TU/e, 2MBA70

Measure and Probability Theory

Pim van der Hoorn and Oliver Tse

Version 1.1 September 1, 2025

The occurrence of any event where the chances are beyond one in ten followed by 50 zeros is an event that we can state with certainty will never happen, no matter how much time is allotted and no matter how many conceivable opportunities could exist for the event to take place.

– Emile Borel

Disclaimer: These are lecture notes for the course *Measure and Probability Theory*. They are by no means a replacement for the lectures, instructions, and/or the books. Nor are they intended to cover every aspect of the field of measure theory or probability theory.

Since these are lecture notes, they also include problems. Each chapter ends with a set of exercises that are designed to help you understand the contents of the chapter better and master the tools and concepts.

These notes are still in progress and they almost surely contain small typos. If you see any or if you think that the presentation of some concepts is not yet crystal clear and might enjoy some polishing feel free to drop a line. The most efficient way is to send an email to us, w.l.f.v.d.hoorn@tue.nl or o.t.c.tse@tue.nl. Comments and suggestions are greatly appreciated.

Contents

1. Introduction

1.1. The need to measure

This course, as the name suggests, is about measuring and probabilities. The field of measure theory is fundamental for many modern versions of mathematical disciplines, such as analysis and probability theory. These lecture notes are designed to introduce the foundations of measure theory and highlight its use in developing the modern theory of probabilities and expectations.

At the core of measure theory is the notion of *measuring*. We do this every day; whether we are measuring the length of a wall to determine if our bed fits, making sure everyone gets an equal size pizza slice, or counting down the minutes till the end of a boring lecture.

The term measuring here refers to the act of assigning a (non-negative) number to every object in some collection. If that was the only requirement, this would be a very brief course. However, such an assignment should satisfy some properties if we want it to be useful. For example, if we want to compute the area of some complex shape, we often (without any thought) divide this shape into smaller more regular pieces for which we can compute the area and then combine these to get the area of the entire shape. If the shape is very complex, we might even approximate it by a collection of easier shapes (think of approximating with little squares for example). The actual area is then computed by taking a finer and finer approximation and considering the limit of these computed areas. In both these cases, we are (implicitly) making use of the fact that our notion of measuring respects these operations. The goal of measure theory is to provide clear mathematical definitions for the operations a proper way of measuring should respect and use these to develop new important concepts and theory. The most fundamental of which is integration.

1.2. A new theory of integration

The ability to properly measure is instrumental for integration. If you think back to your analysis course, integrating a function $f : \mathbb{R} \to [0, +\infty)$ is basically computing the area under the curve. In other words, measuring the set $A \subset \mathbb{R}^2$ given by $A := \{(x, y) \in \mathbb{R} \times \mathbb{R} \,:\, 0 \le y \le f(x)\}$. In most cases, this area is actually very complex and we need to approximate it by looking at very small sections and then combine the outcomes to get the full answer.

Being able to integrate is fundamental in a wide variety of mathematical fields. For example, solving Partial Differential Equations and analyzing their solutions requires a powerful theory of integration, as this is the inverse operation to differentiation. Another example is Harmonic Analysis, eg. Fourier Analysis in $\mathbb{R}^d$, where functions are studied by transforming them using integrals. But also Functional Analysis, which plays a fundamental part in the foundation of modern quantum mechanics, requires integration to map functions to numbers or other objects.

Finally, the field of probability theory heavily relies on being able to measure sets and integrate them (more on this later).

But we already know *Riemann* integration, so why do we need another course on this? Unfortunately, Riemann integration has undesirable issues, highlighted in these two examples:

1. Consider a sequence $(f_n)_{n\geq 1}$ of functions that are each Riemann integrable. Suppose now these functions converge point-wise to a limit function f. Then we would like to say something about whether f is Riemann integrable. Even nicer would be if $\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n = \int f$, i.e., if we can interchange limits and integration. The issue is that both these things or not generally possible, and the conditions for the interchange of limits and integration are very restrictive.

2. In general, it is even difficult to provide a practical characterization for when any function is Riemann integrable. This means that, in the worst case, you have to prove the convergence of the upper and lower Riemann sums for a function f you want to integrate.

One of the main outcomes of this course is a new theory of integration called *Lebesgue integration.* The beauty of this theory is that not only does it not suffer from any of the issues outlined above. We can easily characterize if a function is Riemann-integrable within this new theory. More importantly, any point-wise limit of Lebesgue-integrable functions is, under uniform bounds, again Lebesgue integrable and (most of the time) $\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$. Finally, the theory of Lebesgue integration also generalizes Riemann integration. That is, if you know the Riemann integral $\int f$ of a function f exists, its Lebesgue integral will have the same value. So $\int_0^1 x^2 dx$ is still equal to $1/3$, don't worry.

1.3. Measure theory as the foundation of probability theory

Aside from providing us with a new and powerful theory of integration, measure theory is the true foundation of modern probability theory.

During the first course on probability theory, Probability and Modeling (2MBS10), the concept of probabilities was introduced. The idea here (in its simplest version) is that you have a space Ω of possible outcomes of an experiment, and you want to assign a value in $[0, 1]$ to each set A of potential outcomes. This value would then represent the *probability* that the experiment will yield an outcome in this set A, and was denoted by $\mathbb{P}(A)$.

It turned out that to define these concepts, we needed to impose structure on both the space of events as well as on the probability measure. For example, if we had two sets A, B of possible outcomes, would like to say something about the probability that the outcome is in either A or B. This means that not only do we need to be able to compute $\mathbb{P}(A \cup B)$, but we want that $A \cup B$ is also an event in our space Ω. Another example concerned the probability of the outcome not being in A, which means computing the probability of the event $\Omega \setminus A$, requiring that this set should also be in Ω.

In the end, this prompted the definition of an *event space* which was a collection $\mathcal{F}$ of subsets of Ω satisfying a certain set of properties. In addition, the probability assignment $\mathbb{P}$ was defined as a map $\mathbb{P} : \mathcal{F} \to [0, 1]$ with some addition properties, such as $\mathbb{P}(\Omega) = 1$.

With this setup, it was then possible to define what a *random variable* is. Here a random variable X was defined on a triple $(\Omega, \mathcal{F}, \mathbb{P})$, consisting of a space of outcomes, an event space and a probability on that space. Formally it is a mapping $X : \Omega \to \mathbb{R}$ such that for each $x \in \mathbb{R}$ the set $X^{-1}(-\infty, x) := \{\omega \in \Omega \, : \, X(\omega) \in (-\infty, x)\}$ is in $\mathcal{F}$. This then allowed us to define the *cumulative distribution function* as $F_X(x) := \mathbb{P}\big(X^{-1}(-\infty, x)\big)$.

It is important to note here that already it was needed to make a distinction between how to define a discrete and a continuous random variable. In addition, a separate definition was required to define multivariate distribution functions. All of this limits the extent to which this theory can be applied. For example, let U be distributed uniformly on $[0, 1]$ and Y be distributed uniformly on the set $\{1, 2, \dots, 10\}$ and define the random variable X to be equal to U with probability $1/3$ and equal to Y with probability $2/3$. How would you deal with this random variable, which is both discrete and continuous?

The setting would become even more complex and fuzzy if we were not talking about random numbers in $\mathbb{R}$ but, say, random vectors of infinite length or random functions. Do these even exist? Many other things remain fuzzy or simply impossible in a theory of probability without measure theory. What are conditional probabilities/expectations? How do you define a continuous time Markov Process or a point process? What is a stochastic process? Or, does there exist such a thing as a random probability measure?

The solutions to all these issues come from a generalization of event spaces and probability measures introduced above. These go by the names *sigma-algebra* and *measure*, respectively, which are the core concepts in measure theory. With this, we can then define when any mapping between spaces is *measurable* and use such mappings to define random objects in the space such a function maps to. Finally, armed with the theory of Lebesgue integration, measure theory provides the foundation to define expectations, convergence of random variables, and, most importantly, the notion of conditional probability/expectation.

All of this is to say that a proper study of Probability Theory cannot happen without Measure Theory. By the end of these notes, we hope you will appreciate this and be inspired by the versatility and beauty of measures theory and Lebesgue integration.

2. Measure Spaces (σ-Algebras and Measures)

At the core of measure theory are two things: 1) the objects we want to measure and 2), a way to assign a measure (value) to these objects. The objects are subsets of some given space that satisfy certain properties, which we call *σ-algebras* (sigma-algebras). The structure of these σ-algebras allows us to define the notion of a measure on them, which is a map that assigns to each set a value in $[0, \infty)$. Of course, we will not consider any such map but impose a few properties which will imply many interesting properties of measures and allow us later on to define a general notion of integration. This chapter is concerned with the two basic notions, σ-algebras and measures. We will provide the definitions, important properties, and some key examples that will be fundamental for the remainder of this course.

2.1. Sigma-algebras

2.1.1. Definition and examples

We begin this section with introducing the general structure needed on a collection of sets to be able to assign a notion of measurement to them. Such a collection is called a sigma-algebra, often written as σ-algebra.

Definition 2.1: Sigma Algebra

A *σ-algebra* $\mathcal{F}$ on a set Ω is a collection of subsets of Ω with the following properties:

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$;
2. For every $A \in \mathcal{F}$, it holds that $A^c := \Omega \setminus A \in \mathcal{F}$;
3. For every sequence $A_1, A_2, \dots \in \mathcal{F}$, it holds that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A set $A \in \mathcal{F}$ is called *$\mathcal{F}$-measurable*, or simply *measurable* if it is clear which σ-algebra is meant.

This definition might look very familiar. In the course Probability and Modeling you have been introduced to the concept of an *event space*. It turns out that these concepts are actually the same, see Problem 2.1.

Before we give some examples, we first provide a result that states that any σ-algebra is closed under countable intersections. The proof is left as an exercise to the reader.

Lemma 2.2

Let $\mathcal{F}$ be a σ-algebra on Ω and let $A_1, A_2, \dots \in \mathcal{F}$. Then it holds that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

Proof. See Problem 2.2. ☺

We now give some examples and non-examples of σ-algebras.

Example 2.1 ((Non-)Examples of σ-algebras).

(a) The collection $\mathcal{F} = \{\emptyset, \Omega\}$ is a σ-algebra. This is called the *trivial σ-algebra* or the *minimal* σ-algebra on Ω.

(b) For any subset $A \subset \Omega$ we have that $\mathcal{F} := \{\emptyset, A, \Omega \setminus A, \Omega\}$ is a σ-algebra.

(c) The *power set* $\mathcal{P}(\Omega)$ (the collection of all possible subsets of Ω) is a σ-algebra. This is sometimes called the *maximal σ-algebra* on Ω.

(d) For any subset $A \subset \Omega$, $A \neq \emptyset, \Omega$, we have that $\mathcal{F} := \{\emptyset, A, \Omega\}$ is **not** a σ-algebra.

(e) Let $\Omega = [0, 1]$ and $\mathcal{F}$ be the collections of finite unions of intervals of the form $[a, b]$, $[a, b)$, $(a, b]$ and (a, b) for $0 \leq a < b \leq 1$. Then $\mathcal{F}$ is **not** a σ-algebra.

(f) Let $f : \Omega \to \Omega'$ and let cF' be a σ-algebra on Ω'. Then the collection

$$\mathcal{F} := f^{-1}(\mathcal{F}') = \{f^{-1}(A') \,:\, A' \in \mathcal{F}'\},$$

is a σ-algebra. The converse to this is not always true, see Problem 2.4.

Proving these claims is left as an exercise, see Problem 2.3.

The idea of measure theory is that one can assign a measure to each set in a σ-algebra. In line with this, a pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ a σ-algebra on Ω is called a *measurable space.*

2.1.2. Constructing σ-algebras

We now know what a σ-algebra is and have seen some examples and some non-examples. But the examples we have seen are still quite uninspiring. We will actually discuss a very important σ-algebra in the next section. But for now, we will describe several ways to construct σ-algebras. The first is restricting an existing σ-algebra to a given set.

Lemma 2.3: Restriction of a σ-algebra

Let $(\Omega, \mathcal{F})$ be a measurable space and $A \subset \Omega$. Then the collection defined by

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

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

Proof. We need to check all three properties.

