu(K_n \cap X_n) < \varepsilon$. But if $\mu(B \cap X_n)$ are unbounded, then so are $\mu(K_n \cap X_n)$, and $K_n \cap X_n$ is compact. This proves the supremum is unbounded. Obviously the infimum is also unbounded. ■

9.2 $\mathbf{C}_0^\infty$ functions are dense in $\mathbf{L}^p(\mathbb{R}^n)$

This section is devoted to the result that the space of $\mathbf{C}_0^\infty$ functions is dense in $\mathbf{L}^p(\mathbb{R}^n)$, which was discussed at the end of Section 4.1.

9.2.1 THEOREM

Let $f\colon \mathbb{R}^n \to \mathbb{R} \in \mathbf{L}^p(\mathbb{R}^n)$, $1 \le p < \infty$. Then for any $\varepsilon > 0$, there exists $\psi \in \mathbf{C}_0^\infty$ such that

$$\|\psi - f\|_p = \left(\int_{\mathbb{R}^n} |\psi - f|^p \, d\lambda\right)^{1/p} < \varepsilon\,.$$

Our strategy for proving this theorem is straightforward. Since we already know that the simple functions $\varphi = \sum_i a_i \chi_{E_i}$ are dense in $\mathbf{L}^p$, we should try approximating χ_{E_i} by $\mathbf{C}_0^\infty$ functions. Since $\mathbf{C}_0^\infty$ functions are non-zero on compact sets, it stands to reason that we should approximate the sets E_i by compact sets K_i. If this can be done, then it suffices to construct the $\mathbf{C}_0^\infty$ functions on the sets K_i.

Our constructions start with this last step. You might even have seen some of these constructions before.

9.2.2 LEMMA

Let A be a compact rectangle in $\mathbb{R}^n$. Then there exists $\phi \in \mathbf{C}_0^\infty$ which is positive on the interior of A and zero elsewhere.

Proof Consider the infinitely differentiable function

$$f(x) = \begin{cases} e^{-1/x^2}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

If $A = [0,1]$, then $\phi(x) = f(x) \cdot f(1-x)$ is the desired function of $\mathbf{C}_0^\infty$. (Draw pictures!)

If $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$, then we let

$$\phi_A(x) = \phi\left(\frac{x_1 - a_1}{b_1 - a_1}\right) \cdots \phi\left(\frac{x_n - a_n}{b_n - a_n}\right).$$

■

9.2.3 Lemma

For any $\delta > 0$, there exists an infinitely differentiable function $h\colon \mathbb{R} \to [0,1]$ such that $h(x) = 0$ for $x \le 0$ and $h(x) = 1$ for $x \ge \delta$.

Proof Take the function ϕ from Lemma 9.2.2 for the rectangle $[0, \delta]$, and let

$$h(x) = \frac{\int_{-\infty}^{x} \phi(t)\, dt}{\int_{-\infty}^{\infty} \phi(t)\, dt}.$$

■

9.2.4 Theorem

Let U be open, and $K \subset U$ compact. Then there exists $\psi \in \mathbf{C}_0^\infty$ which is positive on K and vanishes outside some other compact set L, $K \subset L \subset U$.

Proof For each $x \in U$, let $A_x \subset U$ be a bounded open rectangle containing x, whose closure $\overline{A_x}$ lies in U. The $\{A_x\}$ together form an open cover of K. Take a finite subcover $\{A_{x_i}\}$. Then the compact rectangles $\{\overline{A_{x_i}}\}$ also cover K.

From Lemma 9.2.2, obtain functions $\psi_i \in \mathbf{C}_0^\infty$ that are positive on $\overline{A_{x_i}}$ and vanish outside $\overline{A_{x_i}}$. Let $\psi = \sum_i \psi_i \in \mathbf{C}_0^\infty$. ψ vanishes outside $L = \bigcup_i \overline{A_{x_i}}$, which is compact. ■

9.2.5 Corollary

In Theorem 9.2.4, it is even possible to require in addition that $0 \le \psi(x) \le 1$ for all $x \in \mathbb{R}^n$ and $\psi(x) = 1$ for $x \in K$.

Proof Let ψ be from Theorem 9.2.4. Since ψ is positive on the compact set K, it has a positive minimum δ there. Take the function h of Lemma 9.2.3 for this δ. The new candidate function is $h \circ \psi$. ■

Proof (Proof of Theorem 9.2.1) Let $\varphi = \sum_i a_i \chi_{E_i}$, $a_i \ne 0$ be a simple function such that $\|\varphi - f\|_p < \varepsilon/2$. Let $\psi_i \in \mathbf{C}_0^\infty$ such that $\|\psi_i - \chi_{E_i}\|_p < \varepsilon/2|a_i|$. (Note that E_i must have finite measure; otherwise φ would not be integrable.) Let $\psi = \sum_i a_i \psi_i$. Then (Minkowski's inequality),

$$\begin{aligned} \|f - \psi\|_p &\le \|f - \varphi\|_p + \|\varphi - \psi\|_p \\ &\le \|f - \varphi\|_p + \sum_i |a_i| \cdot \|\chi_{E_i} - \psi_i\|_p < \varepsilon. \end{aligned}$$

■

Actually, even the last part of theorem can be generalized to spaces other than $\mathbb{R}^n$: instead of infinitely differentiable functions with compact support, we consider continuous functions, defined on the metric space X, with compact support. In this case, a topological argument must be found to replace Lemma 9.2.2. This is easy:

9.2.6 Lemma

Let A be any compact set in X. Then there exists a continuous function $\phi\colon X \to \mathbb{R}$ which is positive on the interior of A and zero elsewhere.

