An Introduction to Measure Theory and Integration

Oliver Tse

Adapted from the lecture notes by Jim Portegies

Contents

Preface

Why would one study measure theory?

Measure theory is foundational for both modern probability theory and modern analysis. It is used in Martingale theory, large deviation theory, Ordinary and Partial Differential Equations, Harmonic Analysis, Functional Analysis etc.—all are subjects that you might encounter in later stages of your Mathematics study.

Many systems and processes encountered in Physics can be modeled as differential equations. Since integration and differentiation are inverse operations, in order to analyze or solve such equations (often numerically), it is important to have a powerful theory of *integration*. Measure theory provides a theory of integration, usually called Lebesgue integration.

But in Analysis, we have already seen a theory of integration, namely Riemann integration. Riemann integration has some unfortunate aspects, that are dealt with by Lebesgue integration.

- If a sequence of Riemann integrable functions $f_n : [0,1] \to [0,1]$ converge pointwise to a function $f : [0,1] \to [0,1]$, the function f is not necessarily Riemann integrable. In contrast, Lebesgue-integrability is preserved under pointwise convergence of such functions.
- Within the theory of Riemann integration, it is practically impossible to characterize in a simple way when a function is Riemann integrable. Such a characterization of *Riemann*-integrability is easily given within the theory of Lebesgue integration.

Measure theory provides very useful conditions for when a limit and an integral can be interchanged. Again, if $f_n \to f$ pointwise, the important theorems in this course provide conditions that ensure that

$$\lim_{n\to\infty} \int f_n = \int \left(\lim_{n\to\infty} f_n \right) = \int f.$$

Perhaps this interchange of limit and integration seems obvious, but there are many examples where the limit and integration cannot be interchanged. Some of these examples do have significance in other fields. These theorems become very important for instance when one approximates solutions of differential equations.

In a way, probability theory *is* measure theory (restricted to the study of *unit* measures). If one studies discrete spaces or discrete stochastic processes, the technical aspects of measure theory do not need to surface. But when studying continuous stochastic processes, such as the Brownian motion, things become unnecessarily complicated when one does not use the language of measure theory. Such continuous stochastic processes play a large role also in financial mathematics.

References

There are many excellent sources for learning Measure Theory. One of my personal favorites is *Measure Theory and Fine Properties of Functions* by Evans and Gariepy. It only deals with measure theory on $\mathbb{R}^n$, which made me think it would be a bit restricted as a source for probability theory. I also recommend *Real Analysis* by Royden, which is written in a very clear way, and is very comprehensive. As it comes to measure theory from a probability perspective, I can recommend *Probability with Martingales* by David Williams.

Another nice source that is very readable is *Measure, Integral and Probability* by Capinski and Kopp. In my opinion, this book also manages to strike a nice balance between analysis and probability.

CHAPTER 1

Introduction

1.1. Looking ahead

Among the most important objects of this course are measures. Given a set Ω, a measure μ assigns a nonnegative real number $\mu(A)$ to subsets A of Ω. Depending on the situation, this number may be interpreted as the length, area, volume, measure or probability of this subset. For instance, if $\Omega = [0, 1]$, the unit interval, a natural choice for a measure $\mu((a, b])$ of an interval $(a, b] \subset [0, 1]$ would be $b - a$, which we interpret as its length.

More generally, we can view any measure μ as a set function $\mu : 2^\Omega \to [0, +\infty)$, where it assigns numbers to certain subsets of Ω. Recall that 2^Ω is the notation for the power set of Ω, that is, the set of subsets of Ω. To really capture the idea of length, volume, or probability, the set function μ will have to satisfy additional properties. First of all, $\mu(\emptyset) = 0$. Secondly, it should be *σ-additive* in the following sense.

Definition 1.1.1. A set function $\mu : 2^\Omega \to [0, +\infty)$ is said to be

(i) *finitely additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ for any two disjoint sets $A, B \in 2^\Omega$.
(ii) *σ-additivity* if for any family $\{A_i\}_{i\in\mathbb{N}}$ of mutually disjoint subsets of Ω,

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

If a set function μ satisfies $\mu(\emptyset) = 0$ and is σ-additive, then we call μ a *measure* on $(\Omega, 2^\Omega)$.

Example 1.1.2. Let Ω be a finite set. The set function $\mu_c : 2^\Omega \to \mathbb{N} \cup \{0\}$ assigns to every subset $A \subset \Omega$ its cardinality $\mu_c(A) := \#A$. In other words, it counts the number of elements in the set A. The set function μ_c is called the *counting measure* and is a measure on $(\Omega, 2^\Omega)$. $\diamondsuit$

Example 1.1.3. Let Ω be a set and let $x \in \Omega$. The *Dirac measure* at x is defined as

$$2^\Omega \ni x \mapsto \delta_x(A) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \in \{0, 1\}.$$

The Dirac measure is indeed a measure on $(\Omega, 2^\Omega)$. $\diamondsuit$

Example 1.1.4. Let $(p_i)_{i\in\mathbb{N}} \subset [0, 1]$ be sequence of numbers such that

$$\sum_{i=1}^{\infty} p_i = 1$$

for instance $p_i = 2^{-i}$. Define the set function $\mathbb{P} : 2^{\mathbb{N}} \to [0, 1]$ defined by

$$\mathbb{P}(I) = \sum_{i \in I} p_i = \sum_{i=1}^{\infty} \mathbf{1}_I(i)\, p_i$$

for all subsets $I \subset \mathbb{N}$. Here, $\mathbf{1}_I$ is the *indicator function* of the set I, defined by

$$\mathbf{1}_I(i) := \begin{cases} 1 & \text{if } i \in I \\ 0 & \text{if } i \notin I. \end{cases}$$

We claim that $\mathbb{P}$ is a measure on $(\mathbb{N}, 2^{\mathbb{N}})$ with $\mathbb{P}(\mathbb{N}) = 1$. Since $\mathbb{P}$ satisfies the additional property $\mathbb{P}(\mathbb{N}) = 1$, we say that $\mathbb{P}$ is a *probability measure*.

To show that $\mathbb{P}$ is indeed a measure on $(\mathbb{N}, 2^{\mathbb{N}})$, we should show that $\mathbb{P}(\emptyset) = 0$ and that $\mathbb{P}$ is σ-additive. The proof of the first statement is easy:

$$\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbf{1}_{\emptyset}(i)\, p_i = \sum_{i=1}^{\infty} 0 = 0.$$

To show the σ-additivity, let $\{A_i\}_{i \in \mathbb{N}}$ be a family of mutually disjoint subsets of $\mathbb{N}$ and set

$$A := \bigcup_{j=1}^{\infty} A_j. \tag{1.1.1}$$

We need to show that

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(A_j),$$

that is we need to show that

$$\sum_{i=1}^{\infty} \mathbf{1}_A(i)\, p_i = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i. \tag{1.1.2}$$

From the equality (1.1.1) we know that for every $i \in \mathbb{N}$

$$\sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i) = \mathbf{1}_A(i).$$

So all we need to show is that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i. \tag{1.1.3}$$

In other words, we need to show that we can interchange the order of summation. Such an interchange of sums and limits is a central topic in measure theory.

We will first show the inequality "$\geq$". To that end, let $M, N \in \mathbb{N}$ be arbitrary. Then

$$\sum_{j=1}^{M} \sum_{i=1}^{N} \mathbf{1}_{A_j}(i)\, p_i = \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{i=1}^{N} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i.$$

Since N was arbitrary, we can take the limit $N \to \infty$ and find

$$\sum_{j=1}^{M} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i.$$

Similarly, because M is arbitrary, we can take the limit $M \to \infty$ and find

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i.$$

For the opposite inequality "$\leq$", let again $M, N \in \mathbb{N}$. Then,

$$\sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{1}_{A_j}(i)\, p_i = \sum_{j=1}^{M} \sum_{i=1}^{N} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{j=1}^{M} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i.$$

Since $M \in \mathbb{N}$ was arbitrary, we find

$$\sum_{i=1}^{N} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i \leq \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{A_j}(i)\, p_i$$

and because $N \in \mathbb{N}$ is arbitrary, we conclude the opposite inequality by sending $N \to \infty$. $\diamondsuit$

Rule of thumb. It is common practice in measure theory to first consider objects that are *finite* and to give meaning to *infinite* objects via some limit procedure.

1.2. Looking back

The examples above suggest that we may construct measures by assigning to every point $\omega \in \Omega$ a value $a_\omega \geq 0$, and to define a measure μ by

$$\mu(A) := \sum_{\omega \in A} a_\omega, \ \text{ for } A \subset \Omega.$$

The first problem we run into is when Ω consists of *uncountably* many elements. In this case, we do not yet know how to define the sum. Let us first tackle this problem.

In Analysis I, we said that a series $\sum_{k=1}^{\infty} a_k$ converges, if the sequence of partial sums $S_n := \sum_{k=1}^{n} a_k$ converges, and then we set

$$\sum_{k=1}^{\infty} a_k := \lim_{n \to \infty} S_n.$$

We now would like to define the sum over an arbitrary family $\{a_\alpha\}_{\alpha \in \mathcal{I}} \subset [0, \infty)$ of nonnegative elements. Here $\mathcal{I}$ is just some set that labels the elements in the family, we also say that $\mathcal{I}$ is an index set. We then define

$$\sum_{\alpha \in \mathcal{I}} a_\alpha := \sup \left\{ \sum_{\alpha \in \mathcal{K}} a_\alpha : \ \mathcal{K} \text{ finite subset of } \mathcal{I} \right\}. \tag{1.2.1}$$

1.3. Dealing with infinity

In Analysis, we have always stressed that '$\pm\infty$' is not a number, and you should not treat it as one. For example, we have seen the statement that a set $A \subset \mathbb{R}$ is either bounded from above, in which case by the completeness of $\mathbb{R}$, the supremum of A always exists, or A is not bounded from above and the supremum does not exist.

In measure theory, such a splitting in cases is too tedious. Moreover, when summing or multiplying positive numbers, one actually can give a consistent treatment of the symbol '$\pm\infty$'. More precisely, we add to $\mathbb{R}$ two symbols, '$+\infty$' and '$-\infty'$ to get the extended real number line, which we denote by $\overline{\mathbb{R}}$. We extend the usual arithmetic operations as follows

$$\begin{aligned}
(\pm\infty) + a = a + (\pm\infty) &= \pm\infty && \text{for all } a \in \mathbb{R} \\
a - (\mp\infty) = (\pm\infty) - a &= \pm\infty && \text{for all } a \in \mathbb{R} \\
(\pm\infty) \cdot b = b \cdot (\pm\infty) &= \mp\infty && \text{for all } b \in (-\infty, 0) \\
(\pm\infty) \cdot c = c \cdot (\pm\infty) &= \pm\infty && \text{for all } c \in (0, +\infty) \\
\frac{a}{\pm\infty} &= 0 && \text{for all } a \in \mathbb{R} \\
\frac{\pm\infty}{b} &= \mp\infty && \text{for all } b \in (-\infty, 0) \\
\frac{\pm\infty}{c} &= \pm\infty && \text{for all } c \in (0, +\infty) \\
(+\infty) + (+\infty) = (+\infty) - (-\infty) &= +\infty \\
(+\infty) \cdot (\pm\infty) = (-\infty) \cdot (\mp\infty) &= \pm\infty
\end{aligned}$$

We also extend the order on $\mathbb{R}$ to an order on $\overline{\mathbb{R}}$ by saying that $-\infty < a < +\infty$ for all $a \in \mathbb{R}$. We write $[-\infty, a)$ for $\{-\infty\} \cup (-\infty, a)$ and similarly $(a, +\infty] = (a, +\infty) \cup \{+\infty\}$.

Remark 1.3.1. (i) There are several operations that we do *not* define, such as "$(+\infty) - (+\infty)$" and "$(+\infty)/(+\infty)$" as they cannot be assigned consistent values.

(ii) With these definitions, the usual limit theorems generalize[1], with some care, to limits diverging to $+\infty$ or $-\infty$. For instance, if $a_n, b_n \in (-\infty, \infty]$ are two sequences of extended real numbers, that converge to $a, b \in (-\infty, \infty]$ respectively, then $a_n + b_n$ converges to $a + b$.

In measure theory, it will be very convenient to *define* $0 \cdot (+\infty) := 0$. We do need to be a bit careful here, because of course it is in general not true that if $a_n \to 0$, $b_n \to +\infty$ that then $a_n \cdot b_n \to 0$. So even though we cannot formulate a corresponding limit theorem for the multiplication of sequences of extended real numbers, the definition $0 \cdot (+\infty) := 0$ is so useful for other reasons that we also use it in these lecture notes.

With these definitions, we find that the supremum of every subset of the real line exists as an extended real number. Similarly, the series

$$\sum_{n=1}^{\infty} a_n \in [0, +\infty]$$

[1] In fact, this is part of the motivation of the definitions.

for nonnegative $a_n \in [0, +\infty)$ always exists as an extended real number, and the limit of a monotone sequence always exists as an extended real number.

1.4. Back on track

Let us get back to our idea of constructing a measure μ by

$$\mu(A) \coloneqq \sum_{\omega \in A} a_\omega. \tag{1.4.1}$$

First the good news. The set function μ is indeed a measure (cf. Exercise 1.6.2)

However, Exercise 1.6.3 will show that when we want μ to be a probability measure, then, in fact, we are secretly in a familiar situation of summing over only countably many elements: if μ is a probability measure, then $a_\omega > 0$ for only countably many $\omega \in \Omega$.

1.5. Length measure

Some wishful thinking at this stage might lead us to hope that there exists a measure μ on $(\mathbb{R}, 2^{\mathbb{R}})$ which is *translation invariant* and assigns the value $\mu([0, 2]) = 2$ to the interval $[0, 2] \subset \mathbb{R}$. Yet we cannot have our cake and eat it: such a measure does *not* exist. This nonexistence may be part of the reason why measure theory is a subject of study at all, and that is why already at this stage, we mention it. We will not prove the theorem in class but it is included for the curious souls.

THEOREM 1.5.1. *There is no translation invariant, σ-additive set function $\mu : 2^{\mathbb{R}} \to [0, +\infty]$ satisfing $\mu(\emptyset) = 0$ and $\mu([0, 2]) = 2$.*

The idea of the proof is to assume that such a measure μ exists, to then create a horrible subset F of $(0, 1)$ (so-called *Vitali set*, discovered by Vitali himself) for which it is impossible to consistently define its measure and to derive a contradiction this way.

PROOF OF THEOREM 1.5.1. So suppose such a set function μ does exist. As a preparation for the construction of a horrible subset, we define an equivalence relation $\sim$ on $\mathbb{R}$. Two real numbers $a, b \in \mathbb{R}$ are equivalent, and we write $a \sim b$, if and only if $a - b \in \mathbb{Q}$. We denote the class of equivalence classes by $\mathbb{R}/\sim$. Now *choose*[2] a representative $x_\alpha \in (0, 1)$ from every class $\alpha \in \mathbb{R}/\sim$.

Set $F \coloneqq \{x_\alpha\}_{\alpha \in \mathbb{R}/\sim}$. This is our horrible subset of Ω. In the rest of the proof, we basically show that we cannot give a consistent definition of $\mu(F)$.

We will first rule out the possibility that $\mu(F) = 0$. Clearly, for every $y \in \mathbb{R}$, there exists a unique $q \in \mathbb{Q}$ such that $y \in F + q$. Stated differently, $\mathbb{R}$ is the disjoint union of the sets $F + q$ with $q \in \mathbb{Q}$, that is

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (F + q).$$

and $(F + q) \cap (F + r) = \emptyset$ if $q, r \in \mathbb{Q}$ and $q \neq r$.

[2] Here, one assumes the Axiom of Choice, which states that for every family of sets $(A_\alpha)_{\alpha \in \mathcal{I}}$ there exists a family $(x_\alpha)_{\alpha \in \mathcal{I}}$ such that $x_\alpha \in A_\alpha$ for every $\alpha \in \mathcal{I}$.

By the σ-additivity of μ,

$$\begin{aligned} 2 = \mu([0,2]) &= \sum_{q\in\mathbb{Q}} \mu((F+q)\cap[0,2]) \\ &\le \sum_{q\in\mathbb{Q}} \Big(\mu((F+q)\cap[0,2]) + \mu((F+q)\backslash[0,2])\Big) \qquad (1.5.1)\\ &= \sum_{q\in\mathbb{Q}} \mu(F+q). \end{aligned}$$

Since we assumed that μ is translation invariant, $\mu(F+q)$ is independent of q. If $\mu(F+q)=0$, we would get a contradiction with (1.5.1). So $\mu(F+q)=\mu(F)>0$ for all $q\in\mathbb{Q}$.

We will now show that $\mu(F)>0$ is also impossible. Again, by σ-additivity, we have

$$\begin{aligned} 2 = \mu([0,2]) &= \sum_{q\in\mathbb{Q}} \mu((F+q)\cap[0,2]) \\ &\ge \sum_{q\in(0,1)\cap\mathbb{Q}} \mu((F+q)\cap[0,2]) = \sum_{q\in(0,1)\cap\mathbb{Q}} \mu(F) = \infty, \end{aligned}$$

which is a contradiction. □

The way out is to not assign a measure to *all* subsets of Ω, but only to certain subsets. That is, one just restricts the domain of μ to a collection of subsets of Ω, that we denote by $\mathcal{F}$, that is strictly smaller than 2^Ω. It is useful if this collection $\mathcal{F}$ has some additional properties, namely that it contains the empty set and the whole space, it is closed with respect to the formation of complements, and with respect to the formation of countable unions. In that case, we call $\mathcal{F}$ a σ-algebra. The subsets in $\mathcal{F}$ are called *measurable sets* or, in probability theory, *events*.

Even though from an analysis perspective, a fact such as the nonexistence of a measure $\mu : 2^{[0,1]} \to [0,1]$ such that $\mu((a,b]) = b-a$ may be disappointing, from a probability perspective, it is not that strange to assign probabilities only to certain events, and not to every possible event. Moreover, especially in probability theory, the restriction to certain σ-algebras comes hand-in-hand with many powerful ideas, such as conditional expectations and measures.

1.6. Exercises

EXERCISE 1.6.1. Show that in case $\mathcal{I}=\mathbb{N}$, (1.2.1) coincides with the usual definition.

EXERCISE 1.6.2. Show that the set function defined in (1.4.1) is a measure on $(\Omega, 2^\Omega)$.

EXERCISE 1.6.3. Let $\{a_\alpha\}_{\alpha\in\mathcal{I}} \subset (0,+\infty)$ be a family of *positive* real numbers, where $\mathcal{I}$ is an (index) set with uncountably many elements. Show that

$$\sum_{\alpha\in\mathcal{I}} a_\alpha = \infty.$$

CHAPTER 2

Measurability and σ-Algebras

2.1. Definition of a σ-algebra

Let Ω be a set. In probability theory, Ω is often referred to as the *sample space*.

DEFINITION 2.1.1. A collection of subsets of Ω, $\mathcal{F} \subset 2^{\Omega}$, is called a *$\sigma$-algebra* over Ω if

(i) $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$;
(ii) For every $A \in \mathcal{F}$, it holds that $\Omega \backslash A \in \mathcal{F}$;
(iii) For every sequence $A_1, A_2, \cdots \in \mathcal{F}$, it holds that

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

A pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a σ-algebra $\mathcal{F}$ is called a *measurable space*.

REMARK 2.1.2. From the second and third conditions, it follows that

$$\bigcap_{i=1}^{\infty} A_i = \Omega \backslash \left(\bigcup_{i=1}^{\infty} (\Omega \backslash A_i) \right) \in \mathcal{F} \qquad \text{for every sequence } A_1, A_2, \cdots \in \mathcal{F}.$$

In analysis, the sets in the σ-algebra $\mathcal{F}$ are called *measurable sets*. In probability theory, the sets in the σ-algebra $\mathcal{F}$ are usually called *events*.

EXAMPLE 2.1.3. The trivial σ-algebra $\mathcal{F} = \{\emptyset, \Omega\}$ consists of only two subsets of Ω, namely the empty set $\emptyset$ and Ω itself. $\diamondsuit$

EXAMPLE 2.1.4. Given a subset $A \subset \Omega$, the collection

$$\mathcal{F} = \{\emptyset, A, \Omega \backslash A, \Omega\} \qquad \text{is a } \sigma\text{-algebra.}$$

EXAMPLE 2.1.5. The power set 2^{Ω}, the set of all subsets of Ω, is a σ-algebra. $\diamondsuit$

PROPOSITION 2.1.6. *Let $(\Omega, \mathcal{F})$ be a measurable space and let $A \in \mathcal{F}$ be non-empty. The collection of subsets of A defined by*

$$\mathcal{F}_A := \{A \cap B : \ B \in \mathcal{F}\}$$

is a σ-algebra over A called the restriction of $\mathcal{F}$ to A.