1. Since $A \cap \Omega = A$ and $A \cap \emptyset = \emptyset$, it follows that $A, \emptyset \in \mathcal{F}_A$.

2. Consider a set $C \in \mathcal{F}_A$. Then by definition $C = A \cap B$ for some $B \in \mathcal{F}$. Next, we note
$$A \setminus C = A \setminus (A \cap B) = A \cap (\Omega \setminus B).$$
Since $\mathcal{F}$ is a σ-algebra, it follows that $\Omega \setminus B \in \mathcal{F}$ and so $A \setminus C \in \mathcal{F}_A$.

3. Let $C_1, C_2, \ldots$ be sets in $\mathcal{F}_A$. Then there are $B_1, B_2, \cdots \in \mathcal{F}$ such that $C_i = A \cap B_i$. Hence
$$\bigcup_{i=1}^{\infty} C_i = \bigcup_{i=1}^{\infty} A \cap B_i = A \cap \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}_A,$$
since $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$ because this is a σ-algebra. ☺

While it is nice to be able to take a given σ-algebra and create a possibly smaller one by restricting it to a given set, we might also want to start with a given collection of sets $\mathcal{A}$ and then create a σ-algebra that contains this collection. Of course, the powerset $\mathcal{P}(\Omega)$ will always work. However, it is not always desirable to take this maximal σ-algebra. It would be much better if we could create the smallest σ-algebra that contains $\mathcal{A}$. It turns out that this can be done and the resulting σ-algebra is said to be *generated by* $\mathcal{A}$.

Proposition 2.4: Generated σ-algebra

Let $\mathcal{A}$ be a collection of subsets of Ω and denote by $\Sigma_{\mathcal{A}}$ the collection of all σ-algebras on Ω that contain $\mathcal{A}$. Then the collection defined by
$$\sigma(\mathcal{A}) := \bigcap_{\mathcal{F} \in \Sigma_{\mathcal{A}}} \mathcal{F} \qquad \text{is a } \sigma\text{-algebra.}$$
It is called the *σ-algebra generated by* $\mathcal{A}$. Equivalently, $\mathcal{A}$ is called the *generator of* $\sigma(\mathcal{A})$.

Moreover, $\sigma(\mathcal{A})$ is the smallest σ-algebra that contains $\mathcal{A}$. If $\mathcal{F}$ is a σ-algebra on Ω and $\mathcal{A}$ is a collection of subsets such that $\mathcal{F} = \sigma(\mathcal{A})$, we call $\mathcal{A}$ the *generator of* $\mathcal{F}$.

Proof. If we can show that $\sigma(\mathcal{A})$ is a σ-algebra, then the claim about it being the smallest σ-algebra that contains $\mathcal{A}$ follows from its definition. So we will prove that $\sigma(\mathcal{A})$ is a σ-algebra.

Similar to Lemma 2.3 we need to check all the requirements.

1. Since $\emptyset, \Omega \in \mathcal{F}$ holds for every $\mathcal{F} \in \Sigma_{\mathcal{A}}$ it follows that $\emptyset, \Omega \in \sigma(\mathcal{A})$. In particular, we note that $\sigma(\mathcal{A})$ is not empty.

2. Take $A \in \sigma(\mathcal{A})$. Then $A \in \mathcal{F}$ for each $\mathcal{F} \in \Sigma_{\mathcal{A}}$. Since $\mathcal{F}$ is a σ-algebra it holds that $\Omega \setminus A \in \mathcal{F}$ for each $\mathcal{F} \in \Sigma_{\mathcal{A}}$. This then implies that $\Omega \setminus A \in \sigma(\mathcal{A})$.

3. Let $(A_i)_{i\in\mathbb{N}}$ be a sequence of sets in $\sigma(\mathcal{A})$. Then by definition $A_i \in \mathcal{F}$ for each $\mathcal{F} \in \Sigma_{\mathcal{A}}$. Hence
$$\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F},$$
holds for each $\mathcal{F} \in \Sigma_{\mathcal{A}}$ and thus it follows that $\bigcup_{i\in\mathbb{N}} A_i \in \sigma(\mathcal{A})$. 😄

The nice thing about this construction of σ-algebras is that it respects inclusions.

Lemma 2.5: Inclusion property of σ-algebras

If $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$ are collection of subsets of Ω, then also $\sigma(\mathcal{A}) \subset \sigma(\mathcal{B}) \subset \sigma(\mathcal{C})$.

Proof. See Problem 2.5 😄

An example of a generated σ-algebra is to construct products of measurable spaces.

Definition 2.6: Product σ-algebra

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. Then we define $\mathcal{F} \otimes \mathcal{F}'$ to be the σ-algebra on $\Omega \times \Omega'$ generated by sets of the form $A \times B$, with $A \in \mathcal{F}$ and $B \in \mathcal{F}'$.

However, there is a much more important σ-algebra that is constructed from a generator set.

2.1.3. Borel σ-algebra

The Euclidean space $\mathbb{R}^d$ is omnipresent in mathematics and hence pops up in many bachelor courses as well. In particular, the concept of random variables, as given in the course Probability and Modeling, is mainly concerned with objects that have values in $\mathbb{R}$. Based on this, the need to impose a measurable structure on this space, by means of a σ-algebra, should not come as a surprise. It turns out that there is a canonical σ-algebra which is called the *Borel σ-algebra* and is named after the French mathematician Émile Borel, one of the pioneers of measure theory.

In order to define the Borel σ-algebra we need the notion of an open set in $\mathbb{R}^d$. For any $x \in \mathbb{R}$ and $r > 0$, we denote by $B_x(r) := \{y \in \mathbb{R}^d : \|x - y\| < r\}$ the open ball of radius r around x. A set $U \subset \mathbb{R}^d$ is called *open* if and only if for every $x \in U$, there exists an $r > 0$ such that $B_x(r) \subset U$.

Definition 2.7: Borel σ-algebra

The *Borel σ-algebra* on $\mathbb{R}^d$, denoted by $\mathcal{B}_{\mathbb{R}^d}$, is the σ-algebra generated by all open sets in $\mathbb{R}^d$. Elements of $\mathcal{B}_{\mathbb{R}^d}$ are called *Borel sets*.

Remark. From the definition, it should be clear that one can define a *Borel σ-algebra* on any metric space. In fact, we can define it in any topological space. However, this requires the notion of a topology which is beyond the scope of this course.[1]

While this is a perfectly fine definition, it is often cumbersome to work with. It is, therefore, convenient that $\mathcal{B}_{\mathbb{R}^d}$ is generated by other, more compact, collections of sets. At this point, we state the result for the one-dimensional Borel σ-algebra.

Proposition 2.8

The Borel σ-algebra on $\mathbb{R}$ is the σ-algebra generated by any of the following family of sets,

(1) $\{(a, b)\}$,

(2) $\{(a, b]\}$,

(3) $\{[a, b)\}$,

(4) $\{(-\infty, a]\}$,

(5) $\{(-\infty, a)\}$,

(6) $\{[a, \infty)\}$,

(7) $\{(a, \infty)\}$,

where $a, b \in \mathbb{Q}$, or $a, b \in \mathbb{R}$

Proof. See Problem 2.6. 😄

2.2. Measures

2.2.1. Definition and examples

In the previous section, we have seen how we can define and construct collections of sets that we would like to be able to measure. It turned out that this collection should satisfy some properties. Likewise, when defining the notion of a *measure* we also will require it to have certain properties.

The main property we require is called *σ-additive*. Consider any collection $\mathcal{C}$ of subsets of some set Ω. Then a set function $\mu : \mathcal{C} \to [0, \infty]$ is called *σ-additive* if for any countable family $(A_i)_{i\in\mathbb{N}}$ of pairwise disjoint sets in $\mathcal{C}$

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

[1] Those who are interested can have a look at *Foundations of modern probability* by Olav Kallenberg.

Definition 2.9: Measure

Let $(\Omega, \mathcal{F})$ be a measurable space. A set function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure on* $(\Omega, \mathcal{F})$ if the following holds:

(1) $\mu(\emptyset) = 0$ and,

(2) μ is σ-additive.

Any set function $\mu : \mathcal{C} \to [0, \infty]$ on a collection $\mathcal{C}$ of subsets of some set Ω that satisfies the two properties in Definition 2.9 is called a *pre-measure*.

A triple $(\Omega, \mathcal{F}, \mu)$, consisting of a measurable space $(\Omega, \mathcal{F})$ and a measure μ on that space is called a *measure space*. We will often work with measure spaces that have an addition property.

Definition 2.10: σ-finite measure space

A measure space $(\Omega, \mathcal{F}, \mu)$ is called *σ-finite* if there exists a family $(A_i)_{i\geq 1}$ of measurable sets such that $\mu(A_i) < \infty$ for all $i \geq 1$ and $\bigcup_{i\geq 1} A_i = \Omega$.

Let us give some simple examples of measures.

Example 2.2 (Examples of measures).

1. *(Trivial measures)* Let $(\Omega, \mathcal{F})$ be a measurable space. Then the following two set functions are measures:

$$\mu(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ \infty & \text{otherwise.} \end{cases} \quad \text{and} \quad \mu(A) = 0 \quad \text{for all } A \in \mathcal{F}.$$

2. *(Dirac measure)* Let $(\Omega, \mathcal{F})$ be a measurable space and $x \in \Omega$. Then the function

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise,} \end{cases}$$

is a measure called the *Dirac delta measure* or *unit mass* at x.

3. *(Counting measure)* Let $(\Omega, \mathcal{F})$ be a measurable space. Then the function defined as

$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is a finite set,} \\ \infty & \text{otherwise,} \end{cases}$$

is a measure called the *counting measure*.

4. *(Discrete measure)* Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a countable set and consider the measurable space $(\Omega, \mathcal{P}(\Omega))$. Take any sequence of $(a_i)_{i\in\mathbb{N}}$ such that $\sum_{i=1}^{\infty} a_i < \infty$. Then the function
$$\mu(A) = \sum_{j=1}^{\infty} a_j \delta_{\omega_j}(A),$$
is a measure called the *discrete measure*. If the a_i are such that $\sum_{i=1}^{\infty} a_i = 1$ we call this the *discrete probability measure*.

It should be noted that, outside maybe the discrete measure, the examples listed above do not include any interesting measure.

2.2.2. Important properties

Although the number of properties a measure needs to satisfy is very limited, they actually imply a great number of other important properties. We will start with the basic ones, which relate the measure of a set that is obtained from a given set operation on two sets A, B to the measure of these sets.

Proposition 2.11: Basic properties of measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $A, B \in \mathcal{F}$. Then the following properties hold for μ:

1. (finitely additive) If $A \cap B = \emptyset$, then $\mu(A \cup B) = \mu(A) + \mu(B)$.
2. (monotone) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
3. (exclusion) If in addition $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.
4. (strongly additive) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.
5. (subadditive) $\mu(A \cup B) \leq \mu(A) + \mu(B)$.

Proof.

1. Let $A_1 = A$, $A_2 = B$ and $A_i = \emptyset$ for all $i \geq 3$. Then this property follows directly from the fact that μ is σ-additive.
2. Since $A \subseteq B$ we have that $B = A \cup (B \setminus A)$, with A and $B \setminus A$ disjoint sets. It then follows from property 1 that $\mu(B) = \mu(A) + \mu(B \setminus A)$ and thus $\mu(A) \leq \mu(B)$.
3. Since $\mu(A) < \infty$ we can subtract $\mu(A)$ from both sides of the equation $\mu(B) = \mu(A) + \mu(B \setminus A)$ to obtain the desired result.
4. See Problem 2.9.
5. Property 4 implies that $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B) \geq \mu(A \cup B)$. ☺

In addition to properties relating a measure μ to set operations, we also want to understand what happens if we take a limit of the measures of an infinite family of sets. Let $(A_i)_{i\in\mathbb{N}}$ be a family of measurable sets. We say this family is *increasing* if $A_i \subset A_{i+1}$ holds for all $i \in \mathbb{N}$. Because a measure is monotone it follows that the sequence $(\mu(A_i))_{i\in\mathbb{N}}$ is a monotone sequence in $[0,\infty]$. So a natural question would be: what is the limit of this sequence? A similar question can be asked about a *decreasing* sequence, i.e. $A_{i+1} \subseteq A_i$. It turns out that in both cases, the limit can be expressed as the measure of the union or intersection of all sets, respectively.

Proposition 2.12: Continuity of measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(A_i)_{i\in\mathbb{N}}$ be an family of measurable sets.