Proof Let $C = X \setminus \text{interior}\, A$, so C is closed. Then $\phi(x) = d(x, C)$ works. ($d(x, C)$ was defined in the proof of Theorem 9.1.1.) ■

The proof of Theorem 9.2.4 goes through verbatim for metric spaces X, provided that X is **locally compact**. This means: given any $x \in X$ and an open neighborhood U of x, there exists another open neighborhood V of x, such that $\overline{V}$ is compact and $\overline{V} \subseteq U$.

Finally, we need to consider when properties (1) and (2) (in the remarks preceding Theorem 9.1.2) are satisfied. These properties are somewhat awkward to state, so we will introduce some new conditions instead.

9.2.7 Definition

A measure μ on a topological space X is **locally finite** if for each $x \in X$, there is an open neighborhood U of x such that $\mu(U) < \infty$.

It is easily seen that when μ is locally finite, then $\mu(K) < \infty$ for *every* compact set K.

9.2.8 Definition

A topological space X is **strongly sigma-compact** if there exists a sequence of open sets X_n with compact closure, and $\{X_n\} \nearrow X$.

If X is strongly sigma-compact, and μ is locally finite, then properties (1) and (2) are automatically satisfied. It is even true that strong sigma-compactness implies local compactness in a metric space. (The proof requires some topology and is left as an exercise.) Then we have the following theorem:

9.2.9 Theorem

Let X be a strongly sigma-compact metric space, and μ be any locally finite measure on $\mathcal{B}(X)$. Then the space of continuous functions with compact support is dense in $\mathbf{L}^p(X, \mathcal{B}(X), \mu)$, $1 \le p < \infty$.

Chapter 10

Relation of integral with derivative

10.1 Differentiation with respect to volume

10.1.1 DEFINITION (AVERAGE OF FUNCTION)
The **average** of a function $f \in \mathbf{L}^1_{\text{loc}}(\mathbb{R}^n)$ over a cube of width $r > 0$ is:

$$A_r f(x) = \frac{1}{\mu(Q(x,r))} \int_{Q(x,r)} f \, d\mu \, .$$

10.1.2 THEOREM
If f is continuous at x, then $A_r f(x) \to f(x)$ as $r \to 0$.

Proof Same as that of the first fundamental theorem of calculus (Theorem 3.5.1). ■

10.1.3 DEFINITION (HARDY-LITTLEWOOD MAXIMAL FUNCTION)
The **Hardy-Littlewood maximal function** is this operator acting on $\mathbf{L}^1_{\text{loc}}(\mathbb{R}^n)$:

$$Hf(x) = \sup_{r>0} A_r f(x) = \sup_{r>0} \frac{1}{\mu(Q(x,r))} \int_{Q(x,r)} |f| \, d\mu \, .$$

10.1.4 THEOREM
$A_r f$ is continuous for each $r > 0$, and Hf is a measurable function (in fact, lower semi-continuous).

Proof Let $y \to x$; then $\mathbb{I}(Q(y,r)) \to \mathbb{I}(Q(x,r))$ pointwise. Since f is integrable, and $\mathbb{I}(Q(y,r))$ is clearly majorized by $\mathbb{I}(Q(x,2r))$ for y close to x, the dominated convergence theorem implies

$$\int_{Q(y,r)} f \, d\mu \to \int_{Q(x,r)} f \, d\mu \, , \quad \mu(Q(y,r)) \to \mu(Q(x,r)) \, .$$

Hence $A_r f(y) \to A_r f(x)$.

Then we also have $\{Hf > \alpha\} = \bigcup_{r>0} \{A_r |f| > \alpha\}$ expressed as a union of open sets; this shows Hf is measurable. ■

10.1.5 LEMMA (CUBIC SUBCOVERING WITH BOUNDED OVERLAP)
Let $A \subseteq \mathbb{R}^n$ be any bounded set, and $\{Q(x, \delta(x)) \mid x \in A\}$ be a collection of open cubes, one for each centre point $x \in A$, with a side length $\delta(x) > 0$ that is selected from a *finite* list $0 < \delta_1 \leq \ldots \leq \delta_m$.

Then there exists a finite subcover of $\{Q(x, \delta(x)\}$ that also covers A, where the cubes from the finite subcover overlap each other at most 2^n times at any point of A.