PROOF. We show the three properties required for $\mathcal{F}_A$ to be a σ-algebra over A.

(i) Since $\emptyset, \Omega \in \mathcal{F}$, then also $A \cap \emptyset = \emptyset$, $A \cap \Omega = A \in \mathcal{F}_A$.
(ii) Let $E \in \mathcal{F}_A$. Then there is some $B \in \mathcal{F}$ for which $E = A \cap B$. Since $\Omega \backslash B \in \mathcal{F}$, we have that

$$A \backslash E = A \backslash (A \cap B) = A \cap (\Omega \backslash B) \in \mathcal{F}_A.$$

(iii) Let $E_1, E_2, \ldots \in \mathcal{F}_A$. We find $B_1, B_2, \ldots \in \mathcal{F}$ such that $E_i = A \cap B_i$, $i \geq 1$. Since $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$, we have that

$$\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} (A \cap B_i) = A \cap \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}_A.$$

With these three properties, we conclude that $\mathcal{F}_A$ is a σ-algebra over A. □

2.2. Generation of σ-algebras

Suppose we start off with a collection $\mathcal{A}$ of subsets of Ω, which is not necessarily a σ-algebra. We want to let the set $\mathcal{A}$ *generate* a σ-algebra, by adding sufficiently many other subsets of Ω to the collection until we have a σ-algebra. One may be tempted to do this sequentially: first adding all the complements, then adding all countable intersections and unions, and iterating this procedure. However, it turns out that one cannot generate a σ-algebra this way[1]. Instead, we need to follow a more abstract, but in a way easier, approach.

We say that a σ-algebra $\mathcal{F}$ is smaller than a σ-algebra $\mathcal{G}$ if every set in the σ-algebra $\mathcal{F}$ is also contained in $\mathcal{G}$, that is $\mathcal{F} \subset \mathcal{G}$.

PROPOSITION 2.2.1. *Given a collection $\mathcal{C}$ of subsets of Ω, and the collection $\Sigma_{\mathcal{C}}$ of all σ-algebras containing $\mathcal{C}$, the collection defined by*

$$\sigma(\mathcal{C}) := \bigcap_{\mathcal{F} \in \Sigma_{\mathcal{C}}} \mathcal{F} \tag{2.2.1}$$

is the smallest σ-algebra containing $\mathcal{C}$. It is called the σ-algebra generated by $\mathcal{C}$.

PROOF. We need to show that with this definition, $\sigma(\mathcal{C})$ is indeed a σ-algebra. This is completely straightforward but might take some getting used to.

(i) The empty set and Ω are by definition members of *every* σ-algebra $\mathcal{F} \in \Sigma_{\mathcal{C}}$, so also $\emptyset \in \sigma(\mathcal{C})$ and $\Omega \in \sigma(\mathcal{C})$.

(ii) Let $E \in \sigma(\mathcal{C})$. By definition of $\sigma(\mathcal{C})$, we know that $E \in \mathcal{F}$ for all $\mathcal{F} \in \Sigma_{\mathcal{C}}$. Since $\mathcal{F}$ is a σ-algebra, we also know that $\Omega \backslash E \in \mathcal{F}$, i.e. we have shown that $\Omega \backslash E \in \mathcal{F}$ for every $\mathcal{F} \in \Sigma_{\mathcal{C}}$. Thus, $\Omega \backslash E \in \sigma(\mathcal{C})$.

(iii) let $(E_i)_{i \in \mathbb{N}} \subset \sigma(\mathcal{C})$ be a family of sets in $\sigma(\mathcal{C})$. Then by definition, $E_i \in \mathcal{F}$ for every $i \in \mathbb{N}$ and every $\mathcal{F} \in \Sigma_{\mathcal{C}}$. Since $\mathcal{F}$ is a σ-algebra,

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{F} \qquad \text{for every } \mathcal{F} \in \Sigma_{\mathcal{C}}.$$

It then follows that

$$\bigcup_{i=1}^{\infty} E_i \in \sigma(\mathcal{C}).$$

Together, we conclude that $\sigma(\mathcal{C})$ is a σ-algebra.

The fact that $\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$ follows by construction. □

[1]This is not an easy statement and goes beyond the content of this course. If you want to know more, the keyword is "descriptive set theory". If one would use transfinite induction rather than ordinary induction one can sometimes construct the generated σ-algebra along the proposed lines. In a way, the arguments that we present in this section have this transfinite induction implicitly and conveniently built in.

EXAMPLE 2.2.2. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. There is a standard way of constructing a σ-algebra on the (Cartesian) product of sets

$$\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) : \ \omega_1 \in \Omega_1, \ \omega_2 \in \Omega_2\} .$$

The σ-algebra is called the product σ-algebra, it is denoted by $\mathcal{F}_1 \otimes \mathcal{F}_2$ and it is the σ-algebra generated by all sets of the form $A \times B$ where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. $\diamondsuit$

2.3. The Borel σ-algebra

The Borel σ-algebra on Euclidean space $\mathbb{R}^d$ is very important: it is (one of) the standard σ-algebra(s) on $\mathbb{R}^d$, while $\mathbb{R}^d$ is probably the space that we encounter most often, either as sample space or as target space of functions.

DEFINITION 2.3.1. The Borel σ-algebra in $\mathbb{R}^d$, denoted by $\mathcal{B}_{\mathbb{R}^d}$, is the σ-algebra generated by all open sets in $\mathbb{R}^d$.

It is often extremely convenient to use the fact that the Borel σ-algebra is also generated by a much smaller collection of subsets of $\mathbb{R}^d$.

PROPOSITION 2.3.2. *The Borel σ-algebra in $\mathbb{R}^d$ is generated by the sets*

$$(-\infty, a_1] \times \cdots \times (-\infty, a_n] \qquad \text{with } a_i \in \mathbb{Q},\ i = 1, \ldots, d.$$

The Borel σ-algebra on $\mathbb{R}$ has an easy extension to the Borel σ-algebra on the extended real line $\overline{\mathbb{R}}$.

DEFINITION 2.3.3. The Borel σ-algebra on the extended real line, denoted by $\mathcal{B}_{\overline{\mathbb{R}}}$ is the σ-algebra generated by the sets $[-\infty, a]$, for $a \in \mathbb{Q}$. Equivalently, it consists of those subsets $A \subset \mathbb{R}$ such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$.

REMARK 2.3.4. The Borel σ-algebra exists more generally for any topological space.

2.4. The monotone class theorem

The purpose of this section is two-fold. Firstly, we introduce an extremely powerful theorem that allows us to generate σ-algebras. Even though the theorem is incredibly useful, its proof is not much more difficult than the arguments that we have seen before. So the second purpose of this section is to get some practice with this type of argumentation by proving the monotone class theorem along a series of exercises.

A σ-algebra is a collection of subsets of Ω that satisfies certain properties. The monotone class theorem revolves around two other types of collections, algebras, and monotone classes, that each have their own defining properties.

DEFINITION 2.4.1. A collection of subsets $\mathcal{A}$ of Ω is an *algebra* of sets if

(i) $\emptyset \in \mathcal{A}$.
(ii) $\Omega \backslash A \in \mathcal{A}$ for every $A \in \mathcal{A}$.
(iii) $A \cup B \in \mathcal{A}$ for every $A, B \in \mathcal{A}$.

DEFINITION 2.4.2. A collection of subsets $\mathcal{M}$ of Ω is called a *monotone class* if

(i) $\mathcal{M}$ is closed under countable unions of *increasing* sequences, i.e.,

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{M} \qquad \text{for every increasing sequence } A_1 \subset A_2 \subset \cdots \text{ in } \mathcal{M}.$$

(ii) $\mathcal{M}$ is closed under countable intersections of *decreasing* sequences, i.e.,

$$\bigcap_{i=1}^{\infty} B_i \in \mathcal{M} \qquad \text{for every decreasing sequence } B_1 \supset B_2 \supset \cdots \text{ in } \mathcal{M}.$$

Clearly, there are monotone classes that are not algebras, let alone σ-algebras, as seen in the following example.

EXAMPLE 2.4.3. Let $\Omega = \{a, b, c, d\}$ and consider the family $\mathcal{M} = \{\{a\}, \{a, b\}\}$. Then $\mathcal{M}$ is a monotone class, but for obvious reasons ($\Omega \notin \mathcal{M}$) not an algebra of sets. ♢

Analogous to how $\sigma(\mathcal{A})$ was defined, we have the following, whose proof follows the same strategy as the proof of Proposition 2.2.1.

PROPOSITION 2.4.4 (Minimal monotone class). *Given a collection $\mathcal{C}$ of subsets of Ω, and the collection $\Xi_{\mathcal{C}}$ of all monotone classes containing $\mathcal{C}$, the collection defined by*

$$\mathcal{M}(\mathcal{A}) := \bigcap_{\mathcal{F} \in \Xi_{\mathcal{C}}} \mathcal{F} \tag{2.4.1}$$

is the smallest monotone class containing $\mathcal{C}$. It is called the monotone class generated by $\mathcal{C}$.

THEOREM 2.4.5 (The monotone class theorem). *Let $\mathcal{A}$ be an algebra of sets over Ω. Then the monotone class $\mathcal{M}(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$, i.e.*

$$\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A}).$$

We will prove the theorem in several steps (cf. Exercises 2.5.2) in the homework exercises.

2.5. Exercises

EXERCISE 2.5.1. Prove Proposition 2.3.2.

EXERCISE 2.5.2. (i) Show that every σ-algebra is a monotone class.

(ii) Show that if $\mathcal{M}$ is a monotone class and also an algebra of sets over Ω, then $\mathcal{M}$ is a σ-algebra over Ω.

(iii) Conclude the proof of the monotone class theorem (cf. Theorem 2.4.5) by showing that $\mathcal{M}(\mathcal{A})$ is an algebra.

CHAPTER 3

Measures and their Properties

3.1. Definition of a measure

In the introduction, we have seen that it is often *not* possible to define a measure that assigns a nonnegative real number to *every* possible subset of Ω, if we also want it to satisfy some extra properties such as translation invariance. The proposed way out was to define the measure only for certain subsets of Ω. In this chapter we will see whether this approach indeed leads to useful measures.

To do so we will need to define what we mean for a set function on a collection of subsets of Ω to be σ-additive: Let $\mathcal{C} \subset 2^\Omega$ be a collection of subsets of Ω and $\mu : \mathcal{C} \to [0, +\infty]$ be a set function. Then μ is said to be *σ-additive* if for every family $\{A_i\}_{i\in\mathbb{N}}$ of mutually disjoint subsets $A_i \in \mathcal{C}$ of Ω with $\bigcup_{i=1}^\infty A_i \in \mathcal{C}$, we have that

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

Definition 3.1.1. Let $(\Omega, \mathcal{F})$ be a measurable space. A measure μ on $(\Omega, \mathcal{F})$ is a set function $\mu : \mathcal{F} \to [0, +\infty]$ that assigns to each measurable set $A \in \mathcal{F}$ a nonnegative extended real number $\mu(A) \in [0, +\infty]$, and satisfies the following properties:

(i) $\mu(\emptyset) = 0$.
(ii) μ is σ-additive.

We call a triple $(\Omega, \mathcal{F}, \mu)$, consisting of a measurable space $(\Omega, \mathcal{F})$ and a measure μ on $(\Omega, \mathcal{F})$ a *measure space*. If, in addition, $\mu(\Omega) = 1$, then we call μ a *probability measure* and the triple $(\Omega, \mathcal{F}, \mu)$ a *probability space*. We often use the notion $\mathbb{P}$ for a probability measure.

Additionally, we say that a measure μ is *finite* if $\mu(\Omega) < +\infty$ and that it is *σ-finite* if there exists a sequence of sets $B_i \in \mathcal{F}$, such that $\mu(B_i) < \infty$ for every $i \in \mathbb{N}$ and

$$\Omega = \bigcup_{i=1}^\infty B_i.$$

3.2. First properties of measures

Let us derive some simple properties of measures first.

Proposition 3.2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $A, B \in \mathcal{F}$ with $A \subset B$.*

(i) *(finite additivity) If $n \in \mathbb{N}$ and $A_i \in \mathcal{F}$, $i = 1, \ldots, n$ are mutually disjoint, then*

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i).$$

(ii) *(monotonicity) Then*

$$\mu(A) \leq \mu(B)$$

(iii) *(exclusion) If in addition $\mu(B) < +\infty$, then*

$$\mu(B \backslash A) = \mu(B) - \mu(A).$$

(iv) *(σ-subadditivity) If $(A_i) \subset \mathcal{F}$ is a sequence of measurable sets such that*

$$A \subset \bigcup_{i=1}^{\infty} A_i, \qquad \textit{then} \qquad \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

PROOF. (i) The finite additivity is a direct consequence of the σ-additivity of the measure: Given mutually disjoint sets $A_i \in \mathcal{F}$ for $i = 1, \ldots, n$, we can define $A_i := \emptyset$ for $i > n$ and apply the σ-additivity.

(ii) If $A \subset B$, then B is the disjoint union of A and $B \backslash A$. Since the measure μ is countably additive by definition, it is also finitely additive. Hence

$$\mu(A) + \mu(B \backslash A) = \mu(B)$$

so that $\mu(A) \leq \mu(B)$. Hence, we have shown the monotonicity property.

(iii) To show the exclusion property, note that if in addition $\mu(B) < +\infty$, we may subtract $\mu(A)$ from both sides and find

$$\mu(B \backslash A) = \mu(B) - \mu(A).$$

(iv) Finally, to show the σ-subadditivity, we first define the sets

$$B_m := A_m \backslash \bigcup_{i=1}^{m-1} A_i, \qquad m \geq 1.$$

Then the sets B_m are mutually disjoint by construction. Moreover,

$$A = \bigcup_{m=1}^{\infty} (A \cap B_m).$$

By the σ-additivity and monotonicity, it now follows that

$$\mu(A) = \sum_{m=1}^{\infty} \mu(A \cap B_m) \leq \sum_{m=1}^{\infty} \mu(B_m) \leq \sum_{m=1}^{\infty} \mu(A_m),$$

which concludes the proof. □

3.3. Important continuity results

The following continuity results are extremely important: they are used again and again in this course.

PROPOSITION 3.3.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $(A_i)_{i \in \mathbb{N}}$ be an increasing sequence of $\mathcal{F}$-measurable sets, that is $A_i \subset A_{i+1}$ for all $i \in \mathbb{N}$. Then,*

$$\lim_{i \to \infty} \mu(A_i) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right).$$

This property is know as continuity from below of a measure.

PROOF. Define the sets $E_1 = A_1$, $E_i := A_{i+1} \backslash A_i$, $i \geq 2$. Then $(E_i)_{i \in \mathbb{N}}$ is a sequence of mutually disjoint $\mathcal{F}$-measurable sets with

$$A := \bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} A_i.$$

Therefore, by the σ-additivity of the measure μ,

$$\mu(A) = \sum_{i=1}^{\infty} \mu(E_i) = \lim_{k \to \infty} \sum_{i=1}^{k} \mu(E_i) = \lim_{k \to \infty} \mu(A_k). \qquad \square$$

The next proposition is very similar to the previous one. However, there is an important assumption that shouldn't be overlooked, namely that the first set in the sequence has a finite measure. Without this assumption, the conclusion would in general not hold.

PROPOSITION 3.3.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $(B_i)_{i \in \mathbb{N}}$ be a* decreasing *sequence of $\mathcal{F}$-measurable sets, that is $B_i \supset B_{i+1}$ for all $i \in \mathbb{N}$. Let moreover $\mu(B_1) < +\infty$. Then,*

$$\lim_{i \to \infty} \mu(B_i) = \mu\left(\bigcap_{i=1}^{\infty} B_i\right).$$

This property is know as continuity from above of a measure.

3.4. Uniqueness of measures

If two σ-finite measures μ_1 and μ_2 on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ coincide on a family of sets that generate $\mathcal{B}_{\mathbb{R}}$, are they in fact the same measure? That is, do they then assign the same probability to *every* set in the Borel σ-algebra of $\mathbb{R}$? The answer is *yes*, and the proof is a true display of the power of the monotone class theorem. More generally, we have the following result.

THEOREM 3.4.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ_1, μ_2 be finite measures on $(\Omega, \mathcal{F})$. Let $\mathcal{A} \subset \mathcal{F}$ be an algebra of sets over Ω and suppose that $\mu_1 \equiv \mu_2$ on $\mathcal{A}$, i.e., $\mu_1(A) = \mu_2(A)$ for every $A \in \mathcal{A}$. Then $\mu_1 \equiv \mu_2$ on $\sigma(\mathcal{A}) \subset \mathcal{F}$.*

PROOF. Define the collection $\mathcal{M}$ of $\mathcal{F}$-measurable sets

$$\mathcal{M} := \Big\{ E \in \mathcal{F} : \ \mu_1(E) = \mu_2(E) \Big\}.$$

Clearly, $\mathcal{A} \subset \mathcal{M}$. We now claim that $\mathcal{M}$ is a monotone class. From that, it would follow immediately by the monotone class theorem that $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A}) \subset \mathcal{M}$. This simply means that $\mu_1(E) = \mu_2(E)$ every $\mathcal{F}$-measurable set $E \in \sigma(\mathcal{A})$. So all that is left to do is to check that $\mathcal{M}$ is a monotone class.

Let $A_1 \subset A_2 \subset \cdots$ be an increasing sequence in $\mathcal{M}$ with $\mu_1(A_i) = \mu_2(A_i)$ for all $i \in \mathbb{N}$. Since both μ_1 and μ_2 are measures, the continuity from below of measures imply

$$\mu_1\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mu_1(A_i) = \lim_{i \to \infty} \mu_2(A_i) = \mu_2\left(\bigcup_{i=1}^{\infty} A_i\right),$$

thus implying that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$.

In a similar fashion, we can show using the continuity from above of measures that $\mathcal{M}$ is closed under countable intersections of decreasing sequences. Note that, here, we use the fact that μ_1 and μ_2 are finite to guarantee that $\mu_1(B_1) = \mu_2(B_1) < +\infty$ for every decreasing sequence of sets $B_1 \supset B_2 \supset \cdots$ in $\mathcal{M}$. $\square$

3.5. Null sets

Definition 3.5.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We say that a subset $N \subset \Omega$ is a *null set* if there exists a measurable set $B \in \mathcal{F}$ with $N \subset B$ and $\mu(B) = 0$. The set N itself does *not* need to be measurable.

We say that a property holds *almost everywhere* if it holds everywhere except in a null set. If we are considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a property holds almost everywhere, then and only then we say that the property holds *almost surely*. We often abbreviate almost everywhere and almost surely as *a.e.* and *a.s.* respectively. To highlight the dependence on the measure μ, we sometimes write μ-a.e. or μ-a.s.

A measure space $(\Omega, \mathcal{F}, \mu)$ is said to be *complete* if every null set $N \in \mathcal{F}$.

Example 3.5.2. Let $A \subset \Omega$ with $\Omega \backslash A \neq \emptyset$. Consider the σ-algebra $\mathcal{F} = \{\emptyset, A, \Omega\backslash A, \Omega\}$. Now let $\mu = 1$ on $\Omega\backslash A$ and $\mu = 0$ on A. Then every subset $N \subset A$, $N \neq A$ is a null set that is not an element of $\mathcal{F}$. Hence, $(\Omega, \mathcal{F}, \mu)$ is *not* complete. $\diamondsuit$

Proposition 3.5.3. *Countable unions of null sets are null sets.*

The following proposition makes it a bit easier to show that some sets are null sets.

Proposition 3.5.4. *A set N is a null set if and only if for every $\epsilon > 0$ there exists a set $B \in \mathcal{F}$ such that $N \subset B$ and $\mu(B) < \epsilon$.*

Proof. The 'only if' part immediately follows from the definition of a null set. To show the 'if' part, let N be such that for every $\epsilon > 0$ there exists a set $B \in \mathcal{F}$ such that $N \subset B$ and $\mu(B) < \epsilon$. We define inductively a decreasing sequence of sets B_n inductively, such that $N \subset B_n$ and $\mu(B_n) < 1/n$ for every $n \in \mathbb{N}$. First, we set $\epsilon = 1$, and use that there exists a set $B_1 \in \mathcal{F}$ with $N \subset B_1$ and $\mu(B_1) < 1$. Now suppose $B_n \in \mathcal{F}$ has been defined. Then there exists a set $B_{n+1} \subset B_n$ such that $N \subset B_{n+1}$ and $\mu(B_{n+1}) < 1/(n+1)$.