1. If $(A_i)_{i\in\mathbb{N}}$ is an increasing family then
$$\lim_{i\to\infty} \mu(A_i) = \mu\left(\bigcup_{i\in\mathbb{N}} A_i\right).$$
2. If $(A_i)_{i\in\mathbb{N}}$ is an decreasing family and $\mu(A_1) < \infty$, then
$$\lim_{i\to\infty} \mu(A_i) = \mu\left(\bigcap_{i\in\mathbb{N}} A_i\right).$$

The first property in Proposition 2.12 is called *continuity from below*, while the second one is called *continuity from above*.

Proof. See Problem 2.10. 😄

With these continuity properties we can show that the subadditive property of any measure can actually be extended to any countable family of sets.

Lemma 2.13: Measures are σ-subadditive

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(A_i)_{i\in\mathbb{N}}$ be a family of sets in $\mathcal{F}$. Then
$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) \le \sum_{i=1}^{\infty} \mu(A_i),$$
and the measure μ is said to be *σ-subadditive.*

Proof. See Problem 2.11 😄

The continuity properties in 2.12 are actually powerful enough to characterize a measure. Below we provide one such alternative characterization. We refer to problem [??] for more versions.

Theorem 2.14: Alternative definition of a measure

Let $(\Omega, \mathcal{F})$ be a measurable space. A set function $\mu : \mathcal{F} \to [0, \infty]$ is a measure if, and only if,

1. $\mu(\emptyset) = 0$,
2. $\mu(A \cup B) = \mu(A) + \mu(B)$, for any two disjoint sets $A, B \in \mathcal{F}$, and
3. for any increasing family $(A_i)_{i\in\mathbb{N}}$ of measurable sets such that $A_\infty := \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$, it holds that
$$\mu(A_\infty) = \lim_{i\to\infty} \mu(A_i) \quad \left(= \sup_{i\in\mathbb{N}} \mu(A_i)\right).$$

Proof. The fact that any measure satisfies these three properties follows from the definition and Propositions 2.11 and 2.12. So let us now assume that we have a set function μ that satisfies the three properties listed above. Then to show that μ is a measure we only have to prove that it is σ-additive.

To this end, let $(A_i)_{i\in\mathbb{N}}$ be a family of pairwise disjoint measurable sets. Now define $B_k = \bigcup_{i=1}^k A_i$ and note that $B_k \in \mathcal{F}$ for all $k \in \mathbb{N}$ and

$$B_\infty := \bigcup_{k\in\mathbb{N}} B_k = \bigcup_{i\in\mathbb{N}} A_i.$$

Using property 2. we get that $\mu(B_k) = \sum_{i=1}^k \mu(A_i)$ while property 3. now implies that

$$\mu(B_\infty) = \lim_{k\to\infty} \mu(B_k) = \lim_{k\to\infty} \sum_{i=1}^k \mu(A_i) = \sum_{i=1}^\infty \mu(A_i).$$ ☺

Finally, let us discuss a uniqueness result for measures. In Section 2.1.2 we discussed how to construct σ-algebras from a generator set $\mathcal{A}$. Suppose now that we have two measures μ_1 and μ_2 agree on $\mathcal{A}$, that is $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{A}$. Then we would intuitively expect that they should agree on the entire σ-algebra $\sigma(\mathcal{A})$. This turns out to be true, under some small conditions on the generator set.

Theorem 2.15: Uniqueness of measures

Let $(\Omega, \mathcal{F})$ be a measurable space where $\mathcal{F} = \sigma(\mathcal{A})$ with $\mathcal{A}$ satisfying the following properties:

1. for any $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, and
2. there exists a sequence $(A_i)_{i\in\mathbb{N}}$ with $\Omega = \bigcup_{i\in\mathbb{N}} A_i$.

Then any two measure μ_1 and μ_2 that are equal on $\mathcal{A}$ and are finite on every element of the sequence $(A_i)_{i\in\mathbb{N}}$ are equal on the entire σ-algebra $\mathcal{F} = \sigma(\mathcal{A})$.

The proof of this theorem is outside the scope of this course, as it is requires another more technical results. What is important is the implication of Theorem 2.15: to study a measure on $\sigma(\mathcal{A})$ it suffices to look at what it does on the generator $\mathcal{A}$.

2.2.3. Construction of measures

Definition 2.16: semi-Algebra

A family $\mathcal{S}$ of subsets of Ω is called a *semi-algebra* if the following holds:

1. $\emptyset \in \mathcal{S}$;
2. For any $A, B \in \mathcal{S}$ we have $A \cap B \in \mathcal{S}$;
3. For any $A, B \in \mathcal{S}$, there exists a finite family of mutually disjoint sets $C_1, \ldots, C_m \in \mathcal{S}$ such that $A \setminus B = \bigcup_{i=1}^m C_i$.

A simple example of a semi-algebra is the set of singletons $\mathcal{S} = \{\{\omega\} \in \Omega\} \cup \{\emptyset, \Omega\}$ for any finite set Ω. A more important example in the case of $\mathbb{R}$ is the collection of right-closed intervals

$$\mathcal{S} = \{\emptyset\} \cup \{(a, b] : a \leq b, a, b \in \mathbb{R}\} \cup \{(a, \infty) : a \in \mathbb{R}\} \cup \{\mathbb{R}\}, \tag{2.1}$$

see Problem 2.7. Finally, as the name might suggest, any σ-algebra is also a semi-algebra (see Problem 2.8).

The utility of this somewhat seemingly abstract notion of a semi-algebra is that it serves as the starting point to construct measures. Let $\mu : \mathcal{S} \to [0, \infty]$ be a pre-measure on a semi-algebra $\mathcal{S}$, i.e it satisfies the two properties of a measure as given in Definition 2.9 but $\mathcal{S}$ does not have to be a σ-algebra. Then it is possible to extended μ to a proper, and often unique, measure on the generated σ-algebra $\sigma(\mathcal{S})$.

In addition to this, the measure will be defined on a slightly larger σ-algebra that has some nice properties. To formally state this we need to introduce the concept of *null sets* and *complete measure spaces*.

Definition 2.17: Null set and complete measure space

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A set $N \subset \Omega$ is called a *null set* if there exists a $A \in \mathcal{F}$ such that $N \subset A$ and $\mu(A) = 0$.

We call a measure space $(\Omega, \mathcal{F}, \mu)$ *complete* if every null set $N \in \mathcal{F}$.

It is important to note that a null set does not have to be measurable, i.e., be in $\mathcal{F}$.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, not necessarily complete. Then we can construct a new measure space $(\Omega, \overline{\mathcal{F}}, \bar{\mu})$ that is complete and such that the measure $\bar{\mu}$ is equal to μ on $\mathcal{F}$, i.e.

$\bar{\mu}|_{\mathcal{F}} = \mu$. We refer to this construction as *completing* the measure space $(\Omega, \mathcal{F}, \mu)$. The details of this construction are not important. It basically entails adding all null sets to the σ-algebra. For more details see Problem 2.12.

With this setup we can not state the main result that allows us to extend pre-measures to complete measures.

Theorem 2.18: Carathéodory Extention Theorem

Let $\mathcal{S}$ be a semi-algebra on Ω and let $\mu : \mathcal{S} \to [0, \infty]$ be a pre-measure on $\mathcal{S}$. Then there exists a measure μ^* on $\overline{\sigma(\mathcal{S})}$ (the completed version of $\sigma(\mathcal{S})$) that is an extension of μ, i.e. $\mu^*|_{\mathcal{S}} = \mu$. Moreover, if μ is σ-finite the measure μ^* is unique.

The proof of this theorem is outside the scope of this course and will not be covered in these lecture notes. The interested student can look at [REF Appendix] for the details.

Theorem 2.18 is powerful, as it allows us to create measures on the σ-algebra $\sigma(\mathcal{S})$ by only describing a pre-measure on the semi-algebra $\mathcal{S}$ that generates it. This is especially useful if there is no direct description of the $\sigma(\mathcal{S})$.

We will an application of Theorem 2.18 later in Section [REF] for constructing random variables and in Section 5.1 for constructing products of measure spaces. However, its most crucial application is to construct a measure on the Borel sigma-algebra, which will be the key object for defining a new notion of integration.

2.2.4. The Lebesgue measure

The Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ plays a central role in the development of a new theory of integration. However, for this we need to first to turn it into a measure space. This requires us to define a measure on $\mathcal{B}_{\mathbb{R}^d}$. But how can we construct a non-trivial measure on this space? The problem is that the Borel σ-algebra is only defined in terms of its generator. Hence if we want to define what $\mu(A)$ is for any $A \in \mathcal{B}_{\mathbb{R}^d}$ we first have to get a better handle on the full σ-algebra. That might seem daunting, and it really is. The problem becomes even more challenging when we want the measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ to have additional properties. For example, the measure of any rectangle should be simply its volume, which seems like a very natural property to ask for. Moreover, it also seems natural to require that the measure of any rectangle is independent of where it is, i.e., we want the measure to be translational and rotational invariant.

Fortunately, it turns out that such a measure does exist. This fundamental measure is called the *Lebesgue measure*, named after the French mathematician Henri Lebesgue who was the architect of the modern notion of integration we will cover in this course. Moreover, in addition to the measure of any rectangle being equal to its volume, the Lebesgue measure has several other strong features.

We can now state the main result that proves the existence of the Lebesgue measure and its important properties.

Theorem 2.19: Lebesgue measure

There exists a σ-algebra $\mathcal{L}^d \supset \mathcal{B}_{\mathbb{R}^d}$ on $\mathbb{R}^d$ and a unique measure λ such that $(\mathbb{R}^d, \mathcal{L}^d, \lambda)$ is complete and satisfies the following properties for any $B \in \mathcal{B}_{\mathbb{R}^d}$:

1. For any half-open rectangle $I := [a_1, b_1) \times \cdots \times [a_d, b_d)$ it holds that
$$\lambda(I) = \prod_{i=1}^{d}(b_i - a_i);$$
2. For any $x \in \mathbb{R}^d$, $\lambda(B + x) = \lambda(B)$, where $B + x = \{y + x : y \in B\}$;
3. For any combination of translation, rotation and reflection R, $\lambda(R^{-1}(B)) = \lambda(B)$;
4. For any invertible matrix $M \in \mathbb{R}^{d\times d}$, $\lambda(M^{-1}(B)) = |\mathrm{det} M|^{-1}\lambda(B)$.

At its hearth, the proof of this theorem relies on the Carathéodory Extention Theorem (Theorem 2.18) applied to the set of right-closed intervals in $\mathbb{R}^d$ (see (2.1) for the one-dimensional setting). Some additional work is needed to make it complete and then check the four properties. The interested student is referred to the Appendix, where we provide the full details.

It follows from Theorem 2.19 that the Lebesgue measure is formally defined on a larger σ-algebra $\mathcal{L}^d$ than the Borel σ-algebra. This σ-algebra is called the *Lebesgue σ-algebra* and functions that are $\mathcal{L}^d$-measurable are called *Lebesgue measurable*. The Lebesgue measure on $\mathcal{B}_{\mathbb{R}^d}$ is now defined as the restriction of λ to the Borel σ-algebra.

Remark (Lebesgue vs Borel measurable). It should be noted that $\mathcal{B}_{\mathbb{R}^d} \subsetneq \mathcal{L}^d$. That is, there are sets that are Lebesgue measurable but not Borel measurable (eg. subsets of the Cantor set).

Having established the most important measure for this course, we end this section by looking at some important general properties of measures.

2.3. Problems

Problem 2.1. Recall that an *event space* is a collection $\mathcal{F}$ of subsets of Ω such that

(a) $\mathcal{F}$ is non-empty;

(b) If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$;

(c) If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Show that the definition of an *event space* is equivalent to that of a σ-algebra as given in Definition 2.1.

Problem 2.2. Prove Lemma 2.2. [Hint: how are intersections related to the other operations used in the definition of a σ-algebra?]

Problem 2.3. Prove the claims made in Example 2.1.

Problem 2.4. Provide a counter example to the statement: if $(\Omega, \mathcal{F})$ is a measurable space and $f : \Omega \to \Omega'$. Then $f(\mathcal{F})$ is a σ-algebra on Ω'.

Problem 2.5. Prove Lemma 2.5.

Problem 2.6. The goal of this problem is to prove Proposition 2.8. We will do this in several stages. First we will show point 1, that the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ is generated by either the family $\mathcal{A}_1 := \{(a, b), a, b \in \mathbb{Q}\}$ or $\mathcal{A}_1' := \{(a, b), a, b \in \mathbb{R}\}$.