Proof We select the finite subcover

$$\{Q(x_i, \delta(x_i)) \mid x_i \in A\,, i = 1, \ldots, k\}$$

as follows: assuming that $x_1, \ldots, x_{i-1}$ have already been chosen, choose x_i to be any element of $A \setminus \big(Q(x_1, \delta(x_1)) \cup \cdots \cup Q(x_{i-1}, \delta(x_{i-1}))\big)$, with the largest possible $\delta(x_i)$.

We note that this selection process cannot continue indefinitely. For the points x_i chosen will have the open cubes $Q(x_i, \delta(x_i)/2)$ all disjoint; each of these cubes contributes a volume of at least $(\delta_1/2)^n$. On the other hand, these cubes are contained in the δ_m-neighborhood of A, which is bounded and hence has finite volume. So the finite subcovering by the open cubes will eventually exhaust A.

Finally, we have to show that each $y \in A$ is contained in at most 2^n elements of our finite subcover. Draw the 2^n rectangular quadrants with origin at y. In each quadrant, there is at most one cube from our subcover covering y, whose centre x_i lies in that quadrant. For if there were two cubes, then the cube with the larger width would swallow the centre of the other cube, and this is impossible from the way we have chosen the points x_i. ■

10.1.6 THEOREM (THE MAXIMAL INEQUALITY)
For all integrable functions f and $\alpha > 0$,

$$\mu\{Hf > \alpha\} \leq \frac{2^n}{\alpha} \int_{\mathbb{R}^n} |f| \, d\mu\,.$$

Proof Let $E = \{Hf > \alpha\}$. For each $x \in E$, there exists $r(x) > 0$ such that $A_{r(x)}|f|(x) > \alpha$. Since $A_{r(x)}|f|$ is continuous, there is a neighborhood $U(x)$ of x where

$$\frac{1}{\mu\, Q(y, r(x))} \int_{Q(y, r(x))} |f| \, d\mu = A_{r(x)}|f|(y) > \alpha\,, \quad \text{for } y \in U(x) \text{ too.}$$

If K is any compact subset of E, then these neighborhoods $U(x)$ cover K; select a finite number $U(x_1), \ldots, U(x_m)$ that cover K.

Any $y \in K$ is contained in some $U(x_j)$; pick one, and let $\delta(y) = r(x_j)$. The collection $\{Q(y, \delta(y)) \mid y \in K\}$ satisfies the hypotheses of Lemma 10.1.5. So there exists a finite subcollection of cubes $\{Q(y_i, \delta(y_i)) \mid i = 1, \dots, k\}$ that cover K and overlap at most 2^n times. Then we have

$$\mu(K) \leq \sum_{i=1}^{k} \mu\, Q(y_i, \delta(y_i)) < \sum_{i=1}^{k} \frac{1}{\alpha} \int_{Q(y_i,\delta(y_i))} |f|\, d\mu = \frac{1}{\alpha} \int_{\mathbb{R}^n} |f| \sum_{i=1}^{k} \mathbb{I}(Q(y_i), \delta(y_i))\, d\mu$$
$$\leq \frac{2^n}{\alpha} \int_{\mathbb{R}^n} |f|\, d\mu\,.$$

Since $K \subseteq E$ is arbitrary, we obtain the desired result. ■

10.2 Fundamental theorem of calculus

Chapter 11

Spaces of Borel measures

11.1 Integration as a linear functional

11.2 Riesz representation theorem

11.3 Fourier analysis of measures

11.4 Applications to probability theory

11.5 Exercises

11.1 (Glivenko-Cantelli theorem) Let $X_1, X_2, \ldots$ be independently identically distributed random variables representing observation samples. Let

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(y \geq X_i)$$

be the empirical cumulative distribution functions, and μ_n be the corresponding probability measures. Then P-almost surely, μ_n converges weakly to the pull-back measure $\lambda(B) = \mathrm{P}(X_1 \in B)$. Also, P-almost surely, $F_n(x) \to x$ uniformly in x as $n \to \infty$.

Chapter 12

Miscellaneous topics

12.1 Dual space of L^p

12.2 Egorov's Theorem

The following theorem does not really belong in a first course, but it is quite a surprising and interesting result, and I want to record its proof.

12.2.1 THEOREM (EGOROV)
Let (X, μ) be a measure space of finite measure, and $f_n \colon X \to \mathbb{R}$ be a sequence of measurable functions convergent almost everywhere to f. Then given any $\varepsilon > 0$, there exists a measurable subset $A \subseteq X$ such that $\mu(X \setminus A) < \varepsilon$ and the sequence f_n converges uniformly to f on A.

Proof First define

$$B_{n,m} = \bigcap_{k=n}^{\infty} \left[\, |f_k - f| < \tfrac{1}{m} \,\right].$$

Fix m. For most $x \in X$, $f_n(x)$ converges to $f(x)$, so there exists n such that $|f_k(x) - f(x)| < 1/m$ for all $k \geq n$, so $x \in B_{n,m}$. Thus we see $\{B_{n,m}\}_n \nearrow X \setminus C$, C being some set of measure zero.