Now define

$$B \coloneqq \bigcap_{n=1}^{\infty} B_n.$$

Then B is measurable, that is $B \in \mathcal{F}$. Moreover, $N \subset B$ and by the continuity result in Proposition 3.3.2 we have

$$\mu(B) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \frac{1}{n} = 0.$$

Hence, N is a null set. $\square$

Definition 3.5.5. Let $\mathcal{C}$ and $\mathcal{A}$ be two collections of subsets of Ω such that $\mathcal{C} \subset \mathcal{A}$. A measure $\nu : \mathcal{A} \to [0, +\infty]$ is said to be an *extension* of a set function $\mu : \mathcal{C} \to [0, +\infty]$ if $\mu(A) = \nu(A)$ for every $A \in \mathcal{C}$.

Theorem 3.5.6 (The completion of a measure space). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then the collection $\mathcal{G} \subset 2^\Omega$ defined by*

$$\mathcal{G} \coloneqq \Big\{A \cup N : \ A \in \mathcal{F},\ N \subset \Omega \textit{ is a } \mu\textit{-null set}\Big\}$$

is a σ-algebra that contains $\mathcal{F}$, and the set function $\bar{\mu} : \mathcal{G} \to [0, \infty]$ defined by

$$\bar{\mu}(A \cup N) := \mu(A), \qquad A \in \mathcal{F},$$

is well-defined and extends μ. Moreover, $(\Omega, \mathcal{G}, \bar{\mu})$ is a complete measure space.

3.6. Exercises

EXERCISE 3.6.1. (i) Prove Proposition 3.3.2.

(ii) Give an example of a measure space $(\Omega, \mathcal{F}, \mu)$ and a decreasing sequence $(B_i)_{i \in \mathbb{N}}$ of $\mathcal{F}$-measurable sets such that

$$\lim_{i \to \infty} \mu(B_i) \neq \mu\left(\bigcap_{i=1}^{\infty} B_i\right).$$

EXERCISE 3.6.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A_1, \ldots, A_n \in \mathcal{F}$. Show that

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - n + \sum_{i=1}^{n} \mathbb{P}(A_i).$$

EXERCISE 3.6.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A_1, A_2 \ldots \in \mathcal{F}$ with $\mathbb{P}(A_i) = 1$ for all $i \in \mathbb{N}$. Show that $\mathbb{P}\left(\bigcap_{i \in \mathbb{N}} A_i\right) = 1$.

EXERCISE 3.6.4. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Denote by $\nu := \mu|_{\mathcal{G}}$ the *restriction* of μ to $\mathcal{G}$.

(i) Show that ν is again a measure.
(ii) Assume that μ is a finite measure. Is ν still a finite measure?
(iii) Does ν inherit σ-finiteness from μ?

EXERCISE 3.6.5. Prove Theorem 3.5.6.

CHAPTER 4

The Lebesgue Measure

The goal of this chapter is to finally construct the infamous Lebesgue measure on $\mathbb{R}^d$, i.e. we would like to have a measure that satisfies additionally the property of translational invariance. To do so, we start by introducing the semi-algebra of sets and thereafter defining a candidate measure on a semi-algebra of $\mathbb{R}$, which we then *extend* to a measure on $\mathcal{B}_{\mathbb{R}^d}$.

4.1. Semi-algebra of sets

Definition 4.1.1 (Semi-algebra of sets). A collection $\mathcal{S}$ of subsets of Ω is called a *semi-algebra* over Ω if

(i) $\emptyset, \Omega \in \mathcal{S}$.
(ii) $A \cap B \in \mathcal{S}$ for all $A, B \in \mathcal{S}$.
(iii) For all $A \in \mathcal{S}$ with $\Omega \backslash A \notin \mathcal{S}$, there exists a finite family of mutually disjoint sets $\{A_i\}_{i=1,\ldots,n} \subset \mathcal{S}$ such that $\Omega \backslash A = \bigcup_{i=1}^n A_i$.

Example 4.1.2. Let Ω be a finite set. Then the collection $\mathcal{S} = \{\{\omega\} \in \Omega\} \cup \{\emptyset, \Omega\}$ of singletons is a semi-algebra. $\diamondsuit$

Exercise 4.1.3. Show that the collection $\mathcal{I}$ of right-closed real intervals, i.e. intervals of the form $(a, b]$, $a, b \in \mathbb{R}$, taking the form

$$\mathcal{I} = \big\{(a, b] : a \leq b,\ a, b \in \mathbb{R}\big\} \cup \big\{(a, \infty) : a \in \mathbb{R}\big\} \cup \big\{\mathbb{R}\big\}.$$

is a semi-algebra.

As in the case for σ-algebras and monotone classes, one can use a semi-algebra $\mathcal{S}$ over a given set Ω to generate an algebra over Ω. In fact, the generated algebra can be described explicitly in terms of sets from the semi-algebra $\mathcal{S}$.

Proposition 4.1.4. *If $\mathcal{S}$ is a semi-algebra over Ω, then*

$$\mathcal{A}(\mathcal{S}) := \left\{A \subset \Omega : \exists\, n \in \mathbb{N},\ \ A = \bigcup_{i=1}^n A_i,\ \ A_i \in \mathcal{S}\ \ \textit{mutually disjoint}\right\},$$

is the smallest algebra containing $\mathcal{S}$. $\mathcal{A}(\mathcal{S})$ is called the algebra generated by $\mathcal{S}$.

4.2. Premeasures and the outer measure

Given a set function μ defined on a semi-algebra $\mathcal{S}$, we can naturally extend μ to a set function on $\mathcal{A}(\mathcal{S})$ due to Proposition 4.1.4. Indeed, setting

$$\mu(A) := \sum_{i=1}^n \mu(A_i), \qquad A \in \mathcal{A}(\mathcal{S}),$$

we obtain a consistent definition of a set function on $\mathcal{A}(\mathcal{S})$, which is in fact unique.

Definition 4.2.1. Let $\mathcal{A}$ be an algebra over Ω. A *premeasure* $\mu_\circ$ on $\mathcal{A}$ is a set function $\mu_\circ : \mathcal{A} \to [0,\infty]$ such that $\mu_\circ(\emptyset) = 0$ and $\mu_\circ$ is σ-additive.

The *outer measure* μ^* induced by a premeasure $\mu_\circ$ is the set function $\mu^* : 2^\Omega \to [0,+\infty]$ defined for every subset $A \subset \Omega$ by

$$\mu^*(A) \coloneqq \inf\left\{ \sum_{i=1}^{\infty} \mu_\circ(A_i) : A \subset \bigcup_{i=1}^{\infty} A_i,\ A_i \in \mathcal{A} \right\}.$$

Proposition 4.2.2. *The outer measure μ^* induced by a pre-measure $\mu_\circ$ on an algebra $\mathcal{A}$ over Ω satisfies the following properties:*

(i) *μ^* extends $\mu_\circ$, i.e. $\mu^*(A) = \mu_\circ(A)$ for all $A \in \mathcal{A}$.*
(ii) *$\mu^*(\emptyset) = 0$ and $\mu^*(A) \geq 0$ for all $A \subset \Omega$.*
(iii) *μ^* is monotone: If $A, B \subset \Omega$ with $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.*
(iv) *μ^* is σ-subadditive: If $\{A_i\}_{i\in\mathbb{N}}$ is a family of subsets of Ω such that*

$$A \subset \bigcup_{i=1}^{\infty} A_i, \qquad \text{then} \qquad \mu^*(A) \leq \sum_{i=1}^{\infty} \mu^*(A_i).$$

Proof. (i) Since any $A \in \mathcal{A}$ covers itself, we have that $\mu^*(A) \leq \mu_\circ(A)$.

We now prove the reverse inequality: If $\mu^*(A) = +\infty$, then the equality holds true. Suppose that $\mu^*(A) < +\infty$. For any $\varepsilon > 0$ there exists a family $\{A_i\}_{i\in\mathbb{N}}$ with $A_i \in \mathcal{A}$ for all $i \in \mathbb{N}$ such that

$$\sum_{i=1}^{\infty} \mu_\circ(A_i) < \mu^*(A) + \varepsilon \qquad \text{and} \qquad A \subset \bigcup_{i=1}^{\infty} A_i.$$

Since $\mu_\circ$ is σ-subadditive and monotone, we have that

$$\sum_{i=1}^{\infty} \mu_\circ(A_i) \geq \mu_\circ\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \mu_\circ(A),$$

and therefore $\mu_\circ(A) < \mu^*(A) + \varepsilon$. Letting $\varepsilon \to 0$, we conclude that $\mu_\circ(A) \leq \mu^*(A)$. Together, this gives $\mu^*(A) = \mu_\circ(A)$ for all $A \in \mathcal{A}$.

(ii) Obvious.

(iii) Let $A, B \subset \Omega$ with $A \subset B$. Further, let $\{B_i\}_{i\in\mathbb{N}}$ be an arbitrary cover for B with $B_i \in \mathcal{A}$ for all $i \in \mathbb{N}$. Since $A \subset B$, $\{B_i\}_{i\in\mathbb{N}}$ is also a cover for A. Hence, by the definition of μ^*, we have that

$$\mu^*(A) \leq \sum_{i=1}^{\infty} \mu_\circ(B_i).$$

Since the left-hand side is independent of the cover for B, we can minimize over all covers for B to obtain $\mu^*(A) \leq \mu^*(B)$, as desired.

(iv) Due to (iii) it suffices to consider the case $A = \bigcup_{i=1}^{\infty} A_i$. If $\mu^*(A_i) = +\infty$ for some $i \in \mathbb{N}$, then there is nothing to prove. Suppose that $\mu^*(A_i) < +\infty$ for all $i \in \mathbb{N}$. For every $\varepsilon > 0$, we find a cover $\{A^i_j\}_{j\in\mathbb{N}}$ for A_i with $A^i_j \in \mathcal{A}$ such that

$$\sum_{j=1}^{\infty} \mu_\circ(A^i_j) < \mu^*(A_i) + \varepsilon\, 2^{-i}.$$

Clearly, we have that $A \subset \bigcup_{i=1}^\infty \bigcup_{j=1}^\infty A_j^i$, i.e. the family $\{A_j^i\}_{i,j\in\mathbb{N}}$ is a cover for A. Consequently, we have that

$$\mu^*(A) \leq \sum_{i=1}^\infty \sum_{j=1}^\infty \mu_\circ(A_j^i) < \sum_{i=1}^\infty \mu^*(A_i) + \varepsilon.$$

Once again passing $\varepsilon \to 0$, we obtain the σ-subadditivity of μ^*. □

EXERCISE 4.2.3. Let $\mathcal{I}$ be the semi-algebra of right-closed intervals (cf. Example 4.1.3).

(i) Show that the set function $\lambda : \mathcal{I} \to [0, +\infty]$ defined by

$$\lambda((a,b]) := \begin{cases} |b-a| & \text{if } a \leq b \in \mathbb{R}, \\ +\infty & \text{otherwise,} \end{cases}$$

satisfies the properties of a measure ($\lambda(\emptyset) = 0$ and σ-additivity) and is additionally *translational invariant*, i.e. for any $x \in \mathbb{R}$, $\lambda(x + A) = \lambda(A)$ for every $A \in \mathcal{I}$.

(ii) Furthermore, show that λ extends to a pre-measure $\lambda_\circ$ on $\mathcal{A}(\mathcal{I})$ with the same properties by setting

$$\lambda_\circ(A) := \sum_{i=1}^n \lambda(A_i), \qquad A \in \mathcal{A}(\mathcal{I}).$$

(iii) Finally, show that the outer measure λ^* induced by $\lambda_\circ$ is translation invariant.

4.3. Carathéodory's extension theorem

Our next theorem provides a general strategy in constructing a complete measure space from a given pre-measure via its outer measure.

THEOREM 4.3.1 (Carathéodory's extension theorem). *Let $\mu_\circ$ be a pre-measure on an algebra $\mathcal{A}$ over Ω and μ^* be the outer measure induced by $\mu_\circ$. Then the collection of sets $\mathcal{L}_{\mu^*} \subset 2^\Omega$ defined by*

$$\mathcal{L}_{\mu^*} := \left\{ E \subset \Omega : \mu^*(A) \geq \mu^*(E \cap A) + \mu^*(A \backslash E) \;\; \forall\, A \in \mathcal{A} \right\}$$

is a σ-algebra that contains all the null sets of μ^*.

Moreover, the outer measure μ^* is σ-additive on $\mathcal{L}_{\mu^*}$ and its restriction $\bar\mu := \mu^*|\mathcal{L}_{\mu^*}$ on $\mathcal{L}$ is the unique extension of $\mu_\circ$ from $\mathcal{A}$ to $\mathcal{L}_{\mu^*}$. In particular, $(\Omega, \mathcal{L}_{\mu^*}, \bar\mu)$ is a complete measure space and we call $\bar\mu$ the Carathéodory extension of $\mu_\circ$.

Recall the measure λ defined in Example 4.2.3 on the semi-algebra $\mathcal{I}$ of right-closed intervals, which extends to a pre-measure $\lambda_\circ$ on $\mathcal{A}(\mathcal{I})$. Then,

DEFINITION 4.3.2. We define the Lebesgue measure $\bar\lambda : \mathcal{L}_{\lambda^*} \to [0, +\infty]$ as the Carathéodory extension of the pre-measure $\lambda_\circ$ on $\mathcal{A}(\mathcal{I})$.

4.4. Universal construction of measures on the real line

Given a probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, one obtains the cumulative distribution function $F(x) := \mathbb{P}((-\infty, x])$, $x \in \mathbb{R}$. The following exercise shows that F is monotone, non-decreasing, and right-continuous with left limits.

EXERCISE 4.4.1. Show that F is monotone, non-decreasing, and right-continuous with left limits (*cádlág*), i.e.

$$\lim_{y \downarrow x} F(y) = F(x) \qquad \text{and} \qquad \lim_{y \uparrow x} F(y) \ \text{ exists.}$$

It turns out that the converse statement holds true. Let $F : \mathbb{R} \to \mathbb{R}$ be a monotone, non-decreasing, cádlág function. Then F induces a set function μ_F on the semi-algebra $\mathcal{I}$ of right-closed intervals $(a, b]$ by

$$\mu_F((a, b]) := F(b) - F(a),$$

which extends to a pre-measure on $\mathcal{A}(\mathcal{I})$, which we also denote by μ_F.

THEOREM 4.4.2. *Let $F : \mathbb{R} \to \mathbb{R}$ be a monotone, non-decreasing, cádlág function. Then there exists a unique measure $\bar{\mu}_F$ on $(\mathbb{R}, \mathcal{L}_{\mu_F^*})$ with $\bar{\mu}_F((a, b]) = F(b) - F(a)$.*

Moreover, $\bar{\mu}_F$ is a probability measure if

$$\lim_{x \to +\infty} F(x) - \lim_{x \to -\infty} F(x) = 1.$$

In particular, we obtain a full characterization of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

CHAPTER 5

Measurable functions

One of the problems with Riemann integration is that a sequence of continuous functions may have a pointwise limit that is not Riemann integrable. But what kind of functions are a pointwise limit of continuous functions? In this chapter we will introduce the class of *measurable functions*. We will see that every limit of a continuous function is measurable.

Just like we had to accept that there are sets that we cannot assign a measure to, there are functions that we just cannot integrate. But as we will see in later chapters, bounded measurable functions can always be integrated against a finite measure.

Actually, you probably have seen measurable functions before, but you know them by the name of *random variables*. The integral of a random variable against a probability measure is just its expectation value.

5.1. Measurable functions

Given a function $f : \Omega \to E$, recall that the preimage of a subset $A \subset E$ under f is defined as

$$f^{-1}(A) := \{\omega \in \Omega : \ f(\omega) \in A\}.$$

We now give the definition of a *measurable function*, which are called *random variables* in probability theory.

Definition 5.1.1. Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces. We say that a function $f : \Omega \to E$ is $(\mathcal{F}, \mathcal{G})$*-measurable* if $f^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{G}$.

Note that the whether a function $f : \Omega \to E$ is $(\mathcal{F}, \mathcal{G})$-measurable depends on both the σ-algebras $\mathcal{F}$ and $\mathcal{G}$. As an alternative notation, we will sometimes write that a function $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ is measurable, and other times we will leave out (one of) the σ-algebras when they are understood. In particular, when we say that a function $f : \Omega \to \mathbb{R}^d$ is $\mathcal{F}$-measurable, we simply mean that f is $(\mathcal{F}, \mathcal{B}_{\mathbb{R}^d})$-measurable.

Proposition 5.1.2. *Let $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ be a measurable function, and let $A \in \mathcal{F}$ be non-empty. Let $f|_A : \Omega \to E$ be the restriction of f to A. Then $f|_A$ is $(\mathcal{F}_A, \mathcal{G})$-measurable.*

If a function $g_A : A \to E$ is $(\mathcal{F}_A, \mathcal{G})$-measurable, and $p \in E$, then the extension

$$g(\omega) := \begin{cases} g_A(\omega) & \omega \in A \\ p & \omega \notin A, \end{cases} \qquad \text{is } (\mathcal{F}, \mathcal{G})\text{-measurable.}$$

Proof. Let $B \in \mathcal{G}$. Then $f|_A^{-1}(B) = f^{-1}(B) \cap A \in \mathcal{F}_A$.

Because g_A is $(\mathcal{F}_A, \mathcal{G})$-measurable, there exists a set $D \in \mathcal{F}$ such that

$$g^{-1}(B) \cap A = g_A^{-1}(B) = D \cap A.$$

Therefore, if $p \notin B$,

$$g^{-1}(B) = g^{-1}(B) \cap A = D \cap A \in \mathcal{F}.$$

If $p \in B$, then

$$g^{-1}(B) = (g^{-1}(B) \cap A) \cup (g^{-1}(B) \backslash A) = (D \cap A) \cup (\Omega \backslash A) \in \mathcal{F}. \qquad \square$$

5.2. Checking for measurability

The following lemma often greatly reduces the effort in showing that a function is measurable: instead of checking that the preimages of all measurable sets are measurable, it suffices to check the measurability of the preimages of sets that generate the σ-algebra in the target.

LEMMA 5.2.1. *Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces, such that $\mathcal{G} = \sigma(\mathcal{A})$ for some collection $\mathcal{A}$ of subsets of E. Let $f : \Omega \to E$ be such that for every $A \in \mathcal{A}$, the pre-image of A is $\mathcal{F}$-measurable, that is*

$$f^{-1}(A) \in \mathcal{F}.$$

Then, f is a measurable function from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.

PROOF. Again, the strategy of this proof may be surprising. But it is also highly convenient. We are going to show that the collection $\mathcal{H}$ of all sets $B \subset E$ such that $f^{-1}(B) \in \mathcal{F}$ is a σ-algebra. Because this σ-algebra clearly contains $\mathcal{A}$, and $\mathcal{G}$ is the smallest σ-algebra containing $\mathcal{A}$, it follows that $\mathcal{G} \subset \mathcal{H}$. Translating back, this means that for every set $D \in \mathcal{G}$, we have $D \in \mathcal{H}$, and thus that

$$f^{-1}(D) \in \mathcal{F}.$$

Consequently, we would have shown that f is measurable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.

So let us prove that $\mathcal{H}$ is a σ-algebra. First of all, $f^{-1}(\emptyset) = \emptyset$, so $\emptyset \in \mathcal{H}$ and $f^{-1}(E) = \Omega \in \mathcal{F}$ so that also $E \in \mathcal{H}$.

If $B \in \mathcal{H}$, then $f^{-1}(B) \in \mathcal{F}$. Moreover,

$$f^{-1}(E \backslash B) = \Omega \backslash f^{-1}(B) \in \mathcal{F},$$

because $\mathcal{F}$ is a σ-algebra.

Finally, if (B_i) is a sequence of sets in $\mathcal{H}$, then

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i) \in \mathcal{F}. \qquad \square$$

As a first application of the above lemma, we will show that every continuous function is measurable. We first recall the definition of a continuous function.

DEFINITION 5.2.2. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called continuous if for every $x \in \mathbb{R}^n$ and every $\epsilon > 0$ there exists an $r > 0$ such that

$$|f(x) - f(y)| < \epsilon \qquad \text{for every } y \in B(x, r).$$

The following proposition gives an alternative characterization of continuity[1].

[1] In the mathematical field of topology, the alternative characterization that we give here is actually taken as the definition of continuity.