(a) Prove that $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_1') \subset \mathcal{B}_{\mathbb{R}}$. [Hint: what is the relation between an interval (a, b) and an open set?]

(b) We will now focus on the intervals with rational endpoints. Show that for any open set $O \subset \mathbb{R}$
$$O = \bigcup_{I \in \mathcal{A}_1, I \subset O} I$$

(c) Prove that $\sigma(\mathcal{A}_1) = \mathcal{B}_{\mathbb{R}}$. [Hint: You only need one inclusion, for which you can use 2 and the fact that $\mathbb{Q}$ is countable.]

(d) Prove that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A}_1')$.

We now move to the other family of sets. By symmetry of the families in 2 and 3 it suffices to prove only 2, the other proof will be almost identical.

(e) Show that for any $a < b \in \mathbb{R}$
$$(a, b] = \bigcap_{j \in \mathbb{N}} (a, b + \frac{1}{j}).$$

(f) Show that for any $a < b \in \mathbb{R}$
$$(a, b) = \bigcup_{j \in \mathbb{N}} (a, b - \frac{1}{j}].$$

(g) Prove that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A}_2') = \sigma(\mathcal{A}_2)$, where $\mathcal{A}_2 = \{(a, b] \ : \ a, b \in \mathbb{Q}\}$ and $\mathcal{A}_2' = \{(a, b] \ : \ a, b \in \mathbb{R}\}$.

This basically covers the full set of ideas to prove the rest of Proposition 2.8. We invite you to work these out yourself to practice with these kind of arguments. For this problem however we will ask you to explain the idea for the proofs.

(h) Explain what changes in the proof of point 3 of Proposition 2.8 from the proof of point 2 of this proposition outlined above.

(i) Describe the proof strategy to get points 4-8 of Proposition 2.8 using 1-4.

Problem 2.7. Recall the definition of right-closed intervals:

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

(a) Show that $\mathcal{S}$ is a semi-algebra.

(b) Prove that $\sigma(\mathcal{S}) = \mathcal{B}_{\mathbb{R}}$.

Problem 2.8. Prove that any σ-algebra is a semi-algebra.

Problem 2.9. Prove the *strongly additive* property of Proposition 2.11. [Hint: write $A \cup B$ as the union of three disjoint sets.]

Problem 2.10. The goal of this exercise is to prove Proposition 2.12. We will start with part 1. Recall that $(A_i)_{i\geq 1}$ is an increasing family of measurable sets.

(a) Using the family $(A_i)_{i\geq 1}$, construct a new family of measurable sets $(E_k)_{k\geq 1}$ that are mutually disjoint and that satisfy

$$\bigcup_{i\geq 1} A_i = \bigcup_{k=1}^{\infty} E_k \quad \text{and} \quad A_i = \bigcup_{k=i}^{k} E_k.$$

[Hint: Look at the points from Ω that are added when going from A_i to A_{i+1}]

(b) Use σ-additivity to show that

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

(c) Prove part 2 of Proposition 2.12 [Hint: The proof is very similar to that of the first part.]

Problem 2.11. Prove Lemma 2.13. [Hint: Construct a new increasing family of sets and use properties σ-additive, monotonicity and continuity from below.]

Problem 2.12. The goal of this problem is to complete a given measure space. To this end, let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $\mathcal{N}$ be the family of null sets of μ and define the family of sets $\overline{\mathcal{F}}$ as

$$\overline{\mathcal{F}} := \{A \cup N \,:\, A \in \mathcal{F} \text{ and } N \in \mathcal{N}\}.$$

(a) Show that $\overline{\mathcal{F}}$ is a σ-algebra that contains $\mathcal{F}$.

Define the set function $\bar{\mu} : \overline{\mathcal{F}} \to [0, \infty]$ as

$$\bar{\mu}(A \cup N) := \mu(A).$$

(b) Prove that $\bar{\mu}$ is a measure on $\overline{\mathcal{F}}$.

(c) Show that $\bar{\mu}|_{\mathcal{F}} = \mu$.

(d) Conclude that $(\Omega, \overline{\mathcal{F}}, \bar{\mu})$ is a complete measure space.

3. Measurable Functions

Now that we have defined measure spaces $(\Omega, \mathcal{F}, \mu)$, through σ-algebras and measures and studied properties of both these objects, it is time to look at functions that preserve the measurable structure of the spaces involved, called *measurable functions*. We do this first in a general setting and then move to functions that map to the real line $\mathbb{R}$.

3.1. Measurable functions

3.1.1. Definition and properties

We want to consider functions $f : \Omega \to E$ between measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ that preserve the measurable structure, as imposed by the σ-algebras. It turns out that the best way to do this it to look at the preimage of measurable sets in E.

Definition 3.1: Measurable function

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

It is important to note that whether a function is measurable or not depends on the σ-algebras we consider in each of the measurable spaces. This means that a function $f : \Omega \to E$ might be $(\mathcal{F}, \mathcal{G})$-measurable but not $(\mathcal{F}', \mathcal{G})$-measurable for a different sigma algebra $\mathcal{F}'$ on Ω.

We will often omit the explicit reference to the σ-algebras in the definition of a measurable function if it is clear which σ-algebras are considered. That is, we will simply say that the function f between the two measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ is *measurable*. We will sometimes make the choice of σ-algebras explicit by saying that $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ is measurable.

An important example of measurable functions to $\mathbb{R}$ are the indicator functions.

Example 3.1 (Indicator functions are measurable). Let $(\Omega, \mathcal{F})$ be a measurable space, $A \in \mathcal{F}$ and $f : \Omega \to \mathbb{R}$ be defined as $f = \mathbf{1}_A$, that is

$$f(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is measurable.

To see this, first note that $f^{-1}(\{1\}) = A \in \mathcal{F}$ and $f^{-1}(\{0\}) = \Omega \setminus A \in \mathcal{F}$. This implies that for any set $B \in \mathcal{B}_{\mathbb{R}}$ we have that $f^{-1}(B \cap \{x\}) \in \mathcal{F}$ with $x = 0, 1$. Hence

$$f^{-1}(B) = f^{-1}(B \cap \{0\}) \cup f^{-1}(B \cap \{1\}) \in \mathcal{F}.$$

The fact that measurability of f depends on the σ-algebras involved means we need to take a bit of care when considering operations on functions, as these might destroy the measurability. The most natural operation we should check first is composition, as we would like to be able to compose measurable functions into measurable functions. Luckily this is possible.

Proposition 3.2: Composition of measurable functions

Let $(\Omega_i, \mathcal{F}_i)$, for $i = 1, 2, 3$ be three measurable spaces and $f : \Omega_1 \to \Omega_2$, $g : \Omega_2 \to \Omega_3$ be two measurable functions. Then the composition $h := g \circ f : \Omega_1 \to \Omega_3$ is measurable.

Proof. By definition, we need to show that for every $A \in \mathcal{F}_3$ the preimage $h^{-1}(A) \in \mathcal{F}_1$. First note that

$$\begin{aligned} h^{-1}(A) = (g \circ f)^{-1}(A) &= \{x \in \Omega \,:\, g(f(x)) \in A\} \\ &= \{x \in \Omega \,:\, f(x) \in g^{-1}(A)\} = f^{-1}(g^{-1}(A)). \end{aligned}$$

Since g is $(\mathcal{F}_2, \mathcal{F}_3)$-measurable, $g^{-1}(A) \in \mathcal{F}_2$. Then, using that f is $(\mathcal{F}_1, \mathcal{F}_2)$-measurable, we conclude that $h^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{F}_1$ as was required to show. 😄

The next result shows that we can also restrict a measurable function $f : \Omega \to E$ to a measurable subset $A \subset \Omega$, as long as we consider the appropriate (and natural) σ-algebra.

Lemma 3.3: Restriction of measurable functions

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

Proof. Recall that $\mathcal{F}_A = \{A \cap B \,:\, B \in \mathcal{F}\}$. Take $C \in \mathcal{G}$, then

$$f|_A^{-1}(C) = \{\omega \in A : f(\omega) \in C\} = f^{-1}(C) \cap A \in \mathcal{F}_A.$$ 😄

At this stage these are the only general properties of measurable function we can consider. However, if the measurable space a function maps to has more structure we can see if this structure also respect the measurability. For example, we will see later in Section 3.2 that for measurable functions $f, g : \Omega \to \mathbb{R}$ their point-wise product and sum are also measurable, as well as many other operations.

3.1.2. Checking for measurability

Given any function $f : \Omega \to E$ between two measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$, when is this measurable? Definition 3.1 tells us that to answer this question we need to check that the preimage of any measurable set is again measurable. But this can be a cumbersome exercise, or even outright impossible, when we do not have an explicit description of the sigma algebra.

This can happen, for example, when $\mathcal{G}$ is generated by some collection of sets $\mathcal{A}$, which is the case for the important Borel σ-algebra.

Fortunately, the definition of measurability works very well with generated σ-algebras. In particular, to show that a function is measurable, it suffices to only consider sets from the generator set $\mathcal{A}$, instead of the entire σ-algebra $\sigma(\mathcal{A})$.

Lemma 3.4

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces such that $\mathcal{G} = \sigma(\mathcal{A})$. Let $f : \Omega \to E$ be a function such that $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$. Then f is $(\mathcal{F}, \mathcal{G})$-measurable.

Proof. See Problem 3.2. 😃

We thus see that at least for generated σ-algebras we do not need to consider the preimage of every measurable set to check for measurability of a function. However, that still requires some check on preimages to conclude that a given function is measurable, which can still be cumbersome. For example, is the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = e^x$, measurable? It would be much better if we have a more familiar criteria that would imply measurability. Continuity is exactly such a criteria.

Proposition 3.5

Every continuous map $f : \mathbb{R}^d \to \mathbb{R}^m$ is $(\mathcal{B}_{\mathbb{R}^d}, \mathcal{B}_{\mathbb{R}^m})$-measurable.

Proof. Recall from analysis that a map $f : \mathbb{R}^d \to \mathbb{R}^m$ is continuous if for every $x \in \mathbb{R}^d$ and $\varepsilon > 0$, there exists an $r = r(x, \varepsilon)$ such that

$$\|f(x) - f(y)\| < \varepsilon \quad \text{for every } y \in B_x(r).$$

The key step for this proof is to show that this is equivalent to the following condition[1]:

$$\text{for every open set } O \subset \mathbb{R}^m \quad f^{-1}(O) \text{ is open.}$$

If this is true then, since the Borel σ-algebra is generated by the open sets, it follows that $f^{-1}(O) \in \mathcal{B}_{\mathbb{R}^d}$ for each open set $O \subset \mathbb{R}^m$. Lemma 3.4 then implies that f is measurable.

So we are left to show the equivalence of the two conditions for continuity. First, assume that f is continuous and take an arbitrary open set $O \subset \mathbb{R}^m$. We need to show that $f^{-1}(O)$ is open, which means that for every $x \in f^{-1}(O)$ we should find an r such that $B_x(r) \subset f^{-1}(O)$. Since O is open, there exists a $\varepsilon > 0$ such that $B_{f(x)}(\varepsilon) \subset O$. Continuity of f now implies the existence of an r such that $\|f(x) - f(y)\| < \varepsilon$ for all $y \in B_x(r)$. But this simply means that $f(y) \in B_{f(x)}(\varepsilon) \subset O$ for every $y \in B_x(r)$, which implies that $B_x(r) \in f^{-1}(O)$.

[1] Actually, the definition we state here using open sets is the general definition for continuous functions in the mathematical field of topology.

Now assume that $f^{-1}(O)$ is open in $\mathbb{R}^d$, for every open set $O \in \mathbb{R}^m$ and take $x \in \mathbb{R}^d$ and $\varepsilon > 0$. Then the ball $B_{f(x)}(\varepsilon)$ is open in $\mathbb{R}^m$, so that by assumption $f^{-1}(B_{f(x)}(\varepsilon))$ is open in $\mathbb{R}^d$. Since $x \in f^{-1}(B_{f(x)}(\varepsilon))$ there now must exist an $r > 0$ such that $B_x(r) \subset f^{-1}(B_{f(x)}(\varepsilon))$. But this then implies that for every $y \in B_x(r)$, $f(y) \in B_{f(x)}(\varepsilon)$, which is equivalent to $\|f(x) - f(y)\| < \varepsilon$. 😄

With this result we gained access to a vast world of measurable functions $f : \mathbb{R}^d \to \mathbb{R}^m$. It should also be noted that the space of measurable functions is larger than that of continuous functions. For example, the indicator functions are measurable but not continuous.