We construct the set A inductively as follows. Set $A_0 = X \setminus C$. For each $m > 0$, since $\{A_{m-1} \cap B_{n,m}\}_n \nearrow A_{m-1}$, we have $\mu(A_{m-1} \setminus B_{n,m}) \to 0$, so we can choose $n(m)$ such that

$$\mu(A_{m-1} \setminus B_{n(m),m}) < \frac{\varepsilon}{2^m}.$$

Furthermore set

$$A_m = A_{m-1} \cap B_{n(m),m}.$$

Since $A_m \uplus (A_{m-1} \setminus B_{n(m),m}) = A_{m-1}$, we have

$$\begin{aligned} \mu(A_m) &> \mu(A_{m-1}) - \frac{\varepsilon}{2^m} \\ &> \mu(X) - \frac{\varepsilon}{2} - \frac{\varepsilon}{4} - \cdots - \frac{\varepsilon}{2^m} \geq \mu(X) - \varepsilon. \end{aligned}$$

The sets A_m are decreasing, so letting

$$A = \bigcap_{m=1}^{\infty} A_m = \bigcap_{m=1}^{\infty} B_{n(m),m} \,,$$

we have $\mu(A) \geq \mu(X) - \varepsilon$, or $\mu(X \setminus A) \leq \varepsilon$. Finally, for $x \in A$, $x \in B_{n(m),m}$ for all m, showing that $|f(x) - f_k(x)| < 1/m$ whenever $k \geq n(m)$. This condition is uniform for all $x \in A$. ∎

12.3 Integrals taking values in Banach spaces

We give an introduction to the Lebesgue integral generalized to vector functions and vector measures. The vector spaces may be infinite-dimensional; to have vector analogues of fundamental results such as the Lebesgue Dominated Convergence Theorem, we will require that a norm is available — so we assume[1] the vector spaces are Banach spaces.

First, we consider vector-valued functions $f \colon \Omega \to X$, where (Ω, Σ) is a measure space and X is a Banach space, but keep the measure μ to be a real positive measure.

The measurability of vector-valued functions is defined the same way as for real-valued functions. Since there is no concept of "infinity" in a vector space in general, when integrating, we restrict attention to functions f such that $\int_\Omega \|f\| \, d\mu$ is finite. (Note that this is an real-valued integral we have already defined.)

However, if X is infinite-dimensional, even the restriction $\int_\Omega \|f\| \, d\mu < \infty$ does not suffice. First, note that since X has no order relation, Definition 2.4.5 does not carry over; we are forced to define $\int f \, d\mu$ as the limit of $\int \varphi_n \, d\mu$ for simple functions φ_n, which we *can* compute using basic vector addition and scalar multiplication. However, Theorem 2.4.10 fails; there does not seem to be a way (outside of simple spaces such as $\mathbb{R}^n$) to obtain the simple functions φ_n that converge to f. Therefore, we postulate this property:

12.3.1 DEFINITION

Let (Ω, Σ) be a measurable space, and $(X, \mathcal{B}(X))$ be a Banach space together with the Borel measure. A function $f \colon \Omega \to X$ is **strongly Σ-measurable** if there exists a sequence of Σ-measurable simple functions $\varphi_n \colon \Omega \to X$ that converge to f pointwise.

12.3.2 PROPOSITION

If $f_n \colon \Omega \to X$ is a sequence of measurable functions convergent pointwise to $f \colon \Omega \to X$, then f is measurable.

[1] Although a weaker analogue of Lebesgue integration is also available for topological vector spaces.

Proof We prove that $f^{-1}(U)$ is measurable in Ω for every open set U in X.

Let $U_k = \{x \in X \colon d(x, X \setminus U) > 1/k\}$, an open set. Evidently, the countable union of all the sets U_k is U.

Given $x \in X$ and $\omega \in \Omega$, the statement $x = f(\omega) = \lim_n f_n(\omega)$ implies that $\{f_n(\omega)\}_{n \geq N}$ lies inside the ball $B(x; 1/k)$ when N is large enough. Since $B(x; 1/k) \subseteq U_k$ for any $x \in X$, we have

$$f^{-1}(U_k) = \Big\{\omega \in \Omega \colon \lim_{n \to \infty} f_n(\omega) \in U_k\Big\} \subseteq \liminf_{n \to \infty} f_n^{-1}(U_k)\,.$$

(For a sequence of sets $A_n \subseteq \Omega$, the set $\liminf_n A_n$ is defined to be the set of all $\omega \in \Omega$ such that ω is in A_n when n is large enough. Formally, $\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$.)

Moreover, if $\omega \in \Omega$ is such that $f_n(\omega) \in U_k$, then $f(\omega) = \lim_n f_n(\omega) \in \overline{U_k} \subseteq U$. Therefore,

$$f^{-1}(U) = \bigcup_{k=1}^{\infty} f^{-1}(U_k) \subseteq \bigcup_{k=1}^{\infty} \liminf_{n \to \infty} f_n^{-1}(U_k) \subseteq f^{-1}(U)\,,$$

and the expression in the middle shows $f^{-1}(U)$ is measurable. ∎

12.3.3 COROLLARY

A strongly measurable function is measurable.