PROPOSITION 5.2.3. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if for every open set $O \subset \mathbb{R}^m$ the preimage $f^{-1}(O)$ is open.*

PROOF. $\Rightarrow$: We assume that the function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous. Take an arbitrary open set $O \subset \mathbb{R}^m$. We need to show that $f^{-1}(O)$ is open. So let $x \in f^{-1}(O)$ be arbitrary. We need to show that we can find a $r > 0$ such that $B(x, r) \subset f^{-1}(O)$. Since O is open, there exists an $\epsilon > 0$ such that $B(f(x), \epsilon) \subset O$. By the continuity of f, there exists an $r > 0$ such that for every $y \in B(x, r)$,

$$|f(y) - f(x)| < \epsilon,$$

or in other words, $f(y) \in B(f(x), \epsilon)$. But $B(f(x), \epsilon) \subset O$ so $f(y) \in O$ for every $y \in B(x, r)$. That is, $B(x, r) \subset f^{-1}(O)$.

$\Leftarrow$: Let $x \in \mathbb{R}^n$ and $\epsilon > 0$. Clearly, the ball $B(f(x), \epsilon) \subset \mathbb{R}^m$ is open. By assumption, the preimage $f^{-1}(B(X(x), \epsilon))$ is open as well. Since $x \in f^{-1}(B(f(x), \epsilon))$, there exists an $r > 0$ such that

$$B(x, r) \subset f^{-1}(B(f(x), \epsilon)).$$

Hence, for all $y \in B(x, r)$, we have $|f(x) - f(y)| < \epsilon$. Since $x \in \mathbb{R}^n$ and $\epsilon > 0$ were arbitrary, it follows that the function f is indeed continuous. $\square$

The above fact makes it easy to see why continuous functions are measurable.

PROPOSITION 5.2.4. *Every continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ is $(\mathcal{B}_{\mathbb{R}^n}, \mathcal{B}_{\mathbb{R}^m})$-measurable.*

PROOF. By Lemma 5.2.1 and the definition of the Borel σ-algebra as the σ-algebra generated by all open sets, it suffices to check that for every $O \subset \mathbb{R}^m$, the set $f^{-1}(O)$ is $\mathcal{B}_{\mathbb{R}^n}$-measurable. By Proposition 5.2.3, $f^{-1}(O)$ is open, so certainly, $f^{-1}(O) \in \mathcal{B}_{\mathbb{R}^n}$. $\square$

5.3. The composition of measurable functions

The following lemma will say that the composition of two measurable functions is measurable. This seems simple enough, but we need to issue a big **warning** at this stage. Whether a function is measurable or not depends on the σ-algebras on both domain and target.

LEMMA 5.3.1. *Let $(\Omega_i, \mathcal{F}_i)$ for $i = 1, 2, 3$ be three measurable spaces. Let $f : \Omega_1 \to \Omega_2$ be a measurable function from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ and $g : \Omega_2 \to \Omega_3$ be a measurable function from $(\Omega_2, \mathcal{F}_2)$ to $(\Omega_3, \mathcal{F}_3)$. Then the composition $g \circ f$ is a measurable function from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_3, \mathcal{F}_3)$.*

PROOF. Let $E \in \mathcal{F}_3$. Because $g : (\Omega_2, \mathcal{F}_2) \to (\Omega_3, \mathcal{F}_3)$ is measurable, the set $g^{-1}(E)$ is $\mathcal{F}_2$-measurable. Next, because $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ is measurable, the set $f^{-1}(Y^{-1}(E))$ is $\mathcal{F}_3$-measurable. Since $E \in \mathcal{F}_3$ was arbitrary, it follows that for all $E \in \mathcal{F}_3$,

$$(g \circ f)^{-1}(E) = f^{-1}(g^{-1}(E)) \in \mathcal{F}_1.$$

Hence, $g \circ f$ is measurable as a function from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_3, \mathcal{F}_3)$. $\square$

5.4. σ-algebras generated by functions

Sometimes, we are given a space Ω and a family of functions $f_\alpha : \Omega \to E_\alpha$, for α in some (possibly uncountable) index set $\mathcal{I}$. If there is a σ-algebra $\mathcal{G}_\alpha$ on E_α, we can ensure that all the functions f_α are measurable, by choosing a suitable σ-algebra on Ω. Of course, the σ-algebra 2^Ω would always do the job, but we mentioned already that this one usually has way too many sets contained in it. So instead, we want to look for the smallest σ-algebra such that all the f_α are measurable.

PROPOSITION 5.4.1. *Let $\{f_\alpha\}_{\alpha\in\mathcal{I}}$ be a family of functions $f_\alpha : \Omega \to (E_\alpha, \mathcal{G}_\alpha)$, where $\mathcal{G}_\alpha$ is a σ-algebra on E_α, for α in some index set $\mathcal{I}$. The σ-algebra generated by the sets $f_\alpha^{-1}(B)$ where $\alpha \in \mathcal{I}$ and $B \in \mathcal{G}_\alpha$, that is*

$$\mathcal{H} = \sigma\big(\{f_\alpha^{-1}(B) :\ \alpha \in \mathcal{I},\ B \in \mathcal{G}_\alpha\}\big)$$

is the smallest σ-algebra on Ω such that f_α is $(\mathcal{H}, \mathcal{G}_\alpha)$-measurable for every $\alpha \in \mathcal{I}$. It is called the σ-algebra generated by the family $\{f_\alpha\}_{\alpha\in\mathcal{I}}$ and often denoted as

$$\sigma\big(\{f_\alpha :\ \alpha \in \mathcal{I}\}\big) := \sigma\big(\{f_\alpha^{-1}(B) :\ \alpha \in \mathcal{I},\ B \in \mathcal{G}_\alpha\}\big).$$

PROOF. Let $\mathcal{F}$ be a σ-algebra on Ω such that for every $\alpha \in \mathcal{I}$, the function $f_\alpha : \Omega \to E_\alpha$ is $(\mathcal{F}, \mathcal{G}_\alpha)$-measurable. Then for every $\alpha \in \mathcal{I}$ and $B \in \mathcal{G}_\alpha$, we have

$$f_\alpha^{-1}(B) \in \mathcal{F}$$

That is,

$$\{f_\alpha^{-1}(B) :\ \alpha \in \mathcal{I},\ B \in \mathcal{G}_\alpha\} \subset \mathcal{F},$$

and since the σ-algebra generated by the collection on the left-hand side is the smallest σ-algebra that contains the collection, we have

$$\sigma\big(\{f_\alpha^{-1}(B) :\ \alpha \in \mathcal{I},\ B \in \mathcal{G}_\alpha\}\big) \subset \mathcal{F},$$

In particular, f_α is $(\mathcal{H}, \mathcal{G}_\alpha)$-measurable for every $\alpha \in \mathcal{I}$. $\square$

EXERCISE 5.4.2. Let Ω be a set and let $(E, \mathcal{G})$ be a measurable space. Let $f : \Omega \to E$ be a function. Show that the σ-algebra generated by *just* the function f has the following easy characterization

$$\sigma(\{f\}) = \{f^{-1}(B) :\ B \in \mathcal{G}\}.$$

EXAMPLE 5.4.3. Suppose $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measure spaces. We let $\Omega_1 \times \Omega_2$ play the role of Ω in the explanation above. There are two projections, $\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i$, $i = 1, 2$ given by

$$\pi_1(x, y) = x \qquad \pi_2(x, y) = y.$$

The smallest σ-algebra such that π_1 and π_2 are measurable is called the product σ-algebra on $\Omega_1 \times \Omega_2$. We will denote it as $\mathcal{F}_1 \otimes \mathcal{F}_2$, and it is defined in formulas as

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\{\pi_1, \pi_2\}).$$

EXERCISE 5.4.4. We have defined the product σ-algebra before in Example 2.2.2. Show that the two definitions coincide.

PROPOSITION 5.4.5. *Let $\{f_\alpha\}_{\alpha\in\mathcal{I}}$ be a family of functions $f_\alpha : A \to (E_\alpha, \mathcal{G}_\alpha)$, and let $\mathcal{C}_\alpha$ be a collection of sets on E_α such that $\mathcal{G}_\alpha = \sigma(\mathcal{C}_\alpha)$, for α in some index set $\mathcal{I}$. Then,*

$$\sigma(\{f_\alpha : \alpha \in I\}) = \sigma(\{f_\alpha^{-1}(B) :\ \alpha \in \mathcal{I},\ B \in \mathcal{C}_\alpha\}).$$

PROOF. Define for ease of notation

$$\mathcal{F}_1 := \sigma(\{f_\alpha : \alpha \in \mathcal{I}\})$$
$$\mathcal{F}_2 := \sigma(\{f_\alpha^{-1}(B) : \alpha \in \mathcal{I},\ B \in \mathcal{C}_\alpha\}).$$

It is clear from the definition of $\mathcal{F}_1$ that $\mathcal{F}_2 \subset \mathcal{F}_1$. Moreover, f_α is $(\mathcal{F}_2, \mathcal{G}_\alpha)$-measurable for every $\alpha \in \mathcal{I}$ by Lemma 5.2.1. But $\mathcal{F}_1$ is the smallest σ-algebra $\mathcal{F}$ such that f_α is $(\mathcal{F}, \mathcal{G}_\alpha)$-measurable for every $\alpha \in \mathcal{I}$. Therefore also $\mathcal{F}_1 \subset \mathcal{F}_2$. □

EXAMPLE 5.4.6. In probability theory, one often considers sequences of random variables (think of coin tossing problems). Thought of it differently, one considers sequence-valued random variables. Let $(E, \mathcal{G})$ be a measure space, and let Ω be the space of sequences $(x_i)_{i\in\mathbb{N}}$ taking values in E, that is

$$\Omega = E^{\mathbb{N}}.$$

For a sequence $(x_i) \in \Omega$, define the projection $\pi_n : \Omega \to E$ by $\pi_n((x_i)) = x_n$. Then the product σ-algebra on Ω is

$$\bigotimes_{i=1}^{\infty} \mathcal{G} := \sigma(\{\pi_n :\ n \in \mathbb{N}\}).$$

Of course, it is rather unclear what kind of monster this creates at this stage. ♢

DEFINITION 5.4.7. A filtration is an ordered family of σ-algebras $\{\mathcal{F}_\alpha\}$ such that $\mathcal{F}_\alpha \subset \mathcal{F}_\beta$, $\alpha \leq \beta$. A family $\{f_\alpha\}$ of measurable functions such that f_α is $\mathcal{F}_\alpha$-measurable is said to be *adapted* with respect to the filtration $\{\mathcal{F}_\alpha\}$.

EXAMPLE 5.4.8. In the notation of the previous example, probably the most common filtration is

$$\mathcal{F}_n := \sigma(\{\pi_j :\ j \in \{1, \ldots, n\}\}).$$

EXAMPLE 5.4.9. Similar to the previous example, we can consider the space of continuous paths/curves $\Omega := C([0,T]; \mathbb{R}^d)$ the family of functions $\{f_t\}_{t\in[0,T]}$ defined by $f_t(\omega) = \omega_t \in \mathbb{R}^d$. Then, $\mathcal{F} := \sigma(\{f_t : t \in [0,T]\})$ is the smallest σ-algebra that makes the time-evaluation function f_t $(\mathcal{F}, \mathcal{B}_{\mathbb{R}^d})$-measurable for all $t \in [0,T]$ and the family $\{\mathcal{F}_t\}_{t\in[0,T]}$ with $\mathcal{F}_t := \sigma(\{f_s : s \in [0,t]\})$ is a natural filtration for $\mathcal{F}$ since $\{f_t\}_{t\in[0,T]}$ is adapted w.r.t. to $\{\mathcal{F}_t\}_{t\in[0,T]}$.

A large class of continuous stochastic processes can be seen in this way. ♢

5.5. Measurability if the target is the real line

We now specify the case in which the target of the measurable functions is the real line endowed with the Borel σ-algebra, $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The following lemma will make it much easier to check the measurability of functions.

LEMMA 5.5.1. *Let f and g be measurable functions from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then, the following functions are measurable as well*

(i) $f + g$
(ii) αf, *for* $\alpha \in \mathbb{R}$
(iii) fg
(iv) f/g *if* $g(\omega) \neq 0$ *for all* $\omega \in \Omega$.

PROOF. We show (i). By Proposition 2.3.2, we know that $\mathcal{B}_{\mathbb{R}}$ is generated by intervals of the form $(-\infty, a]$ with $a \in \mathbb{Q}$. As a consequence, $\mathcal{B}_{\mathbb{R}}$ is also generated by intervals of the form (a, ∞) with $a \in \mathbb{Q}$. Therefore, by Lemma 5.2.1, it suffices to show that the set

$$\{\omega \in \Omega : f(\omega) + g(\omega) > a\}$$

is measurable for every $a \in \mathbb{Q}$. For brevity, we write $\{f + g > a\}$. The trick is to express this set as a countable union of sets of which we already know are measurable.

In fact, we will show that

$$\{f + g > a\} = \bigcup_{t \in \mathbb{Q}} \Big(\{f > t\} \cap \{g > a - t\} \Big).$$

We first show the inclusion '$\subset$'. If $\omega \in \Omega$ is such that

$$f(\omega) + g(\omega) > a,$$

then

$$f(\omega) > a - g(\omega),$$

so there exists some $t \in \mathbb{Q}$ such that

$$f(\omega) > t > a - g(\omega),$$

and thus $f(\omega) > t$ and $g(\omega) > a - t$. So in that case

$$\omega \in \bigcup_{t \in \mathbb{Q}} \Big(\{f > t\} \cap \{g > a - t\} \Big).$$

Now we will show the inclusion '$\supset$'. Let $\omega \in \Omega$ be such that $f(\omega) > t$ and $g(\omega) > a - t$. Then, by adding the inequalities, we know that $f(\omega) + g(\omega) > a$. □

One can formulate similar statements for functions that take extended real values, but needs to be careful that the sum of two functions taking values in $\overline{\mathbb{R}}$ is in general not defined.

LEMMA 5.5.2. *Let f, g, f_i, for $i \in \mathbb{N}$, be measurable functions from $(\Omega, \mathcal{F})$ to $(\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$. Then the following functions are measurable as well.*

(i) $\max(f, g)$ *and* $\min(f, g)$
(ii) $\sup_i f_i$ *and* $\inf_i f_i$.
(iii) $\limsup_{i \to \infty} f_i$
(iv) $\liminf_{i \to \infty} f_i$

PROOF. We show (ii) and (iii). The other statements can either be shown in a similar way or can be derived easily as a corollary. Let $f_i : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$, for $i \in \mathbb{N}$, be a measurable function. Let $a \in \mathbb{Q}$. Just as in the proof of the previous proposition, by Lemma 5.2.1, and the definition of $\mathcal{B}_{\overline{\mathbb{R}}}$ in Definition 2.3.3, it suffices to check that the set

$$\Big\{\omega \in \Omega : \sup_{i \in \mathbb{N}} f_i(\omega) \leq a \Big\} \qquad \text{is measurable.}$$

Now, this set has an easy description in terms of the sublevel sets of the functions f_i, namely

$$\Big\{\omega \in \Omega : \sup_{i \in \mathbb{N}} f_i(\omega) \leq a \Big\} = \bigcap_{i=1}^{\infty} \big\{\omega \in \Omega : \ f_i(\omega) \leq a \big\}.$$

Because for every $i \in \mathbb{N}$, the function f_i is measurable, the set

$$\big\{\omega \in \Omega : \ f_i(\omega) \le a\big\} \qquad \text{is measurable.}$$

As the intersection of countably many measurable sets, the set

$$\Big\{\omega \in \Omega : \sup_{i \in \mathbb{N}} f_i(\omega) \le a\Big\}$$

is therefore measurable as well.

Since $\inf_i f_i = -\sup_i(-f_i)$, and $(-f_i)$ is measurable by the previous proposition, the function $\inf_i f_i$ is measurable as well.

Finally, recall that

$$\limsup_{i\to\infty} f_i(\omega) = \lim_{i\to\infty} \sup_{k\ge i} f_k(\omega) = \inf_{i\in\mathbb{N}} \left(\sup_{k\ge i} f_k(\omega)\right).$$

By the first part of the proof, we know that the function $g_i := \sup_{k\ge i} f_k$ is measurable for every $i \in \mathbb{N}$, and in turn this implies that the function

$$\limsup_{i\to\infty} f_i = \inf_{i\in\mathbb{N}} g_i$$

is measurable. □

5.6. Lebesgue-measurable functions

We say that a subset of $\mathbb{R}^n$ is Lebesgue-measurable if and only if it is an element of $\mathcal{L}_\lambda$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called Lebesgue-measurable if it is $(\mathcal{L}_\lambda, \mathcal{B}_\mathbb{R})$-measurable.

In a way that can be made quite precise, there are many more Lebesgue-measurable sets than there are Borel-measurable sets, and consequently, there are many more Lebesgue-measurable functions than Borel-measurable functions. It depends a bit on the literature whether a measurable function is meant to be Borel-measurable or Lebesgue-measurable.

The following proposition is one of the advantages of working with Lebesgue-measurable rather than Borel-measurable functions.

Proposition 5.6.1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue-measurable, and $g \equiv f$ almost everywhere, then g is Lebesgue-measurable.*

Exercise 5.6.2. Prove Proposition 5.6.1.

The next proposition is a statement about the difference of the classes of Lebesgue-measurable and Borel-measurable functions.

Proposition 5.6.3. *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue-measurable, then there exists a Borel-measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that $f \equiv g$ almost everywhere*

Exercise 5.6.4. Prove Proposition 5.6.3.

To get back to the warning in Section 5.3, *the composition of two Lebesgue-measurable functions from $\mathbb{R}$ to $\mathbb{R}$ is in general not measurable.*

5.7. The push-forwards and the law of a random variable

If $\mathbb{P}$ is a probability measure and X is a random variable, then the *law* of the random variable under $\mathbb{P}$ is, in measure-theoretic terms, called the push-forward of $\mathbb{P}$ along X. More precisely, we have the following:

DEFINITION 5.7.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(E, \mathcal{G})$ be a measurable space and $f : \Omega \to E$ be $(\mathcal{F}, \mathcal{G})$-measurable. Then the *push-forward* of μ along f, denoted by $f_{\#}\mu$, is defined by

$$f_{\#}\mu(B) \coloneqq \mu(X^{-1}(B)) \qquad \forall\, B \in \mathcal{G},$$

PROPOSITION 5.7.2. *The push-forward $f_{\#}\mu$ is indeed a measure.*

PROOF. It is clear that $f_{\#}\mu(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$. Suppose a sequence of mutually disjoint sets $B_i \in \mathcal{G}$, $i \in \mathbb{N}$, is given, and consider

$$f_{\#}\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu\left(f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mu\left(\bigcup_{i=1}^{\infty} f^{-1}(B_i)\right) = \sum_{i=1}^{\infty} f_{\#}\mu(B_i). \qquad \square$$

EXAMPLE 5.7.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable. Then we say that X has law ν if $X_{\#}\mathbb{P} = \nu$, i.e. if $\nu(B) = \mathbb{P}(X \in B)$ for all $B \in \mathcal{B}_{\mathbb{R}}$.

If X is a *Gaussian random variable* with mean 0 and variance 1, then it simply means

$$\mathbb{P}(X \in B) = \frac{1}{\sqrt{2\pi}} \int_B e^{-x^2/2}\, \lambda(\mathsf{d}x) = \mathscr{N}_{0,1}(B), \qquad B \in \mathcal{B}_{\mathbb{R}},$$

where λ is the Lebesgue measure and $\mathscr{N}_{0,1}$ is the *Gaussian measure* on $\mathbb{R}$ with mean 0 and variance 1. In particular, the definition of the Gaussian measure requires a meaningful way to integrate w.r.t. the Lebesgue measure λ. This is the goal of the following chapter.

CHAPTER 6

The Lebesgue Integral

6.1. Simple functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

DEFINITION 6.1.1. A function $f : \Omega \to \mathbb{R}$ is called *simple* if it takes the form

$$f = \sum_{i=1}^{N} a_i \mathbf{1}_{A_i}$$

for some positive integer $N \in \mathbb{N}$, disjoint measurable sets $A_1, \ldots, A_N \in \mathcal{F}$ and constants $a_1, \ldots, a_N \in \mathbb{R}$.