So on the Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ we have a large class of measurable functions. However, when dealing with functions that map to measurable spaces that are not the Borel space, we still need to carefully check if it is measurable. But what if we can simply construct a σ-algebra such that it makes a function measurable?

3.1.3. σ-algebras generated by measurable functions

Suppose we have a function $f : \Omega \to E$ from a set Ω to some measurable space $(E, \mathcal{G})$. If we want to study the function f in the framework of measure theory, we need to turn Ω into a measurable space $(\Omega, \mathcal{F})$ and have f be $(\mathcal{F}, \mathcal{G})$-measurable. The good news is that we can construct a minimal σ-algebra that does the job for us. It can even be done for multiple functions at the same time.

Proposition 3.6

Let $(\Omega_i, \mathcal{F}_i)$, for $i \in I$ be measurable spaces and $(f_i)_{i \in I}$ be a family of functions $f_i : \Omega \to \Omega_i$. Then the smallest σ-algebra on Ω that makes all f_i simultaneously measurable is

$$\sigma(f_i \,:\, i \in I) := \sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)\right).$$

Remark. The collection of sets $\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)$ may be equivalently expressed as

$$\left\{f_i^{-1}(B) : i \in I,\ B \in \mathcal{F}_i\right\}.$$

Proof of Proposition 3.6. First note that by Proposition 2.4, $\sigma(f_i \,:\, i \in I)$ is a σ-algebra. We will show that any σ-algebra that makes each f_i measurable much contain $\sigma(f_i \,:\, i \in I)$. So let $\mathcal{F}$ be such a σ-algebra. Then in particular, for any $i \in I$ and $B \in \mathcal{F}_i$ we have that $f_i^{-1}(B) \in \mathcal{F}$. This implies that

$$\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i) \subseteq \mathcal{F}.$$

Now since $\sigma(f_i \,:\, i \in I)$ is generated by the collection on the left hand side, Lemma 2.5 implies that

$$\sigma(f_i \,:\, i \in I) := \sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)\right) \subset \sigma(\mathcal{F}) = \mathcal{F}.$$ 😄

Similar to Lemma 3.4, when $\mathcal{F}_i = \sigma(\mathcal{A}_i)$ it turns out that to construct $\sigma(f_i : i \in I)$ it suffices to consider only preimages of the generator sets $\mathcal{A}_i$.

Proposition 3.7

Let $(\Omega, \mathcal{F})$ and $(\Omega_i, \mathcal{F}_i)$, for $i \in I$ be measurable spaces such that $\mathcal{F}_i = \sigma(\mathcal{A}_i)$. Let $(f_i)_{i \in I}$ be a family of functions $f_i : \Omega \to \Omega_i$. Then

$$\sigma(f_i : i \in I) = \sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{A}_i)\right).$$

Proof. Let us write $\mathcal{G}_1 = \sigma(f_i : i \in I)$ and $\mathcal{G}_2 = \sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{A}_i)\right)$. From the definition, it is clear that $\mathcal{G}_2 \subseteq \mathcal{G}_1$. Moreover, each f_i is $(\mathcal{G}_2, \mathcal{F}_i)$-measurable by Lemma 3.4. But by Proposition 3.6 $\mathcal{G}_1$ is the smallest σ-algebra that makes all f_i $(\mathcal{G}_1, \mathcal{F}_i)$-measurable and hence $\mathcal{G}_1 \subseteq \mathcal{G}_2$, which implies the result. 😄

We end this section by going back to the product σ-algebra given in Definition 2.6. There is an alternative way to construct it using functions. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces and consider the functions $\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i$, defined by

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

These are called the *canonical projections*. Following Proposition 3.6 we can construct the σ-algebra $\sigma(\pi_1, \pi_2)$ on $\Omega_1 \times \Omega_2$, which makes both canonical projections measurable. It now follows that, see Problem 3.3,

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\pi_1, \pi_2), \tag{3.1}$$

which shows that the original construction of the product σ-algebra is equal to the one using projection maps.

3.1.4. Push forward measure

Given a measure space $(\Omega, \mathcal{F}, \mu)$ and measurable function $f : \Omega \to E$ to a measurable space $(E, \mathcal{G})$ we can construct a measure on $(E, \mathcal{G})$ using f and μ. This measure is called the *push-forward measure*, as it can be thought of a pushing μ to $\mathcal{G}$ via the function f.

Proposition 3.8: Push-forward measure

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(E, \mathcal{G})$ a measurable space and $f : \Omega \to E$ a measurable function. Then the set function $f_{\#}\mu$ defined as

$$f_{\#}\mu(B) = \mu(f^{-1}(B)) \text{ for every } B \in \mathcal{G},$$

is a measure on $(E, \mathcal{G})$ called the *push-forward measure* of μ under f. Moreover, if μ is a probability measure, so if $f_{\#}\mu$.

Proof. See Problem 3.4. ☺

Push-forward measures play an important role in measure theory, and especially in probability theory. For example, they come up for example when we apply a change of variables in integrals. More importantly, we will see in Section 6.1 that the cumulative distribution function of a random variable is actually defined as the push-forward measure of some probability measure $\mathbb{P}$ under a suitable measurable function.

3.2. Measurable functions on the real line

When studying properties of measurable function we could only do a few things for general measurable spaces. So in this section we will focus on a specific measurable space: the real line $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. We will see that most of the natural operations we can apply to function in a point-wise manner, such as addition and multiplication, preserve their measurability. But we will do even better. We will show that taking point-wise limit operations, such as taking a supremum of a family of measurable functions, preserves measurability as well. This makes the class of measurable functions much more powerful then that of continuous functions, as point-wise limits of continuous functions are not guaranteed to be continuous again. All these properties will be useful when we introduce the concept of integration of measurable functions in Chapter 4 and develop limit theorems for integrals in Chapter 7.

To properly study limit operations on measurable functions, that could diverge, we need to have $+\infty$ be a part of the real line (which it is not). So we first extend the real line to include both $+\infty$ and $-\infty$.

3.2.1. Extended real line

We define $\overline{\mathbb{R}} := [-\infty, +\infty]$ as the *extended real line*. We impose the natural ordering on $\overline{\mathbb{R}}$, inherited from $\mathbb{R}$, with the addition that $-\infty < x$ and $x < +\infty$ for all $x \in \mathbb{R}$. The extended real line also has the same operations of addition and multiplications, which are extended to include the two new elements $\pm\infty$:

1. for every $x \in \mathbb{R}$, $x + (+\infty) = (+\infty) + x = +\infty$ and $x + (-\infty) = (-\infty) + x = -\infty$,
2. $(+\infty) + (+\infty) = +\infty$ and $(-\infty) + (-\infty) = -\infty$,
3. for every $x \in (0, +\infty)$, $x(+\infty) = +\infty, \ x(-\infty) = -\infty$,
4. $0(\pm\infty) = (\pm\infty)0 = 0$ and $1/\pm\infty = 0$.

To turn $\overline{\mathbb{R}}$ into a measurable space we extend the Borel σ-algebra to include $\pm\infty$.

Definition 3.9: Extended real line

The Borel σ-algebra $\overline{\mathcal{B}}$ of the extended real line $\overline{\mathbb{R}}$ is defined by

$$\overline{\mathcal{B}} := \{A \cup S \,:\, A \in \mathcal{B}_{\mathbb{R}} \text{ and } S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$$

The following result, whose proof is left as an exercise, relates $\overline{\mathcal{B}}$ to $\mathcal{B}_{\mathbb{R}}$.

Lemma 3.10

The extended Borel σ-algebra $\overline{\mathcal{B}}$ satisfies

$$\mathcal{B}_{\mathbb{R}} = \overline{\mathcal{B}}_{\mathbb{R}} \quad \text{(i.e., the restriction of } \mathcal{B} \text{ to } \mathbb{R} \text{ according to Lemma 2.3).}$$

Moreover, it is generated by sets of the form $[a, \infty]$, with $a \in \mathbb{Q}$ (or $(a, \infty]$, $[-\infty, a)$, $[-\infty, a]$).

Proof. See Problem 3.5 ☺

3.2.2. Basic operations

For the rest of this section, for any set A we will write $\{f \in A\}$ as a shorthand notation for the set $\{\omega \in \Omega \,:\, f(\omega) \in A\}$. In addition, we write $\{f \leq a\}$ for the set $\{f \in (-\infty, a]\}$ and similarly for $<, \geq, >, =$ and $\neq$.

Lemma 3.11

Let $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ be measurable and take $a \in \mathbb{R}$. Then the following sets

$$\{f < a\},\ \{f \leq a\},\ \{f > a\},\ \{f \geq a\},\ \{f = a\} \text{ and } \{f \neq a\},$$

are also measurable.

Proof. Since f is measurable, it follows immediately from Proposition 2.8 and Lemma 3.4 that $\{f < a\}, \{f \leq a\}, \{f > a\}, \{f \geq a\} \in \mathcal{F}$. This then implies the other two claims since $\{f = a\} = \{f \leq a\} \setminus \{f < a\}$ and $\{f \neq a\} = \Omega \setminus \{f = a\}$. ☺

With this result we can now show that point-wise operations on measurable functions again yield measurable functions.

Lemma 3.12

Let $f, g : (\Omega, \mathcal{F}) \to \mathbb{R}$ be measurable. Then the following functions (where operations are always taken point-wise) are measurable as well:

(1) $f + g$,

(2) $f \vee g := \max\{f, g\}$,

(3) $f \wedge g := \min\{f, g\}$,

(4) fg, and

(5) f/g if $g \neq 0$ on Ω.

Proof. We will prove (2) and (4). The other parts are left as an exercise, see Problem 3.6.

(2) We first note that the sets $\{f > g\}$, $\{g > f\}$, $\{f = g\}$ are measurable. This follows from Lemma 3.11 and the fact that

$$\{f > g\} = \bigcup_{q \in \mathbb{Q}} \{f \geq q\} \cap \{g < q\}, \qquad \{g > f\} = \bigcup_{q \in \mathbb{Q}} \{g \geq q\} \cap \{f < q\}.$$

Next, we observe that for any set $A \subset \mathbb{R}$

$$(f \vee g)^{-1}(A) = \big(f^{-1}(A) \cap \{f > g\}\big) \cup \big(g^{-1}(A) \cap \{g > f\}\big) \cup \{f = g\},$$

which implies that $(f \vee g)^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}_{\mathbb{R}}$.

Lemma 3.10 $\overline{\mathcal{B}}$ is generated by the sets $[a, \infty]$, for $a \in \mathbb{Q}$. Hence, by Lemma 3.4 it suffices to show that

$$(fg)^{-1}([a, \infty]) = \{\omega \in \Omega \,:\, f(\omega)g(\omega) \in [a, \infty]\} \in \mathcal{F}.$$

(4) This proof requires several steps, so please bear with us. We first write

$$\{fg \in (-\infty, t]\} = \{fg \in (-\infty, t \wedge 0)\} \cup \{fg = 0\} \cup \{fg \in (0, t \vee 0]\},$$

were we will disregard the set $\{fg = 0\}$ if $t < 0$. Our goal is to show that each of these three sets is measurable which will then imply the result.

First note $\{fg = 0\} = \{f = 0\} \cup \{g = 0\} \in \mathcal{F}$ by Lemma 3.11.

Now assume that $t > 0$ so that $\{fg \in (0, t \vee 0]\} \neq \emptyset$. Then

$$\{fg \in (0, t \vee 0]\} = \bigcup_{q \in \mathbb{Q}_{>0}} \{f \in (0, q]\} \cap \{g \in (0, t/q]\}.$$

Since for any $x > 0$, $(0, x) = (-\infty, x] \setminus (-\infty, 0] \in \mathcal{B}_{\mathbb{R}}$ and the union above is over a countable number of elements ($\mathbb{Q}$ is countable) it follows that $\{fg \in (0, t \vee 0]\} \in \mathcal{F}$.