Proof By definition, a strongly-measurable function is the pointwise limit of measurable simple functions. ∎

There is another observation about a strongly measurable function f. If φ_n are simple functions convergent to f, then the range of f must be a separable subspace of X. That is, $f(\Omega) \subseteq \overline{D}$ for a *countable* set D — in fact, D can be taken as $\{\varphi_n(\omega) \colon \omega \in \Omega, n \in \mathbb{N}\}$.

Even more interestingly, if we examine the proof of Theorem (2.4.10), we see that it essentially is using the fact that the dyadic rationals (countable in number) are dense in $\mathbb{R}$. So we speculate in general:

12.3.4 THEOREM

A function $f \colon \Omega \to X$ is strongly measurable if and only if it is measurable and its range $f(\Omega)$ is separable in X.

Proof We already know that the "only if" direction is true. For the "if" direction, let $D = \{y_n\}_{n \in \mathbb{N}}$ be a countable set such that $f(\Omega) \subseteq \overline{D}$, and define the functions $\phi_N \colon X \to D \subseteq X$ by the following prescription. For $x \in X$, let $\phi_N(x) = y_k$ where k is the smallest number that minimizes the distance $\|y_k - x\|$ amongst the numbers $1 \leq k \leq N$.

For every $x \in \overline{D}$, we have $\lim_N \phi_N(x) = x$:

$$\|\phi_N(x) - x\| = \min_{0 \leq n \leq N} \|x - y_n\| \searrow 0 \quad \text{as } N \to \infty,$$

because $\{y_n\}$ is dense in $\overline{D}$.

To show that ϕ_N is measurable, it suffices to show that $\phi_N^{-1}(\{y_k\})$ is measurable for $1 \le k \le N$, since ϕ_N takes on only the values $y_1, \dots, y_N$. This just involves translating the description of ϕ_N into formal symbols:

$$[\phi_N = y_k] = \bigcap_{j=1}^{k-1} \Big\{ x \in X \colon \|x - y_k\| < \|x - y_j\| \Big\} \cap \bigcap_{j=k+1}^{N} \Big\{ x \in X \colon \|x - y_k\| \le \|x - y_j\| \Big\} .$$

The above expression describes a Borel set.

Finally, if we set $f_N = \phi_N \circ f$, then $f_N \to f$ pointwise, and $f_N \colon \Omega \to D$ are measurable simple functions. ■

Although we only required pointwise convergence in our definitions, we can automatically obtain the sort of convergence required for the Dominated Convergence Theorem:

12.3.5 THEOREM

A function $f \colon \Omega \to X$ is strongly measurable if and only if there exists a sequence of measurable simple functions $\varphi_n \colon \Omega \to X$ such that $\varphi_n \to f$ and $\|\varphi_n\| \le \|f\|$ pointwise.

Proof Suppose $f_n \colon \Omega \to X$ is a sequence of measurable simple functions that converge to the measurable function f. Let $g_n \colon \Omega \to \mathbb{R}$ be a sequence of measurable simple functions that *increase* to the measurable function $\|f\|$. Then

$$\varphi_n = \begin{cases} \|f_n\|^{-1} g_n \, f_n \,, & f_n \neq 0 \\ 0 \,, & f_n = 0 \end{cases}$$

are measurable simple functions that converge to f pointwise, and $\|\varphi_n\| = g_n \le \|f\|$ for all n. ■

Now we are ready to define vector integration proper:

12.3.6 DEFINITION

Let (Ω, μ) be a measure space, and X be a Banach space. The space $\mathbf{L}^1(\mu, X)$ of **integrable functions** consists of all strongly μ-measurable functions $f \colon \Omega \to X$ such that $\int_\Omega \|f\| \, d\mu < \infty$.

12.3.7 DEFINITION

For a simple function $\varphi \in \mathbf{L}^1(\mu, X)$, its integral is defined by the obvious linear combination

$$\int_\Omega \varphi \, d\mu = \sum_{k=1}^{n} x_k \, \mu(E_k) \,, \quad \varphi = \sum_{k=1}^{n} x_k \, \chi_{E_k} \,, \quad x_k \in X \,,$$

provided all $\mu(E_k)$ are finite quantities.

It follows as before that this integral is well defined, it is linear, and satisfies the triangle inequality.

12.3.8 DEFINITION
To integrate a general function $f \in \mathbf{L}^1(\mu, X)$, compute

$$\int_\Omega f\, d\mu = \lim_{n\to\infty} \int_\Omega \varphi_n\, d\mu$$

using any sequence of simple functions $\varphi_n \in \mathbf{L}^1(\mu, X)$ which converges to f in $\mathbf{L}^1(\mu, X)$ — that is, $\int_\Omega \|\varphi_n \to f\|\, d\mu \to 0$ as $n \to \infty$.