6.2. The integral of a simple function

For a non-negative simple function

$$g = \sum_{i=1}^{N} a_i \mathbf{1}_{A_i}, \qquad a_i \geq 0,$$

we define the integral of g with respect to μ by

$$\int_\Omega g \, \mathsf{d}\mu \coloneqq \sum_{i=1}^{N} a_i \mu(A_i).$$

A priori there could be different representations of the same simple function, so we should check that the integral of a simple function is well-defined. This follows however, because g actually has a unique representation

$$g = \sum_{i=1}^{M} b_i \mathbf{1}_{B_i}, \qquad \text{for which } b_i < b_{i+1}.$$

By the finite additivity of the measure μ,

$$\sum_{i=1}^{N} a_i \mu(A_i) = \sum_{i=1}^{M} b_i \mu(B_i).$$

REMARK 6.2.1. In case $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and X is a simple, real-valued random variable on Ω having the representation

$$X = \sum_{i=1}^{N} a_i \mathbf{1}_{A_i},$$

with mutually disjoint $A_i \in \mathcal{F}$ and $a_i \in \mathbb{R}$, the integral is usually called the *expectation* value of X and is written as

$$\mathbb{E}[X] := \int_\Omega X \,\mathsf{d}\mathbb{P} = \sum_{i=1}^N a_i \mathbb{P}(A_i).$$

6.3. The Lebesgue integral of nonnegative functions

Given a measurable function $f : (\Omega, \mathcal{F}) \to ([0,+\infty], \mathcal{B}_{[0,+\infty]})$, the μ-integral of f over Ω is defined by

$$\int_\Omega f \,\mathsf{d}\mu := \sup \left\{ \int_\Omega g \,\mathsf{d}\mu : \; g \text{ simple}, \; 0 \le g \le f \right\}.$$

For a measurable set $A \in \mathcal{F}$, we use the following notation and definition for integration of f over the set A

$$\int_A f \,\mathsf{d}\mu := \int_\Omega \mathbf{1}_A f \,\mathsf{d}\mu.$$

If we denote by f_A the restriction of f to A, and by μ_A the restriction of μ to $\mathcal{F}_A$, then

$$\int_A f_A \,\mathsf{d}\mu_A = \int_A f \,\mathsf{d}\mu.$$

Similarly, if $f_A : (A, \mathcal{F}_A) \to ([0,+\infty], \mathcal{B}_{[0,+\infty]})$ is measurable, and f is a measurable extension of f_A to the whole of Ω, then

$$\int_A f \,\mathsf{d}\mu = \int_A f_A \, d\mu_A.$$

Proposition 6.3.1 (Properties of the Lebesgue integral of nonnegative functions). *Let f, g be two nonnegative, measurable functions. Let $\lambda \ge 0$ be a constant.*

(i) *(absolute continuity) If $B \in \mathcal{F}$ satisfies $\mu(B) = 0$, then*

$$\int_B f \,\mathsf{d}\mu = 0.$$

(ii) *(monotonicity) If $f \le g$, then*

$$\int_\Omega f \,\mathsf{d}\mu \le \int_\Omega g \,\mathsf{d}\mu.$$

(iii) *(homogeneity)*

$$\lambda \int_\Omega f \,\mathsf{d}\mu = \int_\Omega (\lambda f) \,\mathsf{d}\mu.$$

6.4. The monotone convergence theorem

Theorem 6.4.1 (Monotone convergence theorem I). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f_n : (\Omega, \mathcal{F}) \to ([0,+\infty], \mathcal{B}_{[0,+\infty]})$, $n \in \mathbb{N}$, be a sequence of nonnegative, measurable functions, such that $f_n(\omega) \le f_{n+1}(\omega)$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$. Define the function*

$$f(\omega) := \lim_{n \to \infty} f_n(\omega), \qquad \omega \in \Omega.$$

Then f is measurable and

$$\lim_{n \to \infty} \int_\Omega f_n \,\mathsf{d}\mu = \int_\Omega f \,\mathsf{d}\mu.$$

PROOF. From the monotonicity of the integral, we may immediately conclude that

$$\limsup_{n\to\infty} \int_\Omega f_n \,\mathsf{d}\mu \le \int_\Omega f \,\mathsf{d}\mu.$$

Hence, we are left to show that

$$\liminf_{n\to\infty} \int_\Omega f_n \,\mathsf{d}\mu \ge \int_\Omega f \,\mathsf{d}\mu.$$

This is obvious if $\int_\Omega f \,\mathsf{d}\mu = 0$, so we assume that $\int_\Omega f \,\mathsf{d}\mu := L > 0$.

By the definition of the integral, for every $0 < \varepsilon < L$, there exists a nonnegative simple function $g : \Omega \to \mathbb{R}$ such that $0 \le g \le f$ on Ω and

$$\int_\Omega g \,\mathsf{d}\mu > \int_\Omega f \,\mathsf{d}\mu - \varepsilon.$$

Because g is simple, there exist an $N \in \mathbb{N}$, nonnegative constants $a_i \in (0,\infty)$ and disjoint, measurable sets $A_i \subset \mathcal{F}$ such that

$$g = \sum_{i=1}^N a_i \mathbf{1}_{A_i}.$$

Moreover, we find some $\delta > 0$, such that

$$g_\delta := \sum_{i=1}^N (a_i - \delta)\mathbf{1}_{A_i},$$

satisfies

$$\int_\Omega g_\delta \,\mathsf{d}\mu = \sum_{i=1}^N (a_i - \delta)\,\mu(A_i) \ge \int_\Omega f \,\mathsf{d}\mu - \varepsilon.$$

Now define for $i \in \{1,\dots,N\}$ and $n \in \mathbb{N}$ the measurable set

$$G_n^i := \big\{x \in A_i : \; f_n(x) \ge a_i - \delta\big\}.$$

Then, because $f_n \le f_{n+1}$, we have $G_n^i \subset G_{n+1}^i$ for all $n \in \mathbb{N}$ and by the pointwise convergence of f_n to f, we have

$$\bigcup_{n=1}^\infty G_n^i = A_i, \qquad i = 1,\dots,N.$$

Hence, by the continuity from below of measures

$$\lim_{n\to\infty} \mu(G_n^i) = \mu(A_i).$$

Since for every $n \in \mathbb{N}$,

$$\int_\Omega f_n \,\mathsf{d}\mu \ge \sum_{i=1}^N \int_{A_i} f_n \,\mathsf{d}\mu \ge \sum_{i=1}^N \int_{A_i} f_n \,\mathsf{d}\mu \ge \sum_{i=1}^N \int_{G_n^i} f_n \,\mathsf{d}\mu \ge \sum_{i=1}^N (a_i - \delta)\,\mu(G_n^i),$$

we find that

$$\liminf_{n\to\infty} \int_\Omega f_n \,\mathsf{d}\mu \ge \liminf_{n\to\infty} \sum_{i=1}^N (a_i - \delta)\mu(G_n^i) = \int_\Omega g_\delta \,\mathsf{d}\mu \ge \int_\Omega f \,\mathsf{d}\mu - \varepsilon.$$

Because $\varepsilon > 0$ was arbitrary, it follows that

$$\liminf_{n\to\infty} \int_\Omega f_n \mathsf{d}\mu \geq \int_\Omega f \mathsf{d}\mu. \qquad \square$$

6.5. Approximation by simple functions

In this section, we will give a few explicit approximations to arbitrary measurable functions. First consider a nonnegative measurable function $f : (\Omega, \mathcal{F}) \to ([0,\infty], \mathcal{B}_{[0,\infty]})$. We define the function $(f_n)_{n\in\mathbb{N}}$ by setting $f_n(\omega) = 0$ if $f(\omega) = 0$,

$$f_n(\omega) \coloneqq k\, 2^{-n} \qquad \text{if} \quad f(\omega) \in [k\, 2^{-n}, (k+1)\, 2^{-n}),$$

for some $k \in \mathbb{N} \cup \{0\}$ and setting $f_n(\omega) = +\infty$ if $f(\omega) = +\infty$. Note that we can write

$$f_n = +\infty \mathbf{1}_{\{f=+\infty\}} + \sum_{k=0}^{\infty} k\, 2^{-n} \mathbf{1}_{\{k\, 2^{-n} \leq f < (k+1)\, 2^{-n}\}}, \qquad n \in \mathbb{N}$$

and easily deduce that f_n is measurable for every $n \in \mathbb{N}$.

The advantage of the approximation f_n to f is most clearly seen when $f(\omega) < +\infty$ for all $\omega \in \Omega$. In this case, f_n converges to f uniformly: In fact

$$|f_n(\omega) - f(\omega)| \leq 2^{-n}$$

for all $n \in \mathbb{N}$ and all $\omega \in \Omega$.

The disadvantage of the approximation f_n is that if f is unbounded, the approximation f_n is not simple. To remedy this, we truncate f_n to get the approximation

$$[f]_n \coloneqq \min(2^n, f_n).$$

The function $[f]_n$ is indeed simple.

Both the approximations f_n and $[f]_n$ are nondecreasing in n. Moreover, they are pointwise approximations of the functions f. In particular, the function f_n converges uniformly to f on the set where f is finite, and the functions $[f]_n$ converge uniformly to f on any set on which f is bounded.

Exercise 6.5.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : (\Omega, \mathcal{F}) \to ([0,+\infty), \mathcal{B}_{[0,+\infty)})$. Show that

$$\int_\Omega f \,\mathsf{d}\mu = \lim_{n\to\infty} \int_\Omega [f]_n \,\mathsf{d}\mu.$$

6.6. Additivity of the Lebesgue integral of a nonnegative function

Proposition 6.6.1 (Additivity of the Lebesgue integral of nonnegative functions). *Let f, g be two nonnegative, measurable functions. Then*

$$\int_\Omega (f+g) \,\mathsf{d}\mu = \int_\Omega f \,\mathsf{d}\mu + \int_\Omega g \,\mathsf{d}\mu.$$

Proof. For simple functions, the additivity of the integral is easy to check. Therefore for every $n \in \mathbb{N}$,

$$\int_\Omega ([f]_n + [g]_n) \,\mathsf{d}\mu = \int_\Omega [f]_n \,\mathsf{d}\mu + \int_\Omega [g]_n \,\mathsf{d}\mu.$$

We now take the limit on both sides of the equation. On one hand, the functions $[f]_n + [g]_n$ are increasing in n, and converge pointwise to $(f+g)$. By the monotone convergence theorem,

$$\lim_{n\to\infty} \int_\Omega ([f]_n + [g]_n)\,\mathsf{d}\mu = \int_\Omega (f+g)\,\mathsf{d}\mu.$$

On the other hand, by a limit theorem and Exercise 6.5.1, we know that

$$\lim_{n\to\infty} \left(\int_\Omega [f]_n\,\mathsf{d}\mu + \int_\Omega [g]_n\,\mathsf{d}\mu \right) = \int_\Omega f\,\mathsf{d}\mu + \int_\Omega g\,\mathsf{d}\mu.$$

Therefore,

$$\int_\Omega (f+g)\,\mathsf{d}\mu = \int_\Omega f\,\mathsf{d}\mu + \int_\Omega g\,\mathsf{d}\mu. \qquad \square$$

6.7. Integrable functions

The next goal is to define the integral of functions f that are not necessarily nonnegative. We can only do this if the integral of $|f|$ is finite.

Definition 6.7.1. A measurable function $f : \Omega \to \mathbb{R}$ is *μ-integrable* if

$$\int_\Omega |f|\,\mathsf{d}\mu < +\infty.$$

For any function $f : \Omega \to \overline{\mathbb{R}}$, we define its positive part f^+ and negative part f^- as

$$f^+(\omega) \coloneqq \max(f(\omega), 0), \qquad f^-(\omega) \coloneqq -\min(f(\omega), 0)$$

It follows that $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

The *Lebesgue integral* of a μ-integrable function $f : \Omega \to \mathbb{R}$ is

$$\int_\Omega f\,\mathsf{d}\mu \coloneqq \int_\Omega f^+\,\mathsf{d}\mu - \int_\Omega f^-\,\mathsf{d}\mu.$$

Proposition 6.7.2. *Let f, g be two μ-integrable functions and $\alpha \in \mathbb{R}$ be a constant.*

(i) *(Absolute continuity) If $B \in \mathcal{F}$ satisfies $\mu(B) = 0$, then*

$$\int_B f\,\mathsf{d}\mu = 0.$$

(ii) *(Monotonicity) If $f \le g$ μ-a.e., then*

$$\int_\Omega f\,\mathsf{d}\mu \le \int_\Omega g\,\mathsf{d}\mu.$$

(iii) *(Homogeneity)*

$$\alpha \int_\Omega f\,\mathsf{d}\mu = \int_\Omega (\alpha f)\,\mathsf{d}\mu.$$

(iv) *(Additivity)*

$$\int_\Omega (f+g)\,\mathsf{d}\mu = \int_\Omega f\,\mathsf{d}\mu + \int_\Omega g\,\mathsf{d}\mu.$$

Definition 6.7.3. We say that a measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ is integrable on a set $A \in \mathcal{F}$ if $\mathbf{1}_A f$ is integrable on Ω. Equivalently, we say f is integrable if the restriction $f|_A$ is integrable on the measure space $(A, \mathcal{F}_A, \mu|_A)$.

6.8. Change of variables formula

Proposition 6.8.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ be measurable and $h : (E, \mathcal{G}) \to ([0, \infty], \mathcal{B}_{[0,\infty]})$ be measurable. Then*

$$\int_\Omega h \circ f \, \mathsf{d}\mu = \int_E h \, \mathsf{d}(f_\# \mu).$$

Proof. We first show that statement when h is simple and nonnegative, i.e.

$$h = \sum_{i=1}^N a_i \mathbf{1}_{A_i}$$

for some $N \in \mathbb{N}$, $a_i \in (0, \infty)$, and $A_i \in \mathcal{F}$ mutually disjoint. Then

$$h \circ f = \sum_{i=1}^N a_i \mathbf{1}_{f^{-1}(A_i)}.$$

It follows that

$$\int_\Omega h \circ f \, \mathsf{d}\mu = \sum_{i=1}^N a_i \, \mu(f^{-1}(A_i)) = \sum_{i=1}^N a_i \, (f_\# \mu)(A_i) = \int_E h \, \mathsf{d}(f_\# \mu),$$

which shows the proposition in the case when h is simple and nonnegative.

We now turn to the case in which h is a general, nonnegative measurable function. Note that $[h]_n \circ f$ is a nondecreasing sequence of functions, which converges pointwise to $h \circ f$. By the monotone convergence theorem,

$$\int_\Omega h \circ f \, \mathsf{d}\mu = \lim_{n \to \infty} \int_\Omega [h]_n \circ f \, \mathsf{d}\mu = \lim_{n \to \infty} \int_E [h]_n \, \mathsf{d}(f_\# \mu) = \int_E h \, \mathsf{d}(f_\# \mu). \qquad \square$$

As a direct consequence, we have the following proposition.

Proposition 6.8.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ be measurable and $h : (E, \mathcal{G}) \to (\mathbb{R}, \mathcal{B})$ be measurable. Then $h \circ f$ is integrable with respect to μ if and only if h is integrable with respect to $f_\# \mu$, in which case,*

$$\int_\Omega h \circ f \, \mathsf{d}\mu = \int_E h \, \mathsf{d}(f_\# \mu).$$

Example 6.8.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be a continuous random variable. Assume that the law of X can be represented by a *Lebesgue density function* $\varrho : (\mathbb{R}, \mathcal{B}) \to ([0, \infty), \mathcal{B}_{[0,\infty)})$, i.e. $X_\# \mathbb{P}$ is the (unique) measure given by

$$(X_\# \mathbb{P})((a, b]) = \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \in (a, b]\right\}\right) = \int_a^b \varrho \, \mathsf{d}\lambda.$$

Let $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ be a bounded measurable function. Then by the change of variables formula,

$$\mathbb{E}[h(X)] = \int_\mathbb{R} h \, \mathsf{d}(X_\# \mathbb{P}).$$

On the other hand, the set function

$$\nu : \mathcal{B} \to [0, +\infty], \qquad \nu(A) := \int_A \varrho \, \mathsf{d}\lambda$$

is a measure on the Borel σ-algebra (see exercise). From the uniqueness statement in Theorem 3.4.1, we conclude that $\nu = X_{\#}\mathbb{P}$. For simple functions $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ one can now check that

$$\int_{\mathbb{R}} g \, \mathsf{d}\nu = \int_{\mathbb{R}} g\varrho \, \mathsf{d}\lambda.$$

By approximating an arbitrary bounded Borel-measurable function h by a sequence of simple functions and applying the *dominated convergence theorem* (see next chapter), we conclude that also for arbitrary bounded measurable functions h,

$$\int_{\mathbb{R}} h \, \mathsf{d}\nu = \int_{\mathbb{R}} h\varrho \, \mathsf{d}\lambda$$

and thus

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h \, \mathsf{d}(X_{\#}\mathbb{P}) = \int_{\mathbb{R}} h \, \mathsf{d}\nu = \int_{\mathbb{R}} h\varrho \, \mathsf{d}\lambda.$$

6.9. The Markov inequality

The following Lemma states the Markov inequality. The trick used in the proof can be used to obtain many similar inequalities.

LEMMA 6.9.1 (The Markov inequality). *Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space and let f be a μ-integrable function. For any $t > 0$,*

$$\mu\big(\{\omega \in \Omega : \ |f|(\omega) \geq t\}\big) \leq \frac{1}{t} \int_{\Omega} |f| \, \mathsf{d}\mu.$$

PROOF. The result follows easily from

$$\int_{\Omega} |f| \, \mathsf{d}\mu \geq \int_{\{|f| \geq t\}} |f| \, \mathsf{d}\mu \geq t\, \mu\big(\{|f| \geq t\}\big)$$ □

In probability language, the Markov inequality looks as follows.

LEMMA 6.9.2 (The Markov inequality). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable. For any $t > 0$,*

$$\mathbb{P}(|X| \geq t) \leq \frac{1}{t}\mathbb{E}[|X|].$$

CHAPTER 7

The convergence theorems

7.1. The (first) Borel-Cantelli Lemma

Lemma 7.1.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let (A_i) be a sequence of measurable sets, that is $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$. Assume that*

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

Then for μ-almost every $\omega \in \Omega$, there are only finitely many $i \in \mathbb{N}$ such that $\omega \in A_i$.

Proof. Define the sets

$$B_j := \bigcup_{i \geq j} A_i.$$

Clearly the sequence $(B_j)_{j \in \mathbb{N}}$ is decreasing.

Now let U denote the set of ω such that there are infinitely many $i \in \mathbb{N}$ with $\omega \in A_i$, i.e.

$$U := \Big\{ \omega \in \Omega \,:\, \omega \in A_i \text{ for infinitely many } i \in \mathbb{N} \Big\}.$$

Then $U \subset B_j$ for every $j \in \mathbb{N}$ and $U \in \mathcal{F}$ since U can be written as

$$U = \bigcap_{j \in \mathbb{N}} \bigcup_{i \geq j} A_i.$$

By assumption, and the σ-subadditivity of μ,

$$\mu(B_1) = \mu\left(\bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu(A_i) < +\infty.$$

Moreover, the summability also gives

$$\lim_{j \to \infty} \mu(B_j) \leq \limsup_{j \to \infty} \sum_{i=j}^{\infty} \mu(A_i) = 0.$$

Hence, by the continuity from above of μ, we obtain

$$\mu(U) \leq \mu\left(\bigcup_{j=1}^{\infty} B_j \right) = \lim_{j \to \infty} \mu(B_j) = 0,$$

i.e. U is a null set. $\square$

7.2. Convergence in measure

DEFINITION 7.2.1. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ and $f : \Omega \to \mathbb{R}$ be a function. We say that the sequence $(f_n)_{n\in\mathbb{N}}$ converges to f in μ-measure if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mu(\{\omega \in \Omega : \ |f_n(\omega) - f(\omega)| \geq \epsilon\}) = 0.$$

PROPOSITION 7.2.2. *If a sequence $(f_n)_{n\in\mathbb{N}}$ of measurable functions from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ converges to a function $f : \Omega \to \mathbb{R}$ in μ-measure, then there exists a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that f_{n_k} converges to f μ-almost everywhere.*

PROOF. We claim that there exists an (increasing) index sequence $n_1, n_2, \ldots$ in $\mathbb{N}$, such that for all $k \in \mathbb{N}$

$$\mu\left(\left\{\omega \in \Omega : \ |f_{n_k}(\omega) - f(\omega)| \geq \frac{1}{k}\right\}\right) < \frac{1}{2^k}.$$

We will define the index sequence inductively. Because $(f_n)_{n\in\mathbb{N}}$ converges to f in measure, there exists an $n_1 \in \mathbb{N}$ such that

$$\mu\left(\left\{\omega \in \Omega : \ |f_{n_1}(\omega) - f(\omega)| \geq \frac{1}{1}\right\}\right) < \frac{1}{2^1}.$$

Now assume $n_1, \ldots, n_{k-1}$ are defined for some $k \geq 2$. Again because $f_n \to f$ in measure, there exists an $n_k \in \mathbb{N}$, $n_k > n_{k-1}$ such that

$$\mu\left(\left\{\omega \in \Omega : \ |f_{n_k}(\omega) - f(\omega)| \geq \frac{1}{k}\right\}\right) < \frac{1}{2^k}.$$

This proves the claim. For $k \in \mathbb{N}$, define the set

$$A_k := \left\{\omega \in \Omega : \ |f_{n_k}(\omega) - f(\omega)| \geq \frac{1}{k}\right\}.$$

Then we have by construction that

$$\sum_{k=1}^{\infty} \mu(A_k) \leq \sum_{k=1}^{\infty} 2^{-k} \leq 1 < +\infty.$$

Now let us define

$$U := \Big\{\omega \in \Omega \ : \ \omega \in A_k \text{ for infinitely many } k \in \mathbb{N}\Big\}.$$

It follows by the (first) Borel-Cantelli Lemma that

$$\mu(U) = 0.$$

Finally, let $\omega \in \Omega \backslash U$. Then there are only finitely many k such that $\omega \in A_k$. In other words, there exists a $K \in \mathbb{N}$ such that

$$0 \leq |f_{n_k}(\omega) - f(\omega)| < \frac{1}{k} \qquad \text{for every } k \geq K,$$

and by the squeeze theorem,

$$\lim_{k\to\infty} f_{n_k}(\omega) = f(\omega). \qquad \square$$

7.3. The monotone convergence theorem

THEOREM 7.3.1 (Monotone convergence theorem II). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative, measurable functions, and let f be a nonnegative measurable function such that for μ-almost every $\omega \in \Omega$,*

- $f_n(\omega) \le f_{n+1}(\omega)$ *for all* $n \in \mathbb{N}$
- $\lim_{n\to\infty} f_n(\omega) = f(\omega)$.