We are thus left to show that $\{fg \in (-\infty, t \wedge 0)\}$ is measurable. First, we observe that

$$\{fg \in (-\infty, t \wedge 0)\} = \bigcup_{q \in \mathbb{Q}_{>0}} \{fg \in (-\infty, -q)\},$$

and hence it suffices to show that $\{fg \in (-\infty, -q)\}$ is measurable for any $q \in \mathbb{Q}_{>0}$. To achieve this we further split this event as follows:

$$\left(\{fg \in (-\infty, -q)\} \cap \{f < 0\} \cap \{g > 0\}\right) \cup \left(\{fg \in (-\infty, -q)\} \cap \{f > 0\} \cap \{g < 0\}\right),$$

and observe that due to the symmetry on the right-hand side, it is enough to show that $\{fg \in (-\infty, -q)\} \cap \{f < 0\} \cap \{g > 0\}$ is measurable. For this, we note that

$$\{fg \in (-\infty, -q)\} \cap \{f < 0\} \cap \{g > 0\} = \bigcup_{p \in \mathbb{Q}_{>0}} \{f \in (-\infty, -p)\} \cap \{g \in (0, q/p)\}.$$

Since this is a countable union of measurable sets, it is indeed measurable, and thus so is $\{fg \in (-\infty, t \wedge 0)\}$. This concludes the proof of 4. 😄

3.2.3. Limit operations

In addition to the fact that most of the obvious point-wise operations on measurable functions yields another measurable function, it turns out that this also holds for limit operations.

Lemma 3.13

Let $(f_n)_{n\geq 1}$ be a family of measurable functions from $(\Omega, \mathcal{F})$ to $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$. Then the following functions are also measurable (where again operations are taken point wise):

1. $\sup_{n\geq 1} f_n$,
2. $\inf_{n\geq 1} f_n$,
3. $\limsup_{n\to\infty} f_n$, and
4. $\liminf_{n\to\infty} f_n$.

Moreover, if the point-wise limit $\lim_{n\to\infty} f_n$ exists it is also measurable.

Proof. We will prove 1 and leave the other parts as an exercise, see Problem 3.7.

To this end, we will show that for any $x \in \mathbb{R}$

$$\left\{\sup_{n\geq 1} f_n > x\right\} = \bigcup_{n\geq 1} \{f_n > x\}. \tag{3.2}$$

Note that if this holds then $\{\sup_{n\geq 1} f_n > x\} \in \mathcal{F}$ since each set $\{f_n > x\}$ is measurable by Lemma 3.11 and hence $\sup_{n\geq 1} f_n$ is a measurable function (check this yourself, see Problem 3.7).

Since $x < f_n(\omega) \leq \sup_{n\geq 1} f_n(\omega)$ holds for any ω we get the inclusion $\supset$ for the above two sets. For the other inclusion $\subset$ we will argue by contradiction. Suppose that $f_n(\omega) \leq x$ for all $n \geq 1$, then also $\sup_{n\geq 1} f_n(\omega) \leq x$. This implies that

$$\left\{\sup_{n\geq 1} f_n \leq x\right\} \supset \bigcap_{n\geq 1} \{f_n \leq x\},$$

where each side is the complement of the sets in (3.2). 😄

Although the proof makes the content of Lemma 3.13 look rather trivial, it is actually very important. In particular, it shows the power of the class of measurable functions as it is stable under point-wise limit operations. In contrast, the class of continuous functions is not.

Example 3.2 (Point-wise limits of continuous functions are not continuous). Consider the sequence of functions $(f_n)_{n\geq 1}$ defined by $f_n(x) = \arctan(xn)$. Each f_n is clearly continuous. So let us consider the point-wise limit $f(x) = \lim_{n\to\infty} f_n(x)$. For any $x > 0$ we have that

$$f(-x) = \lim_{n\to\infty} \arctan(-xn) = -\frac{\pi}{2},$$

while

$$f(x) = \lim_{n\to\infty} \arctan(xn) = \frac{\pi}{2}.$$

Moreover, $f(0) = \arctan(0) = 0$. We thus conclude that the point-wise limit of f_n is given by

$$f(x) = \begin{cases} -\frac{\pi}{2} & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ \frac{\pi}{2} & \text{if } x > 0, \end{cases}$$

which is clearly not continuous. However, by Lemma 3.13 it is measurable.

The fact that point-wise limits of continuous functions are not necessary continuous is the reason why one has to be careful when, for example, exchanging limits and integration. Here the notion of uniform continuity is often needed. In contrast, as we will see later, this is not an issue for measurable functions and once we have defined the notion of integration of these functions we obtain a powerful set of limit results for such integrals.

3.3. Problems

Problem 3.1.

(a) Let $(E, \mathcal{G})$ be a measurable space and $f : \Omega \to E$ be constant, i.e. there exists $e \in E$ such that $f(w) = e$ for all $\omega \in \Omega$. Show that f is measurable with respect to the trivial σ-algebra on Ω.

Now suppose that the measurable space $(E, \mathcal{G})$ has the following property: for any $x, y \in E$ there exist $A, B \in \mathcal{G}$ with $x \in A$, $y \in B$ and $A \cap B = \emptyset$.

(b) Suppose $f : \Omega \to E$ is measurable with respect to the trivial σ-algebra on Ω. Show that f is constant.

(c) Construct an example of a function $f : \Omega \to (E, \mathcal{G})$ for some measurable space $(E, \mathcal{G})$ that is measurable with respect to the trivial σ-algebra but is not constant.

Problem 3.2. In this problem you will prove Lemma 3.4. For this we have to show that $f^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{G}$. We will do this by proving the collection of set B that satisfy this property is equal to $\mathcal{G}$.

(a) Consider the following collection of subsets:

$$\mathcal{H} := \{B \subset \mathcal{G} \,:\, f^{-1}(B) \in \mathcal{F}\}.$$

Show that $\mathcal{H}$ is a σ-algebra on E.

(b) Use Lemma 2.5 to show that $\mathcal{G} \subseteq \mathcal{H}$ and use this to finish the proof.

Problem 3.3 (Equivalence of product σ-algebra)**.** Prove equation (3.1).

Problem 3.4 (Push-forward measure)**.** Prove Proposition 3.8.

Problem 3.5. Prove Lemma 3.10.

Problem 3.6. The goal of this problem is to prove points 1, 3, and 5 of Lemma 3.12.

(a) Based on the proofs of points 2 and 4 of Lemma 3.12, explain the general idea behind the proof.

(b) Prove point 1 of Lemma 3.12.

(c) Prove that for any $a \in \mathbb{R}$, the constant function $f : \Omega \to \mathbb{R}$, $f(\omega) = a$ for all $\omega \in \mathbb{R}$ is measurable.

(d) Prove point 3 of Lemma 3.12 (You can do this directly or use the above result and point 2 of the lemma).

(e) Prove that if $g : \Omega \to \mathbb{R}$ is measurable and $g(\omega) \neq 0$ for all $\omega \in \Omega$, then $1/g$ is measurable.

(f) Prove point 5 of Lemma 3.12.

Problem 3.7.

(a) Conclude that if (3.2) holds then $\sup_{n\geq 1} f_n$ is a measurable function.

(b) Prove points 2-4 of Lemma 3.13. [Hint: What is the relation between inf and sup and what is the definition of $\liminf$, $\limsup$ in terms of the infimum and supremum?]

Problem 3.8 (Truncation)**.** Let $f : \Omega \to \overline{\mathbb{R}}$ be a function. For a real number $M > 0$, we define the *truncation* of f to be the function $f_M : \Omega \to \mathbb{R}$ defined by

$$f_M(\omega) := \max\{-M, \min\{f(\omega), M\}\} = \begin{cases} M & \text{if } f(\omega) \geq M, \\ f(\omega) & \text{otherwise}, \\ -M & \text{if } f(\omega) < -M. \end{cases}$$

(a) Show that if f is $\mathcal{F}$-measurable, then f_M is also $\mathcal{F}$-measurable.

(b) Now suppose that f_M is $\mathcal{F}$-measurable for all $M > 0$, show that f is $\mathcal{F}$-measurable.

4. The Lebesgue Integral

We have now arrived at arguably the core aspect of Measure Theory: The Lebesgue integral. Unlike the Riemann integral, the Lebesgue integral can be constructed on any measure space $(\Omega, \mathcal{F}, \mu)$ and facilitates many powerful limit results. The construction will be done in multiple steps, starting with simple functions.

4.1. The integral of a simple function

Definition 4.1

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 $A_1, \dots, A_N \in \mathcal{F}$ and constants $a_1, \dots, a_N \in \mathbb{R}$.

Sometimes, the definition of a simple function will not require the sets to be disjoint. When this is the case, it is referred to as a *standard representation* of a simple functions. But this distinction does not matter.

Lemma 4.2

Let $f : \Omega \to \mathbb{R}$ be a measurable function that attains only finitely many values, i.e. $f(\Omega) = \{a_1, \dots, a_m\}$ with $m < \infty$. Then there exist a finite collection of measurable and mutually disjoint sets $A_1, \dots, A_m$ such that

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

In particular, f is simple.

Proof. See Problem 4.1. 😄

As a consequence of Lemma 4.2 any function of the form

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

with $a_i \in \mathbb{R}$ and $A_1, \dots, A_N$ measurable but necessarily disjoint sets, is simple.

We can now define the Lebesgue integral of a simple function.

Definition 4.3

Let $f : \Omega \to \mathbb{R}_+$ be a non-negative simple function of the form

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

Then we define the μ-*integral* of f by

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f(\omega) \, \mu(\mathrm{d}\omega) := \sum_{i=1}^{N} a_i \mu(A_i).$$

Although this seems like a reasonable definition for an integral of a simple function, one has to be a bit careful. The potential problem is that the value of the integral is defined using the set A_i and values a_i in the representation of the simple function f, which might not be unique. Luckily it turns out that the value of their sum is.

Lemma 4.4

Let $f : \Omega \to \mathbb{R}_+$ be a non-negative simple function such that

$$\sum_{i=1}^{N} a_i \mathbf{1}_{A_i} = f = \sum_{j=1}^{M} b_j \mathbf{1}_{B_j},$$

for two collections $a_i, b_j \in \mathbb{R}$ and two families of mutually disjoint measurable sets A_i, B_j. Then

$$\sum_{i=1}^{N} a_i \mu(A_i) = \sum_{j=1}^{M} b_j \mu(B_j).$$

Proof. See Problem 4.2 😄

This lemma shows that the Lebesgue integral for simple functions given in Definition 4.3 is well-defined. With that, we have made the first important step toward the definition of the Lebesgue integral for general functions. Before we can continue, we will show that any positive measurable function $f : \Omega \to \mathbb{R}$ can be approximated by simple functions.

Lemma 4.5

Let $f, g : \Omega \to \mathbb{R}_+$ be two non-negative simple functions. Then $f + g$ is again a non-negative simple function and moreover,

$$\int_\Omega f + g \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu$$

Proof. 😃

4.2. Approximation by simple functions

Proposition 4.6: Approximation by simple functions

Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to [0, \infty]$ a measurable function. Then there exist a sequence $(f_n)_{n\geq 1}$ of positive simple functions such that for any $\omega \in \Omega$

$$\lim_{n\to\infty} f_n(\omega) = f(\omega).$$

Moreover, the functions f_n form a point-wise non-decreasing sequence and hence $f(\omega) = \sup_{n\geq 1} f_n(\omega)$.

Proof. See Problem 4.3 😃

We will often use the notation $[f]_n$ to denote the simple functions that approximate a given function f.

4.3. The Lebesgue integral of nonnegative functions

We now extend the μ-integral from non-negative simple functions to arbitrary non-negative $\mathcal{F}$-measurable functions.

Definition 4.7

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

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f(\omega) \, \mu(\mathrm{d}\omega) := \sup \left\{ \int_\Omega g \, \mathrm{d}\mu : \; g \text{ simple}, \; 0 \leq g \leq f \right\}.$$

The function f is said to be *μ-integrable* if its μ-integral is finite.

For a measurable set $A \in \mathcal{F}$, we use the following notation and definition for integration of f over the set A
$$\int_A f \, \mathrm{d}\mu := \int_\Omega \mathbf{1}_A f \, \mathrm{d}\mu.$$
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
$$\int_A f_A \, \mathrm{d}\mu_A = \int_A f \, \mathrm{d}\mu.$$
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
$$\int_A f \, \mathrm{d}\mu = \int_A f_A \, d\mu_A.$$

The following lemma summarize some basic properties of the Lebesgue integral for non-negative functions.