The sequence of simple functions φ_n in this definition can be obtained from Theorem 12.3.5. The Dominated Convergence Theorem for real-valued functions and integrals shows that sequence converges to f in $\mathbf{L}^1$. And the limit of $\int \varphi_n$ exists, for

$$\left\| \int \varphi_n\, d\mu - \int \varphi_m\, d\mu \right\| \le \int \|\varphi_n - \varphi_m\|\, d\mu \to 0\,, \quad \text{as } n, m \to \infty$$

($\|\varphi_n - \varphi_m\| \to 0$ pointwise with a dominating factor $2\|f\|$). Thus the sequence $\int \varphi_n$ is a Cauchy sequence and converges in the Banach space X. Similarly, if φ_n' were another sequence satisfying the same conditions,

$$\left\| \int \varphi_n\, d\mu - \int \varphi_n'\, d\mu \right\| \le \int \|\varphi_n - f\|\, d\mu + \int \|\varphi_n' - f\|\, d\mu \to 0\,,$$

so the value of the integral is the same regardless of which sequence is used to compute it.

The vector-valued integral satisfies linearity and the generalized triangle inequality, and these assertions are easily proven by taking limits from simple functions.

12.3.9 THEOREM
Let (Ω, μ) be a measure space, and X and Y be Banach spaces. If $T\colon X \to Y$ is a continuous linear mapping, and $f \in \mathbf{L}^1(\mu, X)$, then $T \circ f \in \mathbf{L}^1(\mu, Y)$ and

$$T\left(\int_\Omega f\, d\mu\right) = \int_\Omega T \circ f\, d\mu\,.$$

Proof The equation is obvious for simple functions, and get the general case by taking limits. ∎

Appendix A

Results from other areas of mathematics

A.1 Extended real number system

A.1.1 DEFINITION (EXTENDED REAL NUMBER SYSTEM)
The **extended real numbers** is the set

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

consisting of the usual real numbers and the infinite quantities in both the positive and negative direction. The quantity $+\infty$ will often be abbreviated as simply "∞".

Although $\mathbb{R}$ together with the infinities do not form a field (in the algebraic sense), some arithmetic operations with infinities can still be reasonably defined, to be compatible with their usual interpretations of limits of ordinary real numbers:

$$\begin{aligned} -\infty \leq a &\leq \infty\,, \quad a \in \overline{\mathbb{R}}\,. \\ \infty + \infty &= \infty\,. \\ -(\pm\infty) &= \mp\infty\,. \\ a \cdot \infty &= +\infty\,, \quad a > 0\,. \\ a \cdot \infty &= -\infty\,, \quad a < 0\,. \\ a/\infty &= 0\,, \quad a \neq \pm\infty\,. \end{aligned}$$

Elementary calculus textbooks ostensibly avoid defining $\overline{\mathbb{R}}$ as to not confuse students with undefined expressions such as "$\infty - \infty$". These expressions will pose no problem for us, as we simply will have no use for them, and we can just disallow them outright.

But the other rules for $\pm\infty$ are convenient for packaging results without having to analyze separately the bounded and unbounded cases. For example, if λ denotes the area of a subset of the plane, then $\lambda([0, w] \times [0, h]) = wh$; while $\lambda(\mathbb{R}^2) = \infty$.

A.2 Linear algebra

A.3 Multivariate calculus

A.3.1 THEOREM (MEAN VALUE THEOREM)
Let $g\colon U \to \mathbb{R}^m$ be a differentiable map on an open subset $U \subseteq \mathbb{R}^n$. If x, y are points in U, and the line segment joining x and y lies entirely in U, then there exists ξ lying on that line segment for which

$$\|g(x) - g(y)\| \leq \|\mathrm{D}g(\xi)(x-y)\| \leq \|\mathrm{D}g(\xi)\| \, \|x-y\| \,.$$

(The operator norm of any linear transformation T between normed vector spaces is defined by:

$$\|T\| = \sup_{u \neq 0} \frac{\|Tu\|}{\|u\|} = \sup_{\|u\|=1} \|Tu\| \,;$$

the supremum is finite if the domain vector space is $\mathbb{R}^n$, for the unit sphere in $\mathbb{R}^n$ is compact.)

A.3.2 THEOREM (INVERSE FUNCTION THEOREM)
Let $f\colon X \to \mathbb{R}^n$ be a continuous differentiable function on an open set $X \subseteq \mathbb{R}^n$. If $\mathrm{D}g(x)$ is *non-singular* at some $x \in X$, then f maps some open neighborhood of x, $U \subseteq X$, *bijectively* to an open set $f(U)$, and the inverse mapping there is also continuously differentiable.

A.3.3 LEMMA (FACTORIZATION OF DIFFEOMORPHISMS)
Let g be a diffeomorphism of open sets in $\mathbb{R}^n$, for $n \geq 2$. For any point in the domain of g, there is a neighborhood A around that point where g can be expressed as the composition:

$$g|_A = u \circ v \,,$$

of a diffeomorphism u that fixes some $0 < m < n$ coordinates of $\mathbb{R}^n$ and another diffeomorphism v that fixes the other $n - m$ coordinates.