Then

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathsf{d}\mu = \int_\Omega f \, \mathsf{d}\mu.$$

PROOF. We will reduce to the case of the previous version of the monotone convergence theorem. By assumption, there exists a set $N \in \mathcal{F}$ such that $\mu(N) = 0$ and for all $\omega \in \Omega\backslash N$

- $f_n(\omega) \le f_{n+1}(\omega)$ for all $n \in \mathbb{N}$
- $\lim_{n\to\infty} f_n(\omega) = f(\omega)$.

Define $g_n(\omega) \coloneqq \max_{1\le k\le n} f_k(\omega)$. Then for *all* $\omega \in \Omega$, and all $n \in \mathbb{N}$, we have that $g_n(\omega) \le g_{n+1}(\omega)$. We define $g(\omega) \coloneqq \lim_{n\to\infty} g_n(\omega)$. Then $g_n(\omega) = f_n(\omega)$ and $g(\omega) = f(\omega)$ for every $\omega \in \Omega\backslash N$. Therefore,

$$\int_\Omega f_n \, \mathsf{d}\mu = \int_\Omega g_n \, \mathsf{d}\mu \qquad \text{and} \qquad \int_\Omega f \, \mathsf{d}\mu = \int_\Omega g \, \mathsf{d}\mu,$$

and the result follows from Theorem 6.4.1. □

THEOREM 7.3.2 (Monotone convergence theorem III). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative random variables on $(\Omega, \mathcal{F})$ and X be a nonnegative random variable on $(\Omega, \mathcal{F})$ such that for all $n \in \mathbb{N}$, $X_n \le X_{n+1}$ $\mathbb{P}$-almost surely and $X_n \to X$ $\mathbb{P}$-almost surely. Then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

7.4. Fatou's Lemma

THEOREM 7.4.1 (Fatou's Lemma I). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative, measurable functions, and define*

$$f \coloneqq \liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{k\ge n} f_k(x).$$

Then

$$\int_\Omega f \, \mathsf{d}\mu \le \liminf_{n\to\infty} \int_\Omega f_n \mathsf{d}\mu.$$

PROOF. We will base our proof on the monotone convergence theorem. Define the measurable functions $g_k : \Omega \to \mathbb{R}$ by

$$g_n(x) \coloneqq \inf_{k\ge n} f_k(x).$$

Then the functions g_n are monotonically nondecreasing, and

$$f = \lim_{n\to\infty} g_n.$$

The monotone convergence theorem yields

$$\int_\Omega f \, \mathsf{d}\mu = \lim_{n\to\infty} \int_\Omega g_n \, \mathsf{d}\mu.$$

Finally, we note that $g_n \le f_n$ for all $n \in \mathbb{N}$, so that

$$\int_\Omega f \, \mathsf{d}\mu \le \liminf_{n\to\infty} \int_\Omega f_n \, \mathsf{d}\mu. \qquad \square$$

REMARK 7.4.2. The inequality in Fatou's lemma can be strict. Consider for example the sequence of functions $f_n = \mathbf{1}_{[n,n+1)}$, $n \ge 1$. Clearly $f = \liminf_{n\to\infty} f_n = 0$, from which one obtains $\int_{\mathbb{R}} f \, \mathsf{d}\lambda = 0$, but $\int_{\mathbb{R}} f_n \, \mathsf{d}\lambda = 1$ for all $n \ge 1$.

THEOREM 7.4.3 (Fatou's Lemma II). *Let (X_n) be a sequence of nonnegative random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the nonnegative random variable X on $(\Omega, \mathcal{F})$ by*

$$X \coloneqq \liminf_{n\to\infty} X_n.$$

Then

$$\mathbb{E}[X] \le \liminf_{n\to\infty} \mathbb{E}[X_n].$$

7.5. The dominated convergence theorem

THEOREM 7.5.1 (The dominated convergence theorem). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions, and let f be a measurable function, such that $f_n \to f$ pointwise μ-almost everywhere. Moreover, assume there is a nonnegative, μ-integrable function $g : \Omega \to [0, \infty]$ such that $|f_n| \le g$ μ-almost everywhere. Then*

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathsf{d}\mu = \int_\Omega f \, \mathsf{d}\mu.$$

Before we give the proof, let us point out that g does not have to equal $|f|$.

PROOF. We will prove the dominated convergence theorem using Fatou's Lemma. We first prove the case where $f_n \to f$ *everywhere* and $|f_n| \le g$ *everywhere*.

Note that in that case the functions $f_n + g$ are nonnegative, so that Fatou's Lemma gives

$$\int_\Omega (f + g) \, \mathsf{d}\mu \le \liminf_{n\to\infty} \int_\Omega (f_n + g) \, \mathsf{d}\mu.$$

Hence, also

$$\int_\Omega f \, \mathsf{d}\mu \le \liminf_{n\to\infty} \int_\Omega f_n \, \mathsf{d}\mu.$$

Similarly, the functions $g - f_n$ are all nonnegative. Fatou's Lemma gives

$$\int_\Omega (g - f) \, \mathsf{d}\mu \le \liminf_{n\to\infty} \int_\Omega (g - f_n) \, \mathsf{d}\mu$$

from which we conclude that

$$\int_\Omega f \, \mathsf{d}\mu \ge \limsup_{n\to\infty} \int_\Omega f_n \mathsf{d}\mu.$$

It follows that

$$\lim_{n\to\infty}\int_\Omega f_n \,\mathsf{d}\mu = \int_\Omega f \,\mathsf{d}\mu.$$

In the general case, note that there exists a null set $N \in \mathcal{F}$ with $\mu(N) = 0$ such that $f_n \to f$ pointwise on $\Omega\backslash N$ as $n \to \infty$, and $|f_n| \leq g$ on $\Omega\backslash N$. Now define measurable functions $\widehat{f}_n, \widehat{f}, \widehat{g} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ as follows

$$\widehat{f}_n(\omega) := \begin{cases} f_n(\omega) & \omega \in \Omega\backslash N \\ 0 & \omega \in N \end{cases}, \qquad \widehat{f}(\omega) := \begin{cases} f(\omega) & \omega \in \Omega\backslash N \\ 0 & \omega \in N \end{cases},$$

and

$$\widehat{g}(\omega) := \begin{cases} g(\omega) & \omega \in \Omega\backslash N \\ 0 & \omega \in N \end{cases}.$$

Then $\widehat{f}_n \to \widehat{f}$ pointwise and for all $n \in \mathbb{N}$, $|\widehat{f}_n| \leq \widehat{g}$, and $\widehat{g}$ is integrable, so we can apply the first part of the proof to conclude that

$$\lim_{n\to\infty}\int_\Omega f_n \,\mathsf{d}\mu = \lim_{n\to\infty}\int_\Omega \widehat{f}_n \,\mathsf{d}\mu = \int_\Omega \widehat{f} \,\mathsf{d}\mu = \int_\Omega f \,\mathsf{d}\mu. \qquad \square$$

REMARK 7.5.2. The uniform boundedness assumption in the dominated convergence theorem is vital. To see this, we consider $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \lambda)$ and sequence $f_n = n\,\mathbf{1}_{[0,1/n]}$ such that

$$f_n \to f = +\infty\mathbf{1}_{\{0\}} = 0 \qquad \lambda\text{-almost everywhere.}$$

However,

$$\int_\mathbb{R} f_n \,\mathsf{d}\lambda = n\lambda([0,1/n]) = 1 \longrightarrow 1 \neq 0 = \int_\mathbb{R} +\infty\mathbf{1}_{\{0\}} \,\mathsf{d}\lambda.$$

EXAMPLE 7.5.3. Let us determine the value $\displaystyle\lim_{n\to\infty}\int_\mathbb{R} \frac{n\sin(x/n)}{x(x^2+1)}\,\lambda(\mathsf{d}x)$.

To do so, we set $f_n(x) := \dfrac{n\sin(x/n)}{x(x^2+1)}$ and $g(x) := \dfrac{1}{x^2+1}$. Since

$$|\sin(y)| \leq |y| \quad \text{for all } |y| > 0 \qquad \text{and} \qquad \lim_{y\to 0}\frac{\sin(y)}{y} = 1,$$

we find that $|f_n(x)| \leq g$ for all $n \geq 1$, and $f_n(x) \to g(x)$ for all $x \in \mathbb{R}\setminus\{0\}$.

Assuming that g is λ-integrable for the moment, we then obtain from DCT that

$$\lim_{n\to\infty}\int_\mathbb{R} f_n \,\mathsf{d}\lambda = \int_\mathbb{R} g \,\mathsf{d}\lambda.$$

The fact that g is λ-integrable can be shown by means of a limit theorem and Riemann integration, outlined in the next chapter.

CHAPTER 8

From Riemann to Lebesgue integration

8.1. Recalling Riemann integration

DEFINITION 8.1.1. A partition $P = (x_0, \dots, x_n)$ of $[a, b]$ is an $(n+1)$-tuple of real numbers x_i such that $a = x_0 < x_1 < \dots < x_n = b$, and we denote by $\Delta x_i = x_i - x_{i-1}$ the length of the interval $[x_{i-1}, x_i]$, $i = 1, \dots, n$. Furthermore, we say that a partition $Q = (y_0, y_1, \dots y_m)$ of $[a, b]$ is a refinement of P if $\{x_0, \dots, x_n\} \subset \{y_0, \dots, y_m\}$.

Recall that given a partition $P = (x_0, \dots, x_n)$, the upper and lower sum of a bounded function $f : [a, b] \to \mathbb{R}$ with respect to P are defined as

$$U(P, f) := \sum_{i=1}^{n} \sup\big\{f(x) : \ x \in [x_{i-1}, x_i]\big\}\Delta x_i$$

and

$$L(P, f) := \sum_{i=1}^{n} \inf\big\{f(x) : \ x \in [x_{i-1}, x_i]\big\}\Delta x_i$$

Note that if a partition Q is a refinement of a partition P, then

$$L(Q, f) \geq L(P, f) \quad \text{ and } U(Q, f) \leq U(P, f).$$

Finally, if P and R are two partitions of $[a, b]$, there exists a partition Q of $[a, b]$ such that Q is both a refinement of P and a refinement of R.

The upper and lower Riemann integral of f are respectively defined as

$$\overline{\int_a^b} f(x)\, \mathsf{d}x := \inf\big\{U(P, f) : \ P \text{ partition of } [a, b]\big\}$$

and

$$\underline{\int_a^b} f(x)\, \mathsf{d}x := \sup\big\{L(P, f) : \ P \text{ partition of } [a, b]\big\}.$$

DEFINITION 8.1.2. Recall that a bounded function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if

$$\overline{\int_a^b} f(x)\, \mathsf{d}x = \underline{\int_a^b} f(x)\, \mathsf{d}x.$$

If f is Riemann integrable, the Riemann integral of f is defined as

$$\int_a^b f(x)\, \mathsf{d}x := \sup\big\{U(P, f) : \ P \text{ partition of } [a, b]\big\}.$$

8.2. Riemann vs Lebesgue integration

THEOREM 8.2.1. *If a bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable, then f is* Lebesgue-*measurable and integrable. Moreover*

$$\int_a^b f(x)\,\mathsf{d}x = \int_{[a,b]} f\,\mathsf{d}\lambda.$$

PROOF. Let $f : [a, b] \to \mathbb{R}$ be Riemann-integrable. We can then find a sequence of partitions (P_n), $P_n = (x_1^n, \dots, x_{N_n}^n)$, such that for every $n \in \mathbb{N}$, P_{n+1} is a refinement of P_n and such that

$$\lim_{n\to\infty} U(P_n, f) = \lim_{n\to\infty} L(P_n, f) = \int_a^b f(x)\,\mathsf{d}x. \tag{8.2.1}$$

The details on how to find such P_n are as follows: By the definition of the upper and lower Riemann integral and by the assumption that f is bounded and Riemann integrable, we know that there exist partitions Q_1 and R_1 such that

$$\int_a^b f(x)\,\mathsf{d}x - 1 < L(R_1, f) \quad \text{and} \quad U(Q_1, f) < \int_a^b f(x)\,\mathsf{d}x + 1.$$

We then choose the partition P_1 as a common refinement of the partitions Q_1 and R_1. Hence,

$$\begin{aligned}\int_a^b f(x)\,\mathsf{d}x - 1 &< L(R_1, f) \le L(P_1, f)\\ &\le U(P_1, f) \le U(Q_1, f) < \int_a^b f(x)\,\mathsf{d}x + 1.\end{aligned}$$

Now suppose the partition P_k has been defined for some $k \in \mathbb{N}$. Again by the definition of the upper and lower Riemann integral and by the assumption that f is bounded and Riemann integrable, there exist partitions Q_{k+1} and R_{k+1} such that

$$\int_a^b f(x)\,\mathsf{d}x - \frac{1}{k+1} < L(R_{k+1}, f) \qquad \text{and} \qquad U(Q_{k+1}, f) < \int_a^b f(x)\,\mathsf{d}x + \frac{1}{k+1}.$$

Now define P_{k+1} as a common refinement of P_k, Q_{k+1} and R_{k+1}. Then

$$\begin{aligned}\int_a^b f(x)\,\mathsf{d}x - \frac{1}{k+1} &< L(R_{k+1}, f) \le L(P_{k+1}, f)\\ &\le U(P_{k+1}, f) \le U(Q_{k+1}, f) < \int_a^b f(x)\,\mathsf{d}x + \frac{1}{k+1}.\end{aligned}$$

It follows that for every $n \in \mathbb{N}$, the partition P_{n+1} is a refinement of the partition P_n and

$$L(P_n, f) \le \int_a^b f(x)\,\mathsf{d}x < L(P_n, f) + \frac{1}{n}$$
$$U(P_n, f) - \frac{1}{n} < \int_a^b f(x)\,\mathsf{d}x \le U(P_n, f)$$

from which the limits (8.2.1) follow.

Now define the functions

$$u_n := \sum_{i=1}^{N_n} \sup\{f(x) :\ x \in [x_{i-1}^n, x_i^n]\}\mathbf{1}_{(x_{i-1}^n, x_i^n]},$$
$$\ell_n := \sum_{i=1}^{N_n} \inf\{f(x) :\ x \in [x_{i-1}^n, x_i^n]\}\mathbf{1}_{(x_{i-1}^n, x_i^n]}.$$

Because P_{n+1} is a refinement of P_n, we find that $\ell_n \le \ell_{n+1}$, $u_{n+1} \le u_n$, and therefore

$$\ell(x) := \lim_{n\to\infty} \ell_n(x) \quad \text{exists}, \qquad u(x) := \lim_{n\to\infty} u_n(x) \quad \text{exists}.$$

The functions $\ell, u : [a,b] \to \mathbb{R}$ are clearly Borel-measurable. Moreover, $\ell \le f \le u$. Note that

$$U(P_n, f) = \int_{[a,b]} u_n \,\mathsf{d}\lambda \qquad L(P_n, f) = \int_{[a,b]} \ell_n \,\mathsf{d}\lambda.$$

By the dominated convergence theorem (recall that f is bounded),

$$U(P_n, f) \to \int_{[a,b]} u \,\mathsf{d}\lambda \qquad L(P_n, f) \to \int_{[a,b]} \ell \,\mathsf{d}\lambda \qquad \text{as } n \to \infty.$$

However, since f is Riemann integrable, both the upper and lower sums also converge to the Riemann integral of f, so

$$\int_{[a,b]} u \,\mathsf{d}\lambda = \int_a^b f(x)\,\mathsf{d}x = \int_{[a,b]} \ell \,\mathsf{d}\lambda.$$

Now since $\ell \le f \le u$, we then obtain, by linearity of the integral,

$$0 \le \int_{[a',b']} \big(u - \ell\big)\,\mathsf{d}\lambda = 0 \qquad \Longrightarrow \qquad \ell \equiv u \quad \lambda\text{-almost everywhere},$$

from which we also obtain $\ell \equiv f \equiv \ell$ λ-almost everywhere. Moreover, since u and ℓ are both Borel-measurable, f is Borel-measurable, and particularly Lebesgue-measurable. ☐

EXAMPLE 8.2.2. Continuing from Example 7.5.3, let $g_n \coloneqq g\mathbf{1}_{[-n,n]}$. Then clearly, $g_n \to g$ pointwise monotonically. By MCT, we have that

$$\lim_{n\to\infty} \int_{\mathbb{R}} g_n \,\mathsf{d}\lambda = \int_{\mathbb{R}} g \,\mathsf{d}\lambda.$$

On the other hand, for every $n \ge 1$,

$$\int_{\mathbb{R}} g_n \,\mathsf{d}\lambda = \int_{[-n,n]} g \,\mathsf{d}\lambda = \int_{-n}^{n} g \,\mathsf{d}x = \int_{-n}^{n} \frac{1}{1+x^2}\,\mathsf{d}x = \arctan(n) - \arctan(-n).$$

Hence,

$$\int_{\mathbb{R}} g \,\mathsf{d}\lambda = \lim_{n\to\infty} \int_{\mathbb{R}} g_n \,\mathsf{d}\lambda = \frac{\pi}{2} + \frac{\pi}{2} = \pi,$$

thus implying that g is λ-integrable.

The following theorem provides a full characterization of Riemann-integrable functions.

THEOREM 8.2.3. *A bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if it is continuous λ-almost everywhere.*

CHAPTER 9

Littlewood's principles

9.1. Regularity of measures

Definition 9.1.1. We say that a measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ is *outer regular* if for all $A \in \mathcal{B}_{\mathbb{R}^d}$

$$\mu(A) = \inf\big\{\mu(O) : A \subset O \text{ open}\big\}.$$

We say that μ is *locally finite* if for every $R > 0$, $\mu(B(0,R)) < \infty$.

A measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ is *inner regular* if for all $A \in \mathcal{B}_{\mathbb{R}^d}$

$$\mu(A) = \sup\big\{\mu(K) : K \subset A \text{ compact}\big\}.$$

A measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ is called a *Radon measure* if it is locally finite and inner regular.

Surprisingly, locally finite measures on $(\mathbb{R}^d, \mathcal{B})_{\mathbb{R}^d}$ are always inner regular, and they are therefore always Radon measures.