Lemma 4.8: Properties of the Lebesgue integral of non-negative functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f, g : \Omega \to \mathbb{R}_+$ two non-negative, measurable functions and $\alpha \geq 0$ be a constant. The the following holds:

1. (absolute continuity) If $B \in \mathcal{F}$ satisfies $\mu(B) = 0$, then
$$\int_B f \, \mathrm{d}\mu = 0.$$
2. (monotonicity) If $f \leq g$, then
$$\int_\Omega f \, \mathrm{d}\mu \leq \int_\Omega g \, \mathrm{d}\mu.$$
3. (homogeneity)
$$\alpha \int_\Omega f \, \mathrm{d}\mu = \int_\Omega (\alpha f) \, \mathrm{d}\mu.$$

Proof. See Problem 4.5 😄

It should be noted that one key property seems to be missing from the above lemma, namely
$$\int_\Omega f + g \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu.$$
We know that the Riemann integral is linear and hence, if the aim of the Lebesgue integral is to extend it, it should at the very least also be linear. The main issue is that it is difficult to proof this directly using Definition 4.7, due to the supremum. However, recall that we did have linearity for the Lebesgue integral of simple functions, see Lemma 4.5. So how do we use this to

show that also the integral of non-negative functions is linear? This is where we first make use of the approximation of non-negative functions by simple functions. Here is how we envision this to go.

$$\begin{aligned}\int_\Omega (f+g)\,\mathrm{d}\mu &= \int_\Omega \lim_{n\to\infty} ([f]_n + [g]_n)\,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_\Omega [f]_n + [g]_n\,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_\Omega [f]_n\,\mathrm{d}\mu + \lim_{n\to\infty} \int_\Omega [g]_n\,\mathrm{d}\mu \\ &= \int_\Omega \lim_{n\to\infty} [f]_n\,\mathrm{d}\mu + \int_\Omega \lim_{n\to\infty} [g]_n\,\mathrm{d}\mu \\ &= \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu.\end{aligned}$$

The problem with this argument is that in both the second and fourth line, we exchanged the limit and the integral. This is something that is not generally possible, at least for Riemann integration, and needs a proof. For this we will use one of the classical limit theorems for Lebesgue integrals: the Monotone Convergence Theorem.

4.4. The monotone convergence theorem

This first convergence result tells us essentially that the point-wise limit of monotone sequences of μ-integrable functions is again μ-integrable, highlighting the difference with Riemann integration.

Theorem 4.9: Monotone convergence theorem I

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f_n \colon \Omega \to [0, +\infty]$, $n \in \mathbb{N}$, be a sequence of nonnegative $(\mathcal{F}, \mathcal{B}_{[0,+\infty]})$-measurable functions, such that $f_n(\omega) \leq f_{n+1}(\omega)$ 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 $(\mathcal{F}, \mathcal{B}_{[0,+\infty]})$-measurable and

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

Proof. From the monotonicity of the integral, we immediately conclude that

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

Hence, we are left to show that

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

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

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

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

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

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

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

$$G_n^i := \Big\{ x \in A_i : \ f_n(x) \geq a_i - \delta \Big\}.$$

Then, because $f_n \leq f_{n+1}$, 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

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

Hence, by the continuity from below of measures

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

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

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

we find that

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

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

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

😄

4.5. Additivity of the Lebesgue integral of nonnegative functions

Armed with monotone convergence, we can now proof that the Lebesgue integral of non-negative functions is linear.

Lemma 4.10: Additivity of the Lebesgue integral of nonnegative functions

Let f, g be two non-negative $(\mathcal{F}, \mathcal{B}_{[0,+\infty]})$-measurable functions. Then,

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

Proof. For simple functions, the additivity of the integral follows from Lemma 4.5. Therefore,

$$\int_\Omega ([f]_n + [g]_n)\,\mathrm{d}\mu = \int_\Omega [f]_n\,\mathrm{d}\mu + \int_\Omega [g]_n\,\mathrm{d}\mu \qquad \text{for every } n \in \mathbb{N}.$$

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 point-wise to $(f+g)$. Hence, by the monotone convergence theorem (Theorem 4.9),

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

On the other hand, by linearity of limits we know that

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

Since both $[f]_n$ and $[g]_n$ are increasing and converge point-wise to f and g, respectively, monotone convergence implies that

$$\lim_{n\to\infty} \int_\Omega [f]_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu, \text{ and } \lim_{n\to\infty} \int_\Omega [g]_n\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu.$$

Therefore, we conclude that

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

😄

4.6. Integrable functions

The final step we need to take is to define the integral of functions f that are not necessarily non-negative. We can only do this if the integral of $|f|$ is finite.

Definition 4.11

A $(\mathcal{F}, \mathcal{B}_{\mathbb{R}})$-measurable function $f : \Omega \to \mathbb{R}$ is *μ-integrable* if

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

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

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

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

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

$$\int_\Omega f \,\mathrm{d}\mu := \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu.$$

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

As in the case for non-negative measurable functions, we have the following properties for μ-integrable functions.

Proposition 4.12

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

1. (absolute continuity) If $B \in \mathcal{F}$ satisfies $\mu(B) = 0$, then
$$\int_B f \,\mathrm{d}\mu = 0.$$
2. (monotonicity) If $f \leq g$ μ-a.e., then
$$\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu.$$
3. (homogeneity)
$$\alpha \int_\Omega f \,\mathrm{d}\mu = \int_\Omega (\alpha f) \,\mathrm{d}\mu.$$
4. (additivity)
$$\int_\Omega (f + g) \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu.$$

Proof. See Problem 4.7 😊

4.7. Riemann vs Lebesgue integration

A fundamental fact about the Lebesgue integral is its relationship with the Riemann integral, which allows us to make use of the integration techniques we know from Calculus and Analysis to compute the Lebesgue integral of a Lebesgue integrable function.

We state an important result, which we will not prove, but will be essential for computing integrals (cf. Appendix B). The first part of the result provides a full characterization of Riemann-integrable functions, while the second provides the means to compute Lebesgue integrals.

Theorem 4.13: Riemann vs Lebesgue

A bounded function $f\colon [a,b] \to \mathbb{R}$ on a compact set $[a,b] \subset \mathbb{R}$ is Riemann integrable if and only if it is continuous λ-almost everywhere.

If a bounded function $f\colon [a,b] \to \mathbb{R}$ is Riemann integrable, then f is $\mathcal{L}$-measurable and λ-integrable. Moreover,

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

where the left-hand side denotes the Riemann integral of f.

Example 4.1. Let us determine the value $\displaystyle\int_{\mathbb{R}} \frac{1}{x^2+1}\,\lambda(\mathrm{d}x)$.

To do so, we set $g(x) := \frac{1}{x^2+1} \geq 0$ and let $g_n := g\mathbf{1}_{[-n,n]}$. Then clearly, g_n is monotone and $g_n \to g$ point-wise. Thus, by monotone convergence we have that

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

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

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

where the second equality follows from the fact that g is continuous on the compact set $[-n,n]$ and from Theorem 4.13. Hence,

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

thus implying that g is λ-integrable.

Remark. The main take-away is that when considering a measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ that is Riemann integrable, the Lebesgue integral is simply the same as the Riemann integral. The difference is that the Lebesgue integral is applicable to any measurable space and a much larger class of functions.

4.8. Change of variables formula

A key tool for computing Riemann integrals was the so-called change of variables formula

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(y)\,\mathrm{d}y.$$

This allowed one to pick a different parametrization of the function in terms of its variables to compute the integral. As you might expect by now, we also have this for the Lebesgue integral. Here we encounter the push-forward measure, see Proposition 3.8.

Proposition 4.14

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(E, \mathcal{G})$ be a measurable space. Further, let $f\colon \Omega \to E$ and $h\colon E \to [0, +\infty]$ be $(\mathcal{F}, \mathcal{G})$- and $(\mathcal{G}, \mathcal{B}_{[0,+\infty]})$-measurable maps respectively. Then,

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

In particular, $h \circ f$ is integrable with respect to μ if and only if h is integrable with respect to $f_\# \mu$.

Proof. The idea of the proof is to first prove the statement for the case where h is simple and nonnegative. Then we use the approximation result, Proposition 4.6, and monotone convergence to prove the statement for general non-negative functions h. Finally, we use linearity of the integral to prove it for general functions.

Following this strategy, we first prove the 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

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

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 $h \circ f$. Thus, by the monotone convergence theorem,

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

Finally, let $h : \Omega \to \mathbb{R}$ be a general measurable function and denote by h^+ and h^- its positive and negative part, respectively. Then, since each of these is non-negative and $h \circ f = h^+ \circ f - h^- \circ f$, we get

$$\begin{aligned} \int_\Omega h \circ f \, \mathrm{d}\mu &= \int_\Omega h^+ \circ f \, \mathrm{d}\mu - \int_\Omega h^- \circ f \, \mathrm{d}\mu \\ &= \int_E h^+ \, \mathrm{d}(f_\# \mu) - \int_E h^- \, \mathrm{d}(f_\# \mu) = \int_E h \, \mathrm{d}(f_\# \mu). \end{aligned}$$ 😄

Remark (General proof strategy). The architecture of the above proof is prototypical for many statements concerning Lebesgue integration. That is, we first proof it for simple non-negative functions. Then we extend this to general non-negative functions using monotone convergence and finally to general functions using linearity. Within this scheme, proving the required result for simple functions will be the main step, and hence the one that requires the most work.

4.9. Problems

Problem 4.1. Here you will prove Lemma 4.2.

(a) Construct a collection of sets $A_1, \dots, A_m$ such that $\omega \in A_i \iff f(\omega) = a_i$.

(b) Prove that these are measurable and mutually disjoint sets.

(c) Conclude that
$$f = \sum_{i=1}^m a_i \mathbf{1}_{A_i}.$$

Problem 4.2. Prove Lemma 4.4.

Problem 4.3. In this exercise we will prove Proposition 4.6

(a) Fix $n \geq 1$, set $N_n = n2^n$ and define the sets
$$A_k^n = \begin{cases} \{k2^{-n} \leq f < (k+1)2^{-n}\} & \text{for } k = 0, 1, \dots, N_n - 1, \\ \{n \leq f < +\infty\} & \text{for } k = N_n, \end{cases}$$

Prove that these are measurable and mutually disjoint.

We now define the following sequence of simple functions:
$$f_n = 2^n \mathbf{1}_{\{f=+\infty\}} + \sum_{k=0}^{N_n} k\, 2^{-n} \mathbf{1}_{A_k^n}.$$

(b) Prove that for any $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$. [Hint: consider the case $f(\omega) = \infty$ and $f(\omega) < \infty$ separately.]

The final thing to do is to show that f_n is a point-wise monotone sequence, i.e. for any $\omega \in \Omega$, $f_n(\omega) \leq f_{n+1}(\omega)$ holds for all $n \geq 1$.

(e) Prove that $f_n(\omega) \leq f_{n+1}(\omega)$ holds for all ω such that $f(\omega) = +\infty$.

(f) Assume now that $f(\omega) < +\infty$ and that $\omega \in A_k^n$ for some $k < n2^n$. Prove that $f_n(\omega) \leq f_{n+1}(\omega)$. [Hint: split the interval A_k^n into two parts of equal length and consider the case that ω is in one of these separately.]

(g) Now consider the case that $\omega \in A_k^n$ with $k = n2^n$. Show that $f_n(\omega) \leq f_{n+1}(\omega)$.

Problem 4.4. Consider the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$, where μ is the counting measure on $\mathbb{N}$. Show that for any function $f : \mathbb{N} \to [0, +\infty]$,

$$\int_{\mathbb{N}} f \, d\mu = \sum_{n \geq 1} f(n).$$

Problem 4.5. Prove Proposition 4.8. [Hint: first prove it for simple functions and then use the definition of the integral.]

Problem 4.6. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f\colon (\Omega, \mathcal{F}) \to ([0, +\infty), \mathcal{B}_{[0,+\infty)})$ be a nonnegative measurable function. Show that

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

Problem 4.7. Proof Proposition 4.12.

Problem 4.8. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and suppose that f is a non-negative $(\mathcal{F}, \mathcal{B})$-measurable function such that $\int_{\Omega} f \, d\mu = 1$. Define the set function $\nu_f\colon \mathcal{F} \to [0, +\infty]$ by

$$\nu_f(A) := \int_A f \, d\mu, \qquad \forall A \in \mathcal{F}.$$

(a) Show that ν_f is a probability measure on $(\Omega, \mathcal{F})$.

(b) Show that for all nonnegative $(\mathcal{F}, \mathcal{B}_{[0,+\infty]})$-measurable functions $g\colon \Omega \to [0, +\infty]$,

$$\int_{\Omega} g \, d\nu_f = \int_{\Omega} gf \, d\mu.$$

Hint: Start with simple functions and then approximate.