Proof We have to solve the above equation for the appropriate diffeomorphisms $v\colon A \to v(A)$ and $u\colon v(A) \to g(A)$. Let $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{n-m}$ be the coordinate values. Labelling the first batch and second batch of coordinates with the subscripts "1" and "2" respectively, we expand:

$$\begin{aligned}
g(x,y) &= u\big(v(x,y)\big) \\
&= \Big(u_1\big(v_1(x,y), v_2(x,y)\big), u_2\big(v(x,y)\big)\Big) \\
&= \Big(u_1\big(v_1(x,y), y\big), u_2\big(v(x,y)\big)\Big) && \text{since } v \text{ fixes last coords} \\
&= \Big(v_1(x,y), u_2\big(v(x,y)\big)\Big) && \text{since } u \text{ fixes first coords.}
\end{aligned}$$

Or, succinctly:

$$g_1 = v_1\,, \quad g_2 = u_2 \circ v\,.$$

The first equation determines the solution function v trivially. The second equation can be inverted by the inverse function theorem:

$$u_2 = g_2 \circ v^{-1}\,,$$

for

$$\mathrm{D}v(x,y) = \begin{bmatrix} \mathrm{D}_1 g_1(x,y) & \mathrm{D}_2 g_1(x,y) \\ 0 & I \end{bmatrix}\,, \quad \det \mathrm{D}v(x,y) = \det \mathrm{D}_1 g_1(x,y) \neq 0\,.$$

So given a starting point (x_0, y_0) in the domain of g, we can define v^{-1} on some open set B containing $v(x_0, y_0)$. Then take $A = v^{-1}(B)$. ■

A.4 Point-set topology

A.4.1 Theorem (Lindelöf's theorem)
If X is a second-countable topological space, then every open cover of X has a countable subcover.

A.4.2 Corollary
Every open cover of an open subset of $\mathbb{R}^n$ has a countable subcover.

A.5 Functional analysis

A.6 Fourier analysis

Appendix B

Bibliography

First, you need a respectable first-year calculus course, dealing with limits rigorously. The course I took used [Spivak1] (possibly the best math book ever).

You should be at least somewhat familiar with multi-dimensional calculus, if only to have a motivation for the theorems we prove (e.g. Fubini's Theorem, Change of Variables). I learned multi-dimensional calculus from [Spivak2] and [Munkres]. As you'd expect, these are theoretical books, and not very practical, but we will need a few elementary results that these books prove.

Point-set topology is also introduced in the study of multi-dimensional calculus. We will not need a deep understanding of that subject here, but just the basic definitions and facts about open sets, closed sets, compact sets, continous maps between topological spaces, and metric spaces. I don't have particular references for these, as it has become popular to learn topology with Moore's method (as I have done), where you are given lists of theorems that you are supposed to prove alone.

The last book, [Rosenthal], (not a prerequesite) is what I mostly referred to while writing up Section 5.1. It contains applications to probability of the abstract measure stuff we do here, and it is not overly abstract. I recommend it, and it's cheap too.

I don't mention any of the standard real analysis or measure theory books here, since I don't have them handy, and this text is supposed to supplant a fair portion of these books anyway. But surely you can find references elsewhere.

Bibliography

[Spivak1] Michael Spivak. *Calculus*, 3^{rd} edition. Publish or Perish, 1994; ISBN 0-914098-89-6.

[Spivak2] Michael Spivak. *Calculus on Manifolds*. Perseus, 1965; ISBN 0-8053-9021-9.

[Munkres] James R. Munkres. *Analysis on Manifolds*. Westview Press, 1991; ISBN 0-201-51035-9.

[Rosenthal] Jeffrey S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, 2000; ISBN 981-02-4303-0.

[Guzman] Miguel De Guzman. "A Change-of-Variables Formula Without Continuity". *The American Mathematical Monthly*, Vol. 87, No. 9 (Nov 1980), p. 736–739.

[Schwartz] J. Schwartz. "The Formula for Change in Variables in a Multiple Integral". *The American Mathematical Monthly*, Vol. 61, No. 2 (Feb 1954), p. 81–85.

[Ham] M. W. Ham. "The Lecture Method in Mathematics: A Student's View". *The American Mathematical Monthly*, Vol. 61, No. 2 (Feb 1973), p. 195–201.

[Kline] Morris Kline. "Logic versus Pedagogy". *The American Mathematical Monthly*, Vol. 77, No. 3 (Mar 1970), p. 264–282.

[Gamelin] Theodore W. Gamelin, Robert E. Greene. *Introduction to Topology*, 2^{nd} edition. Dover, 1999.

[Folland] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*, 2^{nd} edition. Wiley-Interscience, 1999.