Theorem 9.1.2. *Every finite measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ is both inner and outer regular.*

9.2. Littlewood's first principle

Every measurable set is "practically open"

Theorem 9.2.1. *Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Let $A \in \mathcal{B}_{\mathbb{R}^d}$ and let $\epsilon > 0$. Then there exists a set O of a finite union of open rectangles in $\mathbb{R}^d$, such that the measure of the symmetric difference $A\Delta O := (A\backslash O) \cup (O\backslash A)$ is smaller than ϵ, that is*

$$\mu(A\Delta O) = \mu(A\backslash O) + \mu(O\backslash A) < \epsilon.$$

Proof. We make use of Theorem 9.1.2 for the proof. Let $\epsilon > 0$ be arbitrary. Then the inner regularity of μ provides a compact set $K \subset A$ such that

$$\mu(K) > \mu(A) - \epsilon/2.$$

Moreover, the outer regularity of μ provides a family of open rectangles $(O_i)_{i\in\mathbb{N}}$ such that

$$K \subset \bigcup_{i=1}^{\infty} O_i \qquad \text{and} \qquad \sum_{i=1}^{\infty} \mu(O_i) \leq \mu(K) + \epsilon/2.$$

However, K is compact and therefore there exists an $N \in \mathbb{N}$ such that

$$K \subset \bigcup_{i=1}^{N} O_i =: O.$$

Clearly,

$$\sum_{i=1}^{N} \mu(O_i) \leq \sum_{i=1}^{\infty} \mu(O_i) \leq \mu(K) + \epsilon/2,$$

and hence

$$\mu(A\Delta O) = \mu(A\backslash O) + \mu(O\backslash A) \leq \mu(A\backslash K) + \mu(O\backslash K) \leq \epsilon/2 + \epsilon/2 = \epsilon. \qquad \square$$

9.3. Littlewood's second principle: Egorov's Theorem

Pointwise almost everywhere convergence is "practically uniform convergence".

THEOREM 9.3.1 (Egorov's Theorem I). *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions, that converges pointwise λ-almost everywhere to a (Lebesgue-measurable) function f. Then for all $\epsilon > 0$ there exists a set $C \subset \Omega$ such that $\mu(\Omega \setminus C) \leq \epsilon$ and $f_n \to f$ uniformly on C.*

PROOF. First, for $k, n \in \mathbb{N}$ we investigate for which points $x \in \Omega$, the value $f_m(x)$ is $(1/k)$-close to $f(x)$ for every $m \geq n$. More precisely, we introduce the sets

$$E_{k,n} := \left\{ x \in \Omega : \ |f_m(x) - f(x)| < \frac{1}{k} \text{ for all } m \geq n \right\}.$$

Because $f_n \to f$ pointwise λ-almost everywhere, there exists a set $N \subset \Omega$ with $\mathbb{P}(N) = 0$ such that for every $k \in \mathbb{N}$ and for every $x \in \Omega\backslash N$, for $n \gg 1$ large enough (depending on x),

$$x \in E_{k,n}.$$

We may also write this as

$$\Omega\backslash N = \bigcup_{n=1}^{\infty} E_{k,n} \qquad \text{for every } k \in \mathbb{N}.$$

Now choose n_k such that $\mathbb{P}(\Omega\backslash E_{k,n_k}) < \epsilon\, 2^{-k}$. Finally, define

$$C = \bigcap_{k=1}^{\infty} E_{k,n_k}.$$

Then, by the subadditivity of μ, we obtain

$$\mu(\Omega\backslash C) = \mu\left(\bigcup_{k\in\mathbb{N}} (\Omega\backslash E_{k,n_k})\right) \leq \sum_{k\in\mathbb{N}} \mathbb{P}(\Omega\backslash E_{k,n_k}) = \epsilon.$$

Moreover for every $k \in \mathbb{N}$ and every $x \in C$,

$$|f_m(x) - f(x)| \leq \frac{1}{k} \qquad \text{for all } m \geq n_k.$$

It follows that $f_n \to f$ uniformly on C. $\square$

THEOREM 9.3.2 (Egorov's Theorem II). *Let X_n be a sequence of random variables such that $X_n \to X$ almost surely. For every $\epsilon > 0$ there exists a decreasing function $\eta_\epsilon : \mathbb{N} \to (0, \infty)$ converging to zero as $n \to \infty$, and an event $C \in \mathcal{F}$ with $\mathbb{P}(C) > 1 - \epsilon$, such that*

$$\mathbf{1}_C |X_n - X| \leq \eta_\epsilon(n).$$

9.4. Littlewood's third principle: Lusin's Theorem

Every Borel measurable function is "practically continuous".

THEOREM 9.4.1 (Lusin's Theorem). *Let μ be a finite measure on the measurable space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable. Then for every $\epsilon > 0$ there exists a compact set $K \subset \mathbb{R}^d$ and a continuous function $g : \mathbb{R}^d \to \mathbb{R}$ such that $\mu(\mathbb{R}^d \backslash K) < \epsilon$ and $f \equiv g$ on K.*

PROOF. Let $\epsilon > 0$. Define for $n \in \mathbb{N}$ and $k \in \mathbb{Z}$ the set

$$A_k^n \coloneqq \Big\{\omega \in \mathbb{R}^d : \ (k-1)2^{-n} < f(\omega) \le k2^{-n}\Big\}.$$

Now there exist open sets $U_k^n \supset A_k^n$ and compact sets $K_k^n \subset A_k^n$ such that

$$\mu(U_k^n \backslash A_k^n) < \frac{1}{n2^{|k|}} \qquad \mu(A_k^n \backslash K_k^n) < \frac{1}{n2^{|k|}}.$$

Define continuous functions $\varphi_k^n : \mathbb{R}^d \to \mathbb{R}$ such that φ_k^n is compactly supported in U_k^n, moreover $0 \le \varphi_k^n \le 1$ and $\varphi_k^n(\omega) = 1$ for $\omega \in K_k^n$. Define

$$\varphi^n \coloneqq \sum_{k=-2^n}^{2^n} k2^{-n}\varphi_k^n.$$

which is continuous for all $n \ge 1$. Since the functions $\varphi^n \to f$ pointwise a.e., by Egorov's Theorem, there is a compact set K such that φ^n converge to f uniformly on K. Because uniform convergence preserves continuity, f restricted to K, denoted by $f|_K$, is continuous.

We now need to construct $Y : \mathbb{R}^d \to \mathbb{R}$. For $\omega \in K$, we set $g(\omega) = f(\omega)$. Because $f|_K$ is continuous, and K is compact, $g|_K$ is uniformly continuous. Therefore, there exists a continuous, increasing function $\eta : [0, \infty) \to [0, \infty)$ such that

$$|g(x) - g(y)| \le \eta(d(x, y)) \qquad x, y \in K.$$

Finally, we set

$$g(x) \coloneqq \sup_{y \in K}\big\{f(y) - \eta(d(x, y))\big\}.$$

Note that g is continuous and, agrees with f on K. $\square$

CHAPTER 10

L^p-spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $p \in [1, +\infty)$. Throughout this chapter, we denote

$$\|f\|_p := \left(\int_\Omega |f|^p \, \mathsf{d}\mu \right)^{1/p} \qquad \text{for any measurable function } f : \Omega \to \mathbb{R}.$$

For $p = +\infty$, we set

$$\|f\|_\infty := \operatorname{ess\,sup}\{|f(\omega)| : \ \omega \in \Omega\} = \inf\{t \in [0, \infty) : \ \mu(\{|f| > t\}) = 0\}.$$

10.1. The Hölder inequality

PROPOSITION 10.1.1 (Hölder's inequality). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $p, q \in [1, +\infty]$ be conjugate exponents, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \qquad \textit{for all measurable functions } f, g : \Omega \to \mathbb{R}.$$

PROOF. If the right-hand side is $+\infty$, there is nothing to prove.

Now we will see a very important trick in proving inequalities like this. We note that it is enough to show the inequality for the case in which

$$\int_\Omega |f|^p \, \mathsf{d}\mu = \int_\Omega |g|^q \, \mathsf{d}\mu = 1.$$

By Young's inequality, i.e. for conjugate exponents $p, q \in (1, +\infty)$,

$$ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q \qquad \text{for any } a, b \in [0, +\infty),$$

we have for every $x \in \Omega$, that

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

Hence

$$\int_\Omega |fg| \, \mathsf{d}\mu \leq \frac{1}{p} \int_\Omega |f|^p \, \mathsf{d}\mu + \frac{1}{q} \int_\Omega |g|^q \, \mathsf{d}\mu = 1.$$

For the case $p = 1$, $q = +\infty$, we easily get

$$\int_\Omega |fg| \, \mathsf{d}\mu \leq \int_\Omega |f| \|g\|_\infty \, \mathsf{d}\mu = \|f\|_1 \|g\|_\infty. \qquad \square$$

10.2. The Minkowski inequality

PROPOSITION 10.2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $p \in [1, +\infty]$ be conjugate exponents. Then the 'triangle inequality' holds:*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \qquad \text{for all measurable functions } f, g : \Omega \to \mathbb{R}.$$

PROOF. As before, if the right-hand side is $+\infty$, then there is nothing to prove. Now suppose that $\|f\|_p, \|g\|_p < +\infty$. Then from the binomial formula for $p \in [1, +\infty)$

$$|a + b|^p \leq 2^{p-1}\big(|a|^p + |b|^p\big),$$

we have that

$$\int_\Omega |f + g|^p \, \mathsf{d}\mu \leq 2^{p-1} \left(\int_\Omega |f|^p \, \mathsf{d}\mu + \int_\Omega |g|^p \, \mathsf{d}\mu \right),$$

and hence $\|f + g\|_p < +\infty$. Next,

$$\begin{aligned}
\|f + g\|_p^p &= \int_\Omega |f + g|^p \, \mathsf{d}\mu \\
&= \int_\Omega |f + g||f + g|^{p-1} \, \mathsf{d}\mu \\
&\leq \int_\Omega (|f| + |g|)|f + g|^{p-1} \, \mathsf{d}\mu \\
&= \int_\Omega |f||f + g|^{p-1} \, \mathsf{d}\mu + \int_\Omega |g||f + g|^{p-1} \, \mathsf{d}\mu.
\end{aligned}$$

Now we apply Hölder's inequality (with exponents p and $p/(p-1)$) on both terms to obtain

$$\|f + g\|_{L^p}^p \leq \left(\int_\Omega |f|^p \mathsf{d}\mu \right)^{1/p} \|f + g\|_p^{p-1} + \left(\int_\Omega |g|^p \mathsf{d}\mu \right)^{1/p} \|f + g\|_p^{p-1}.$$

Finally, we divide both sides by $\|f + g\|_p^{p-1}$ and find

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

As for the case $p = +\infty$, we use the triangle inequality to obtain $|f + g| \leq |f| + |g|$, and hence,

$$|f(\omega) + g(\omega)| \leq \|f\|_\infty + \|g\|_\infty \qquad \text{for } \mu\text{-almost every } \omega \in \Omega.$$

Taking the essential supremum then yields the required inequality. □

10.3. Normed and semi-normed vector spaces

Recall that a norm $\|\cdot\|$ on a vector space V is a function $V \to [0, \infty)$ such that

(i) $\|v\| = 0 \Leftrightarrow v = 0$ for all $v \in V$

(ii) $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$

(iii) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

If only the last two properties hold, $\|.\|$ is instead called a *seminorm.*

Let $(V, \|\cdot\|)$ be a semi-normed space. We say that a sequence $(v_n)_{n\in\mathbb{N}} \subset V$ is a Cauchy sequence if for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $m, n \geq N$,

$$\|v_m - v_n\| < \epsilon.$$

We say that a semi-normed space is *complete*, if and only if every Cauchy sequence converges, that is, for every Cauchy sequence $(v_n)_{n\in\mathbb{N}} \subset V$ there exists a $v \in V$ such that

$$\lim_{n\to\infty} \|v_n - v\| = 0.$$

To every semi-normed space $(V, \|\cdot\|)$ one can associate a normed linear space in a standard way. One defines the equivalence relation $\sim$ by $v \sim w$ if and only if $\|v - w\| = 0$. Denote by W the linear space of equivalence classes. One defines a norm on equivalence classes $[v]$ and $[w]$ in W by $\|[w] - [v]\| = \|w - v\|$. If $(V, \|\cdot\|)$ is a complete semi-normed space, then W is a *Banach space*, which is a complete normed linear space.

10.4. Semi-normed $\mathbb{L}^p$ spaces

We have seen in Section 6.7 that the set of μ-integrable functions form a vector space (over $\mathbb{R}$). For $p \in [0, +\infty)$, we define the vector space $\mathbb{L}^p$ of integrable functions f for which

$$\|f\|_p < +\infty.$$

By the Minkowski inequality, $\|\cdot\|_p$ is a seminorm on $\mathbb{L}^p$ for every $p \in [1, \infty]$.

Clearly, the seminorm $\|.\|_p$ is not a norm on $\mathbb{L}^p$: indeed $\|f - g\|_p = 0$ if and only if $f(\omega) = g(\omega)$, for μ-almost every $\omega \in \Omega$. We follow the standard construction described in Section 10.3 to create an associated normed linear space. We say that $f \sim g$ if and only if f is equal to g μ-almost everywhere. We denote by L^p the vector space of equivalence classes

$$L^p := \mathbb{L}^p / \sim .$$

10.5. Completeness of L^p-spaces

Theorem 10.5.1 (Completeness of L^p spaces). *The normed linear space L^p is complete, and is thus a Banach space, for every $p \in [1, +\infty]$.*

Proof. First let $p \in [1, \infty)$ and let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. The trick is to select a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that

$$\|f_{n_{k+1}} - f_{n_k}\|_{L^p(\Omega)} < 4^{-k-1}.$$

For ease of notation we set $g_k := f_{n_k}$. Note that by a telescoping argument, for all $\ell \geq k$,

$$\|g_\ell - g_k\|_{L^p(\Omega)} < 4^{-k}.$$

Then

$$\mu\left(\left\{x \in \Omega : \ |g_{k+1}(x) - g_k(x)| > 2^{-k}\right\}\right) < \frac{1}{2^{-kp}} \|g_{k+1} - g_k\|^p_{L^p(\Omega)} < 2^{-kp}.$$

In particular, by the Borel-Cantelli Lemma, for μ-a.e. $x \in \Omega$, there is an $N \in \mathbb{N}$ such that

$$|g_{k+1}(x) - g_k(x)| \leq 2^{-k} \qquad \text{for all } k > N.$$

For such x, the sequence $(g_k(x))_{k\in\mathbb{N}}$ is Cauchy. So by the completeness of $\mathbb{R}$, a limit exists, which we call $f(x)$.

By Fatou's Lemma,

$$\|g_k - f\|_p \leq \liminf_{\ell\to\infty} \|g_k - g_\ell\|_p \leq 4^{-k}.$$

To see that this implies that f_n converges to f in L^p, we take an arbitrary $\epsilon > 0$. Since f_n is a Cauchy sequence, there exists an $N_1 \in \mathbb{N}$ such that for all $m, n \geq N_1$,

$$\|f_n - f_m\|_p < \frac{\epsilon}{2}.$$

Now there exists an $K \in \mathbb{N}$, with $K > N_1$ (and therefore $n_K > N_1$) such that for all $k \geq K$,

$$\|f_{n_k} - f\|_p < \frac{\epsilon}{2}.$$

Set $N_2 := \max(N_1, n_K)$. Then, for $n \geq N_2$, we find

$$\|f_n - f\|_p \leq \|f_n - f_{n_K}\|_p + \|f_{n_K} - f\|_{L^p(\Omega)} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which gives the required convergence.

The proof of completeness of $L^\infty(\Omega)$ follows similar lines but is in a way easier. Let again $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence and select a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that

$$\|f_{n_{k+1}} - f_{n_k}\|_{L^\infty(\Omega)} < 4^{-k-1}.$$

We define again $g_k = f_{n_k}$. Then

$$\mu\left(\left\{x \in \Omega : \ |g_{k+1}(x) - g_k(x)| \geq 4^{-k-1}\right\}\right) = 0.$$

So, $(g_k(x))_{k\in\mathbb{N}}$ is a Cauchy-sequence for almost every $x \in \Omega$. For such x, the limit as $k \to \infty$ of $g_k(x)$ exists, and we denote it by $f(x)$. Moreover,

$$\mu\left(\left\{x \in \Omega : |g_k(x) - f(x)| \geq 4^{-k}\right\}\right) = 0.$$

It follows that g_k converges to f in $L^\infty(\Omega)$ as $k \to \infty$, and therefore that f_n converges to f in $L^\infty(\Omega)$ as $n \to \infty$ using the same argument as above. $\square$

CHAPTER 11

The Radon-Nikodym Theorem

11.1. A General Radon-Nikodym Theorem

Lemma 11.1.1. *Let μ, ν be finite measures on $(\Omega, \mathcal{F})$ satisfying $\nu \le \mu$ on $\mathcal{F}$, i.e. $\nu(A) \le \mu(A)$ for every $A \in \mathcal{F}$. Then there exists an $(\mathcal{F}, \mathcal{B}_{\mathbb{R}})$-measurable function f_0 with $0 \le f_0 \le 1$ such that*

$$\nu(E) = \int_E f_0 \, \mathsf{d}\mu \qquad \text{for all } E \in \mathcal{F}.$$

Proof. Let

$$H := \Big\{ f \text{ measurable} \,:\, 0 \le f \le 1, \ \int_E f \, \mathsf{d}\mu \le \nu(E) \ \text{ for all } E \in \mathcal{F} \Big\}.$$

Note that $H \neq \emptyset$ since 0 belongs to H. Moreover, when f_1, $f_2 \in H$, also $\max\{f_1, f_2\} \in H$. Indeed, if $A = \{x \in \Omega \,:\, f_1(x) \ge f_2(x)\}$, then

$$\begin{aligned}\int_E \max\{f_1, f_2\} \, \mathsf{d}\mu &= \int_{E\cap A} \max\{f_1, f_2\} \, \mathsf{d}\mu + \int_{E \cap A^c} \max\{f_1, f_2\} \, \mathsf{d}\mu \\ &= \int_{E\cap A} f_1 \, \mathsf{d}\mu + \int_{E\cap A^c} f_2 \, \mathsf{d}\mu \le \nu(E \cap A) + \nu(E \cap A^c) = \nu(E).\end{aligned}$$

Now let $M = \sup\{\int_\Omega f \, \mathsf{d}\mu \,:\, f \in H\}$. Then, $0 \le M < +\infty$ and we find from the previous argument a sequence of measurable functions $(f_n)_{n\in\mathbb{N}} \subset H$ with $0 \le f_1 \le \cdots \le 1$ such that

$$\int_\Omega f_n \, \mathsf{d}\mu > M - \frac{1}{n}.$$

Define $f_0 := \lim_{n\to\infty} f_n$. Then f_0 is measurable. By the Monotone Convergence Theorem, $f_0 \in H$ and $\int_\Omega f_0 \, \mathsf{d}\mu \ge M$. Hence, $\int_\Omega f_0 \, \mathsf{d}\mu = M$.

To complete the proof, we show that $\nu(E) = \int_E f_0 \, \mathsf{d}\mu$ for all $E \in \mathcal{F}$. Suppose otherwise, i.e. there is a set $E \in \mathcal{F}$ for which $\nu(E) > \int_E f_0 \, \mathsf{d}\mu$. Then we can write $E = E_0 \cup E_1$, where $E_1 := \{x \in \Omega \,:\, f_0(x) = 1\}$ and $E_0 := E \setminus E_1$. Since

$$\nu(E) = \nu(E_0) + \nu(E_1) > \int_E f_0 \, \mathsf{d}\mu = \int_{E_0} f_0 \, \mathsf{d}\mu + \mu(E_1) \ge \int_{E_0} f_0 \, \mathsf{d}\mu + \nu(E_1),$$

it follows that $\nu(E_0) > \int_{E_0} f_0 \, \mathsf{d}\mu$. Let $F_n := \{f_0 < 1 - n^{-1}\} \cap E_0$, which gives a sequence of increasing measurable sets. Due to the continuity from below of ν, we obtain

$$\lim_{n\to\infty} \nu(F_n) = \nu\left(\bigcup_{n\ge 1} F_n \right) = \nu(E_0) > \int_{E_0} f_0 \, \mathsf{d}\mu.$$

In particular, there exists some n_0 such that

$$\begin{aligned}\nu(F_{n_0}) > \int_{E_0} f_0 \,\mathsf{d}\mu &= \int_{F_{n_0}} f_0 \,\mathsf{d}\mu + \int_\Omega f_0 \,\mathbf{1}_{\{f_0 \geq 1-n_0^{-1}\}\cap E_0} \,\mathsf{d}\mu \\ &\geq \int_{F_{n_0}} f_0 \,\mathsf{d}\mu + \big(1+n_0^{-1}\big)\mu(\{f_0 \geq 1-n_0^{-1}\}\cap E_0) \\ &= \int_{F_{n_0}} f_0 + \varepsilon \mathbf{1}_{F_{n_0}} \,\mathsf{d}\mu,\end{aligned}$$

with $\varepsilon \coloneqq \big(1+n_0^{-1}\big)\mu(\{f_0 \geq 1-n_0^{-1}\}\cap E_0)/\mu(F_{n_0}) > 0$. Based on this fact, we claim the existence of a measurable set $F \subset F_{n_0}$ with $\mu(F) > 0$ such that $f_0 + \varepsilon \mathbf{1}_F \in H$. If not, then every measurable set $F \subset F_{n_0}$ with $\mu(F) > 0$ contains a measurable subset $G \subset F$ with $\int_G f_0 + \varepsilon \mathbf{1}_F \,\mathsf{d}\mu > \nu(G)$. By an exhaustion argument, we can find a disjoint partition $\bigcup_{m\geq 1} G_m = F_{n_0}$ of F_{n_0} such that $\int_{G_m} f_0 + \varepsilon \mathbf{1}_F \,\mathsf{d}\mu > \nu(G_m)$ for all $m \geq 1$. Consequently,

$$\nu(F_{n_0}) > \int_{F_{n_0}} f_0 + \varepsilon \mathbf{1}_F \,\mathsf{d}\mu = \sum_{m\geq 1} \int_{G_m} f_0 + \varepsilon \mathbf{1}_F \,d\mu > \sum_{m\geq 1} \nu(G_m) = \nu(F_{n_0}),$$

which is a contradiction, i.e. such a measurable set $F \subset F_{n_0}$ must exist.

However, since $\int_\Omega f_0 + \varepsilon \mathbf{1}_F \,\mathsf{d}\mu = M + \varepsilon\mu(F) > M$, this leads to another contradiction, and hence $\nu(E) = \int_E f_0 \,\mathsf{d}\mu$ for all $E \in \mathcal{F}$ as desired. □

THEOREM 11.1.2 (Lebesgue-Radon-Nikodym). *Let μ, ν be finite measures on $(\Omega, \mathcal{F})$. Then there exists a μ-null set $D \in \mathcal{F}$ and a nonnegative μ-integrable function f_0 such that*

$$\nu(E) = \nu(E \cap D) + \int_E f_0 \,\mathsf{d}\mu \qquad \text{for all } E \in \mathcal{F}.$$

PROOF. Let $\lambda = \mu + \nu$. Then $0 \leq \nu \leq \lambda$, so by Lemma 11.1.1, there exists a measurable function g with $0 \leq g \leq 1$ such that $\nu(E) = \int_E g \,\mathsf{d}\lambda$ for all $E \in \mathcal{F}$. It follows that $\mu(E) = \int_\Omega (1-g)\,\mathsf{d}\lambda$ for all $E \in \mathcal{F}$. Let $D = \{g = 1\}$. Then $\mu(D) = 0$.

Moreover, since $\nu(E) = \int_E g \,\mathsf{d}\nu + \int_E g \,\mathsf{d}\mu$, we have $\int_E (1-g)\,\mathsf{d}\nu = \int_E g \,\mathsf{d}\mu$ for all $E \in \mathcal{F}$. In particular, $\int_\Omega (1-g) f \,\mathsf{d}\nu = \int_\Omega g f \,\mathsf{d}\mu$ for all nonnegative measurable functions f. Taking $f = (1 + g + \cdots g^n)\mathbf{1}_E$, we learn that

$$\int_E (1-g^{n+1})\,\mathsf{d}\nu = \int_E g(1 + g \cdots + g^n)\,\mathsf{d}\mu \qquad \text{for all } E \in \mathcal{F} \text{ and } n \geq 1.$$

Now since $0 \leq g < 1$ on D^c, the Monotone Convergence Theorem yields

$$\begin{aligned}\nu(E \cap D^c) = \lim_{n\to\infty} \int_{E\cap D^c} (1-g^{n+1})\,\mathsf{d}\nu &= \lim_{n\to\infty} \int_{E\cap D^c} g(1 + g \cdots + g^n)\,\mathsf{d}\mu \\ = \int_{E\cap D^c} g(1-g)^{-1}\,\mathsf{d}\mu &= \int_E f_0 \,d\mu,\end{aligned}$$

where $f_0 \coloneqq g(1-g)^{-1}\mathbf{1}_{D^c}$, thus concluding the proof. □

11.2. Absolute Continuity

Definition 11.2.1. Let μ, ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say that ν is *absolutely continuous* with respect to μ, if for every $E \in \mathcal{F}$ with $\mu(E) = 0$, we also have that $\nu(E) = 0$, i.e. μ-null sets are ν-null sets. In this case, we write $\nu \ll \mu$.

Theorem 11.2.2 (Lebesgue-Radon-Nikodym II). *Let μ, ν be finite measures on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$. Then there exists a nonnegative μ-integrable function $\frac{\mathsf{d}\nu}{\mathsf{d}\mu}$, called the Radon-Nikodym derivative of ν w.r.t. μ such that*

$$\nu(E) = \int_E \frac{\mathsf{d}\nu}{\mathsf{d}\mu} \, \mathsf{d}\mu \qquad \text{for all } E \in \mathcal{F}.$$

The function $\frac{\mathsf{d}\nu}{\mathsf{d}\mu}$ is also often called the μ-density of ν.

Proof. We apply Theorem 11.1.2 to obtain a μ-measurable function f_0 such that

$$\nu(E) = \nu(E \cap D) + \int_E f_0 \, \mathsf{d}\mu \qquad \text{for all } E \in \mathcal{F},$$

where D is a μ-null set. In particular, $E \cap D$ is a μ-null set. Since $\nu \ll \mu$, we have also that $\nu(E \cap D) = 0$. Setting $\frac{\mathsf{d}\nu}{\mathsf{d}\mu} := f_0$, we then obtain the assertion. $\square$

Theorem 11.2.3 (Lebesgue-Radon-Nikodym III). *Let $\mathbb{P}$, $\mathbb{A}$ be probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{A} \ll \mathbb{P}$. Then there exists a nonnegative $\mathbb{P}$-integrable function $\frac{\mathsf{d}\mathbb{A}}{\mathsf{d}\mathbb{P}}$, called the Radon-Nikodym derivative of $\mathbb{A}$ w.r.t. $\mathbb{P}$ such that*

$$\mathbb{A}(E) = \int_E \frac{\mathsf{d}\mathbb{A}}{\mathsf{d}\mathbb{P}} \, \mathsf{d}\mathbb{P} \qquad \text{for all } E \in \mathcal{F}.$$

11.3. Conditional Expectation

As a first application of the Radon-Nikodym theorem, we may use it to construct the conditional expectation with respect to a sub-σ-algebra in probability theory.

Theorem 11.3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$. For every $\mathbb{P}$-integrable random variable X, there exists an $\mathcal{H}$-measurable random variable $\mathbb{E}[X|\mathcal{H}]$ such that*

$$\int_B \mathbb{E}[X|\mathcal{H}] \, \mathsf{d}\mathbb{P} = \int_B X \, \mathsf{d}\mathbb{P} \qquad \text{for every } B \in \mathcal{H}.$$

Proof. Define the measure $\mathbb{A}$ on the measurable space $(\Omega, \mathcal{H})$ by

$$\mathbb{A}(B) := \int_B X \, \mathsf{d}\mathbb{P} \qquad \text{for every } B \in \mathcal{H}.$$

The measure $\mathbb{A}$ is absolutely continuous with respect to the restriction of $\mathbb{P}$ to $\mathcal{H}$, which we denote by $\mathbb{P}|_{\mathcal{H}}$. By the Radon-Nikodym theorem, there exists an $\mathcal{H}$-measurable random variable, which we denote by $\mathbb{E}[X|\mathcal{H}]$, such that for all $B \in \mathcal{H}$,

$$\int_B X \, \mathsf{d}\mathbb{P} = \mathbb{A}(B) = \int_B \mathbb{E}[X|\mathcal{H}] \, \mathsf{d}\mathbb{P},$$

thereby concluding the proof. $\square$

Definition 11.3.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let X, Y be random variables, where X is $\mathbb{P}$-integrable. Then the conditional expectation of X given Y is defined as

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)].$$

11.4. Martingales

In this section we will say a few words about martingales. Martingales have very powerful applications inside and outside of probability theory. Unfortunately, we do not have the time to deeply go into the theory of martingales. This section is included since with all the work in the course up to now, we can finally give a *definition* of a martingale and surrounding concepts. Moreover, with the language of measure theory developed, this definition is relatively clear and concise.

Let $\mathcal{I}$ be some (possibly uncountable) index set. A *filtered* probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_\alpha)_{\alpha \in \mathcal{I}}$, where $\mathcal{F}_\alpha$ is a sub-σ-algebra of $\mathcal{F}$ for each $\alpha \in \mathcal{I}$.

A *stochastic process* $X = (X_\alpha)_{\alpha \in \mathcal{I}}$ is a family of random variables $X_\alpha : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, $\alpha \in \mathcal{I}$. We say that a process X is *adapted* to the filtration $\mathbb{F}$ (or simply $\mathbb{F}$-*adapted*) if the random variable X_α is $\mathcal{F}_\alpha$-measurable for every $\alpha \in \mathcal{I}$.

We a say that a stochastic process X is an $(\mathbb{F}, \mathbb{P})$-*martingale* if

(i) X is $\mathbb{F}$-adapted
(ii) $\mathbb{E}(|X_\alpha|) < \infty$ for all $\alpha \in \mathcal{I}$,
(iii) For every $\alpha, \beta \in \mathcal{I}$ with $\beta \leq \alpha$, $\mathbb{E}[X_\alpha|\mathcal{F}_\beta] = X_\beta$ $\mathbb{P}$-almost surely.

A function $\tau : \Omega \to \mathcal{I}$ is called an $\mathbb{F}$-*stopping time* if for every $\alpha \in \mathcal{I}$,

$$\{\omega \in \Omega : \ \tau(\omega) \leq \alpha\} \qquad \text{is } \mathcal{F}_\alpha\text{-measurable},$$

or in other words, the stochastic process $(\mathbf{1}_{\{\tau \leq \alpha\}})_{\alpha \in \mathcal{I}}$ is $\mathbb{F}$-adapted.

CHAPTER 12

Product measures and independent random variables

12.1. Product measures and independent random variables

The following theorem defines the product of two measure spaces.

THEOREM 12.1.1 (Construction of product measure). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, for $i = 1, 2$, be two σ-finite measure spaces. Let $\mathcal{F}_1 \otimes \mathcal{F}_2$ denote the product σ-algebra. Then there exists a unique measure $\mu_1 \otimes \mu_2$ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that*

$$(\mu_1 \otimes \mu_2)(A \times B) = \mu_1(A) \cdot \mu_2(B)$$

for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{G}_i)$ are random variables, then recall that the *law* of X_i is defined as $\mu_i := (X_i)_{\#}\mathbb{P}$. The random variables X_1 and X_2 are *independent* if and only if the law of the random variable $(X_1, X_2) : \Omega \to E_1 \times E_2$ is the product measure $\mu_1 \otimes \mu_2$.

Let us be a bit more precise, and *define* what we mean by independence of σ-algebras, express independence of random variables and events in this language, and then leave it as an exercise to show that these are generalizations of the concepts that we know from elementary probability theory.

DEFINITION 12.1.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{I}$ be some index set and let $\{\mathcal{F}_\alpha\}_{\alpha \in \mathcal{I}}$ be a family of sub-σ-algebras. We say that the family is a family of independent sub-σ-algebras if for every finite subset $J \subset \mathcal{I}$, and sets $A_j \in \mathcal{F}_j$ for $j \in J$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

We can now express independence of random variables in this language.

DEFINITION 12.1.3. Let $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ be a family of random variables, with $X_\alpha : (\Omega, \mathcal{F}) \to (E_\alpha, \mathcal{G}_\alpha)$. We say that the random variables X_α are independent if the family of sub σ-algebras $\{\sigma(X_\alpha)\}_{\alpha \in \mathcal{I}}$, is independent.

The independence of events can also be expressed in the measure-theoretic language.

DEFINITION 12.1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A family of events $\{A_\alpha\}_{\alpha \in \mathcal{I}}$ is called independent if the family $\{\mathcal{F}_\alpha\}_{\alpha \in \mathcal{I}}$ of sub-σ-algebras

$$\mathcal{F}_\alpha := \{\emptyset, A_\alpha, \Omega \backslash A_\alpha, \Omega\}$$

is independent.

The following exercise shows that these concepts are really generalizations of concepts that you have seen in elementary probability theory.

Exercise 12.1.5. Let $X_1, X_2 : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be two random variables. Show that X_1 and X_2 are independent according to the above definition, if and only if

$$\mathbb{P}(X_1 \leq a,\ X_2 \leq b) = \mathbb{P}(X_1 \leq a)\,\mathbb{P}(X_2 \leq b)$$

for every $a \in \mathbb{R}$, $b \in \mathbb{R}$.

12.2. Construction of product measures

Given two measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ we recall the definition of the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ defined on the set $\Omega := \Omega_1 \times \Omega_2$ as the σ-algebra generated by the collection

$$\mathcal{S} = \big\{ A \times B \subset \Omega :\ A \in \mathcal{F}_1,\ B \in \mathcal{F}_2 \big\}.$$

We define the set function $\mu_\circ : \mathcal{A}(\mathcal{S}) \to [0, \infty]$ by

$$\mu_\circ(A \times B) := \mu_1(A) \cdot \mu_2(B) \qquad \text{for } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2.$$

In Section 12.1 we stated in Theorem 12.1.1 that in case the measure spaces are σ-finite, there exists a unique measure $\mu_1 \otimes \mu_2$ on $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that

$$(\mu_1 \otimes \mu_2)(A \times B) = \mu_\circ(A \times B) = \mu_1(A) \cdot \mu_2(B) \qquad \text{for all } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2.$$

This theorem is of course proven by an application of the Carathéodory machinary. In order to apply the Carathéodory extension theorem, we should verify that $\mathcal{S}$ is a semi-algebra of sets and that $\mu_\circ$ is a premeasure. As always, this will take some effort. In view of time, we will state the outcome of the Carathéodory extension theorem.

Theorem 12.2.1 (Construction of product measure). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, for $i = 1, 2$, be two measure spaces. Then the set function $\mu_\circ : \mathcal{A}(\mathcal{S}) \to [0, +\infty]$ given by*

$$\mu_\circ(A \times B) := \mu_1(A) \cdot \mu_2(B) \qquad \textit{for } A \in \mathcal{F}_1 \textit{ and } B \in \mathcal{F}_2 \textit{ is a premeasure.}$$

Let μ^* be the Carathéodory outer measure induced by $\mu_\circ$. Then, the Carathéodory extension of $\mu_\circ$, denoted by $\overline{\mu_1 \otimes \mu_2}$, is a measure defined on the σ-algebra of μ^*-measurable subsets of Ω, which we denote by $\overline{\mathcal{F}_1 \otimes \mathcal{F}_2}$. The σ-algebra $\overline{\mathcal{F}_1 \otimes \mathcal{F}_2}$ is larger than $\mathcal{F}_1 \otimes \mathcal{F}_2$.

If the measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ are σ-finite, the measure $\mu_1 \otimes \mu_2$, which is defined as the restriction of $\overline{\mu_1 \otimes \mu_2}$ to $\mathcal{F}_1 \otimes \mathcal{F}_2$, is the unique measure on $\mathcal{F}_1 \otimes \mathcal{F}_2$ extending $\mu_\circ$.

12.3. Fubini's Theorem and Tonelli's Theorem

In this section, we will present four versions of a very similar statement, which basically states that in order to integrate over a product measure space, one may use iterated integration, and one may also change the order of integration, very similar to what you have seen for Riemann integration.

Example 12.3.1. Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be continuous. Then

$$\int_{[0,1]\times[0,1]} f \,\mathsf{d}A = \int_0^1 \int_0^1 f(x,y) \,\mathsf{d}x \,\mathsf{d}y = \int_0^1 \int_0^1 f(x,y) \,\mathsf{d}y \,\mathsf{d}x.$$

The left-hand side can either be interpreted as the two-dimensional Riemann integral, or the integral against the two-dimensional Lebesgue measure: these integrals agree.

Four versions sounds a bit excessive, but let us try to see what is satisfactory, and what is unsatisfactory about each of the versions, so that we have at least some justification for the amount of variation.

Theorem 12.3.2 (Fubini, version for non-complete measure spaces). *Let* $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ *be two measure spaces. Let* $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ *be integrable on the product space* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$. *Then, for every* $x \in \Omega_1$, *the function* $y \mapsto f(x, y)$ *is* $(\mathcal{F}_2, \mathcal{B}_{\overline{\mathbb{R}}})$*-measurable. Moreover, the function*

$$x \mapsto \int_{\Omega_2} f(x, y)\, \mu_2(\mathsf{d}y)$$

is $(\mathcal{F}_i, \mathcal{B}_{\overline{\mathbb{R}}})$*-measurable and*

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f \, \mathsf{d}\mu_1 \otimes \mu_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\, \mathsf{d}\mu_2(y) \right) \mathsf{d}\mu_1(x) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y)\, \mathsf{d}\mu_1(x) \right) \mathsf{d}\mu_2(y).
\end{aligned}$$

Checking whether you can apply Fubini's theorem usually comes with (at least) two difficulties. First of all, you need to check that the function $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ is measurable with respect to the product σ-algebra. This somehow falls into the category "if f is not constructed by some crazy procedure, it's probably fine". What can at times be more difficult, is to check that f is integrable. This is an important check, because if f is not integrable, the conclusion of the theorem may not hold.

This second difficulty is somehow alleviated by Tonelli's theorem, stated next. Tonelli's theorem is about nonnegative functions, and does not assume integrability. It does need the additional assumption of σ-finiteness of the measure spaces.

Theorem 12.3.3 (Tonelli, version for non-complete measure spaces). *Let* $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ *be two* σ*-finite measure spaces. Let* $f : \Omega_1 \times \Omega_2 \to [0, +\infty]$ *be measurable on the product space* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$. *Then, for every* $x \in \Omega_1$, *the function* $y \mapsto f(x, y)$ *is* $(\mathcal{F}_2, \mathcal{B}_{\overline{\mathbb{R}}})$*-measurable. Moreover, the function*

$$x \mapsto \int_{\Omega_2} f(x, y)\, \mathsf{d}\mu_2(y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$*-measurable and*

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f \, \mathsf{d}\mu_1 \otimes \mu_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\, \mathsf{d}\mu_2(y) \right) \mathsf{d}\mu_1(x) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y)\, \mathsf{d}\mu_1(x) \right) \mathsf{d}\mu_2(y).
\end{aligned}$$

Both the Fubini Theorem and the Tonelli theorem just stated have the annoying property, that they do not hold for the Lebesgue measure, since for instance the Lebesgue measure on $\mathbb{R}^2$ is not the product of the Lebesgue measures on $\mathbb{R}$ (why?). Luckily, there are versions of Fubini's and Tonelli's theorems for complete measure spaces.

Theorem 12.3.4 (Fubini, version for complete measure spaces). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ be two complete measure spaces. Let $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ be integrable on the product space $(\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \otimes \mathcal{F}_2}, \overline{\mu_1 \otimes \mu_2})$. Then, for μ_1-almost every $x \in \Omega_1$, the function $y \mapsto f(x, y)$ is $(\mathcal{F}_2, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable. Moreover, the function*

$$x \mapsto \int_{\Omega_2} f(x, y) \mathsf{d}\mu_2(y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \, \mathsf{d}\overline{\mu_1 \otimes \mu_2} &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, \mathsf{d}\mu_2(y) \right) \mathsf{d}\mu_1(x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) \, \mathsf{d}\mu_1(x) \right) \mathsf{d}\mu_2(y).\end{aligned}$$

Note that this version of Fubini's theorem assumes that the measure spaces involved are complete. Note also that the function $y \mapsto f(x, y)$ is in general no longer measurable for *all* $x \in \Omega_1$. Just as the version of Fubini's theorem for not necessarily complete measures, the integrability of the function f is assumed. Tonelli's theorem can sometimes be useful to establish this integrability.

Theorem 12.3.5 (Tonelli, version for complete measure spaces). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ be two complete, σ-finite, measure spaces. Let $f : \Omega_1 \times \Omega_2 \to [0, +\infty]$ be measurable on the product space $(\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \otimes \mathcal{F}_2}, \overline{\mu_1 \otimes \mu_2})$. Then, for μ_1-aalmost every $x \in \Omega_1$, the function $y \mapsto f(x, y)$ is $(\mathcal{F}_2, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable. Moreover, the function*

$$x \mapsto \int_{\Omega_2} f(x, y) \mathsf{d}\mu_2(y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \, \mathsf{d}\overline{\mu_1 \otimes \mu_2} &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, \mathsf{d}\mu_2(y) \right) \mathsf{d}\mu_1(x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) \, \mathsf{d}\mu_1(x) \right) \mathsf{d}\mu_2(y).\end{aligned}$$