[Zeidler] Eberhard Zeidler. *Applied Functional Analysis: Main Principles and Their Applications*. Springer, 1995.

[Kuo] Hui-Hsiung Kuo. *Introduction to Stochastic Integration*. Springer, 2006.

[Doob] Joseph L. Doob. *Measure Theory*. Springer, 1994.

[Shilov] G. E. Shilov, B. L. Gurevich. *Integral, Measure and Derivative: A Unified Approach*. Trans. Richard A. Silverman. Prentice-Hall, 1966.

[Pesin] Ivan N. Pesin. *Classical and Modern Integration Theories*. Trans. Samuel Kotz. Academic Press, 1970

[Wheedan] Richard L. Wheedan, Antoni Zygmund. *Measure and Integral: An Introduction to Real Analysis*. Marcel Dekker, 1977.

[Bouleau] Nicolas Bouleau, Dominique Lépingle. *Numerical Methods for Stochastic Processes*. Wiley-Interscience, 1994.

Appendix C

List of theorems

2.2.5 Additivity and monotonicity of positive measures . . . 9
2.2.6 Countable subadditivity . . . 9
2.2.7 Continuity from below and above . . . 10
2.3.3 Composition of measurable functions . . . 11
2.3.4 Measurability from generated σ-algebra . . . 11
2.3.7 Measurability of real-valued functions . . . 12
2.3.9 Measurability of limit functions . . . 13
2.3.10 Measurability of positive and negative part . . . 13
2.3.11 Measurability of sum and product . . . 13
2.4.8 Linearity of integral for simple functions . . . 16
2.4.10 Approximation by simple functions . . . 17
2.4.15 Monotonicity of integral . . . 19
2.4.17 Vanishing integrals . . . 20
3.1.1 Monotone convergence theorem . . . 25
3.1.2 Linearity of integral . . . 26
3.1.3 Fatou's lemma . . . 27
3.1.5 Lebesgue's dominated convergence theorem . . . 27
3.2.1 Beppo-Levi . . . 29
3.2.2 Generalization of Beppo-Levi . . . 30
3.3.1 Measures generated by mass densities . . . 30
3.4.1 Change of variables . . . 32
3.4.4 Differential change of variables in $\mathbb{R}^n$ . . . 33
3.5.1 First fundamental theorem of calculus . . . 34
3.5.2 Continuous dependence on integral parameter . . . 35
3.5.3 Differentiation under the integral sign . . . 35
4.1.3 Hölder's inequality . . . 39
4.1.4 Minkowski's inequality . . . 39
4.1.8 Dominated convergence in $\mathbf{L}^p$ . . . 41
5.1.3 Properties of induced outer measure . . . 44
5.1.5 Measurable sets of outer measure form σ-algebra . . . 44
5.1.6 Countable additivity of outer measure . . . 46

5.2.2 Carathéodory extension process . . . 47
5.2.3 Uniqueness of extension for finite measure . . . 47
5.2.6 Uniqueness of extension for σ-finite measure . . . 48
5.4.2 Semi-algebra of rectangles . . . 51
5.5.1 Vitali's non-measurable set . . . 54
6.1.3 Measurability of cross section . . . 60
6.1.4 Measurability of function with fixed variable . . . 60
6.1.5 Generating sets of product σ-algebra . . . 60
6.1.6 Product Borel σ-algebra . . . 61
6.2.2 Monotone Class Theorem . . . 61
6.2.3 Measurability of cross-sectional area function . . . 62
6.3.1 Product measure as integral of cross-sectional areas . . . 63
6.3.2 Fubini . . . 63
6.3.3 Tonelli . . . 64
7.1.1 Riemann integrability implies Lebesgue integrability . . . 69
7.1.2 Absolutely-convergent improper Riemann integrals . . . 69
7.2.1 Volume differential . . . 70
8.1.5 Hahn Decomposition . . . 85
8.1.6 Finiteness of total variation . . . 86
8.1.9 Jordan decomposition . . . 87
8.2.3 Positive finite case of Radon-Nikodým decomposition . . . 88
8.2.4 Radon-Nikodým decomposition . . . 89
10.1.5 Cubic subcovering with bounded overlap . . . 99
10.1.6 The maximal inequality . . . 99
12.2.1 Egorov . . . 102
A.3.1 Mean value theorem . . . 108
A.3.2 Inverse function theorem . . . 108
A.3.3 Factorization of diffeomorphisms . . . 108
A.4.1 Lindelöf's theorem . . . 109

Appendix D

List of figures

List of Figures

2.1 Schematic of the counting measure on the integers. 8
2.2 Additivity, monotonicity, and continuity of measures 9
2.3 Approximation by simple functions 18
2.4 Integral of function taking both signs 18
2.5 Symmetric set difference 23

3.1 Witches' hat functions 28
3.2 Volume scaling factor for differential transformations on $\mathbb{R}^n$ 33

Appendix E

Copying and distribution rights to this document

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Index

Beppo-Levi, 29

densities, 30
Dirac delta, 8

multiple integrals, 59