(c) Show that g is ν_f-integrable if and only if gf is μ-integrable, in which case

$$\int_{\Omega} g \, d\nu_f = \int_{\Omega} gf \, d\mu.$$

Problem 4.9. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and μ be a finite measure. Show that an $(\mathcal{F}, \mathcal{B}_{\mathbb{R}})$-measurable function $f\colon \Omega \to \mathbb{R}$ is integrable if and only if

$$\lim_{n\to\infty} \int_\Omega |f| \, \mathbb{1}_{\{|f|\geq n\}} \, d\mu = 0.$$

Problem 4.10 (Continuity property of the integral)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f be μ-integrable. Show that for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\int_A |f| \, d\mu \leq \varepsilon \quad \text{for all} \quad A \in \mathcal{F} \quad \text{with} \quad \mu(A) < \delta.$$

Hint: If f is bounded, things are easy, so consider the set where $|f|$ is larger than some value and where $|f|$ is smaller than such value.

5. Product spaces and Lebesgue integration

In this chapter we will discuss how to turn two measure space into a new measure space by taking a *product*. We will then consider integration on this product space and relate it to integration over the individual spaces. In particular, we will discuss several theorems that basically say that integrating a function $f(x, y)$ with respect to (x, y) is the same as first integrating with respect to x and then to y, or the other way around. In other words, the order does not matter.

5.1. Product measures and independent random variables

Recall that in Chapter 2 we defined the notion of the product of two measurable spaces, see Definition 2.6. Later, in Chapter 3 we gave a alternative definition using the σ-algebra generated by projections, see (3.1).

Given that we can take two measurable space $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ and turn them into a new space, by taking a "product", can we do the same for two measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$? Here of course we would like to have our product measure relate to measures μ_1 and μ_2. One natural case would be that $\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B)$, for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. However, it is not obvious how to explicitly define a measure on the entire product sigma-algebra such that this property holds.

Luckily, the following theorem states that such a measure does exist. Moreover, if we assume the spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ to be σ-finite the product measure is actually the unique measure with this property.

Theorem 5.1: Construction of product measure space

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 complete measure $\mu_1 \otimes \mu_2$ on $(\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \otimes \mathcal{F}_2})$ 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 $B \in \mathcal{F}_2$.

Proof. Recall that $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$, where $\mathcal{F}_1 \times \mathcal{F}_2 := \{A \times B \,:\, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$.

The idea of the proof is to use the Carathéodory Extension theorem (Theorem 2.18). For this we need to show two things:

1. The collection $\mathcal{F}_1 \times \mathcal{F}_2$ is a semi-algebra, and
2. the set function $\mu : \mathcal{F}_1 \times \mathcal{F}_2 \to [0, \infty]$ defined by $\mu(A \times B) = \mu_1(A)\mu_2(B)$ is a pre-measure on $\mathcal{F}_1 \times \mathcal{F}_2$.

Once we have proven these two properties, the result follows immediately from Theorem 2.18.

The first part is left as an exercise, see Problem 5.1. So let us focus on the second part. It is clear that $\mu(\emptyset) = 0$ and thus we only need to show that μ is σ-additive on $\mathcal{F}_1 \times \mathcal{F}_2$. To this end, let $(A_i \times B_i)_{i \geq 1}$ be a family of disjoint sets in $\mathcal{F}_1 \times \mathcal{F}_2$ such that $\bigcup_{i \geq 1} A_i \times B_i = A \times B \in \mathcal{F}_1 \times \mathcal{F}_2$. We then have to show that

$$\mu(A \times B) = \sum_{i \geq 1} \mu_1(A_i)\mu_2(B_i).$$

For this we will consider the indicator functions of the sets and use monotone convergence. First observe that

$$\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A \times B} = \sum_{i \geq 1} \mathbb{1}_{A_i \times B_i} = \sum_{i \geq 1} \mathbb{1}_{A_i} \mathbb{1}_{B_i}. \tag{5.1}$$

Now fix some $\omega_1 \in \Omega_1$ and define the function $f : \Omega_2 \to \{0, 1\}$ as $f(\omega) = \mathbb{1}_A(\omega_1)\mathbb{1}_B(\omega)$. Then f is measurable and integrable and for the integral we have

$$\int_{\Omega_2} f \, \mathrm{d}\mu_2 = \mathbb{1}_A(\omega_1)\mu_2(B).$$

Since the function $g(\omega) = \mathbb{1}_A(\omega)\mu_2(B)$ is again integrable we get that

$$\int_{\Omega_1} g \, \mathrm{d}\mu_1 = \mu_1(A)\mu_2(B).$$

We thus conclude that

$$\mu(A \times B) = \int_{\Omega_1} \left(\int_{\Omega_2} \mathbb{1}_A \mathbb{1}_B \right) \mathrm{d}\mu_1.$$

We now need to show that by first integrating the right hand side in (5.1) with respect to μ_2 and then μ_1 gives us $\sum_{i \geq 1} \mu_1(A_i)\mu_2(B_i)$.

For this we again fix a $\omega_1 \in \Omega_1$ and now consider the function $f_n : \Omega_2 \to \{0, 1\}$, defined by $f_n(\omega) = \sum_{i=1}^n \mathbb{1}_{A_i \times B_i}(\omega_1, \omega)$ which is a measurable and integrable. Moreover, $(f_n)_{n \to \infty}$ is a monotone sequence of functions whose point-wise limit is

$$\sum_{i \geq 1} \mathbb{1}_{A_i}(\omega_1)\mathbb{1}_{B_i}(\omega).$$

Therefore, applying monotone convergence we get

$$\int_{\Omega_2} \sum_{i \geq 1} \mathbb{1}_{A_i}(\omega_1)\mathbb{1}_{B_i} \, \mathrm{d}\mu_2 = \sum_{i \geq 1} \int_{\Omega_2} \mathbb{1}_{A_i}(\omega_1)\mathbb{1}_{B_i} \, \mathrm{d}\mu_2 = \sum_{i \geq 1} \mathbb{1}_{A_i}(\omega_1)\mu_2(B_i).$$

For the final step we note that

$$\int_{\Omega_2} \sum_{i \geq 1} \mathbb{1}_{A_i}(\omega_1)\mathbb{1}_{B_i} \, \mathrm{d}\mu_2 = \int_{\Omega_2} \mathbb{1}_A(\omega_1)\mathbb{1}_B \, \mathrm{d}\mu_2 = \mathbb{1}_A(\omega_1)\mu_2(B),$$

which is measurable and integrable. Thus integrating this function with respect to μ_1 and using monotone convergence implies that

$$\int_{\Omega_1} \left(\int_{\Omega_2} \sum_{i \geq 1} \mathbb{1}_{A_i} \mathbb{1}_{B_i} \, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 = \sum_{i \geq 1} \mu_1(A_i)\mu_2(B_i)$$

which finishes the proof. 😄

Suppose now we have two measures spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ for $i = 1, 2$ such that $\mathcal{F}_i = \sigma(\mathcal{A}_i)$, i.e. the σ-algebras are generated by some sets $\mathcal{A}_i$. Then one might expect that the product $\mathcal{F}_1 \otimes \mathcal{F}_2$ is generated by $\mathcal{A}_1 \times \mathcal{A}_2$. This turns out to indeed be true

Lemma 5.2

Let $(\Omega_i, \mathcal{F}_i)$, for $i = 1, 2$, be two measurable spaces, such that $\mathcal{F}_i = \sigma(\mathcal{A}_i)$ for some generator sets $\mathcal{A}_1, \mathcal{A}_2$. Suppose further there exists sequences $(A_n)_{n \geq 1}$ and $(B_m)_{m \geq 1}$ with $A_n \in \mathcal{A}_1$, $B_m \in \mathcal{A}_2$ and $\bigcup_{n \geq 1} A_n = \Omega_1$ and $\bigcup_{m \geq 1} B_m = \Omega_2$. Then

$$\sigma(\mathcal{A}_1 \times \mathcal{A}_2) = \sigma(\mathcal{F}_1 \times \mathcal{F}_2) = \mathcal{F}_1 \otimes \mathcal{F}_2$$

Proof. See Problem 5.2. 😄

This lemma is very useful because it allows us to analyze product measure $\mu_1 \otimes \mu_2$ by only considering their values on sets from $\mathcal{A}_1 \times \mathcal{A}_2$. This is what we were already used to for general measures.

5.2. Fubini's Theorem and Tonelli's Theorem

Now that we can construct product measure spaces we take a closer look at integration with respect to the product measure, and how it relates to integration with respect to the marginal measures. In particular, 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 on each of the marginal measures, and one may also change the order of integration. This is very similar to what you have seen for Riemann integration. To illustrate that this is not obvious, consider the following example.

Example 5.1. Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be continuous. Then

$$\int_{[0,1]\times[0,1]} f \, \mathrm{d}A = \int_0^1 \int_0^1 f(x,y) \, \mathrm{d}x \, \mathrm{d}y = \int_0^1 \int_0^1 f(x,y) \, \mathrm{d}y \, \mathrm{d}x.$$

In the example above, the left-hand side can either be interpreted as the two-dimensional Riemann integral or the integral against the two-dimensional Lebesgue measure. Since f is continuous, we know (see Theorem 4.13) that these integrals agree. However, the two integrals to the right of this two-dimensional integral are different. Consider the first:

$$\int_0^1 \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

This can be seen as the one-dimensional integral of the function $y \mapsto \int_0^1 f(x,y)\,\mathrm{d}x$. However, it is not clear at all whether this function is measurable. Moreover, why would the one-dimensional Lebesgue integral of this function be the same as the one-dimensional Lebesgue integral of $x \mapsto \int_0^1 f(x,y)\,\mathrm{d}y$? These are exactly the kind of statements the four theorems we will discuss next cover.

Four versions sound excessive, but we will see that each of them has some satisfactory, and some unsatisfactory aspects to them. This will provide at least some justification for the amount of variation we shall cover. We only present the statements of the theorems as the proof is of a technical nature that does not teach us anything useful. The interested student can look up the proofs in [REF].

The first result is the so-caled Fubini's theorem:

Theorem 5.3: 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

$$x \mapsto \int_{\Omega_2} f(x,y)\,\mu_2(\mathrm{d}y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}\mu_1 \otimes \mu_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x,y)\,\mu_2(\mathrm{d}y) \right) \mu_1(\mathrm{d}x)\\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y)\,\mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y).\end{aligned}$$

Checking whether you can apply Fubini's theorem usually comes with (at least) two difficulties. First, 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 some crazy procedure does not construct f, it is 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 5.4: 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

$$x \mapsto \int_{\Omega_2} f(x, y)\, \mu_2(\mathrm{d}y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \,\mathrm{d}\mu_1 \otimes \mu_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\, \mu_2(\mathrm{d}y) \right) \mu_1(\mathrm{d}x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y)\, \mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y).\end{aligned}$$

While both the Fubini Theorem and the Tonelli theorem tell us that when integrating with respect to the product measures, the order does not matter, they are not yet as useful as you would like. For example, both 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?). Instead, the Lebesgue measure is given as the completion of the product measure defined on the completion of the Borel σ-algebra. Luckily, there are versions of Fubini's and Tonelli's theorems for complete measure spaces. Recall that for a given measure space $(\Omega, \mathcal{F}, \mu)$ we denote its completion by $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$.

Theorem 5.5: 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

$$x \mapsto \int_{\Omega_2} f(x, y)\, \mu_2(\mathrm{d}y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \,\mathrm{d}\overline{\mu_1 \otimes \mu_2} &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y)\, \mu_2(\mathrm{d}y) \right) \mu_1(\mathrm{d}x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y)\, \mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y).\end{aligned}$$

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 5.6: 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-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

$$x \mapsto \int_{\Omega_2} f(x, y) \, \mathrm{d}\mu_2(y)$$

is $(\mathcal{F}_1, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}\overline{\mu_1 \otimes \mu_2} &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, \mu_2(\mathrm{d}y) \right) \mu_1(\mathrm{d}x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) \, \mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y).\end{aligned}$$

The takeaway of these four theorems is that if you have to integrate either with respect to the product measures or the completion of the product measures (as is the case for the Lebesgue measure) then, as long as your function is integrable (in the case of Fubini) or measurable (in the case of Tonelli), this can be done by iterative integration and the order does not matter.

5.3. Problems

Problem 5.1. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces and define

$$\mathcal{F}_1 \times \mathcal{F}_2 := \{A \times B \,:\, A \in \mathcal{F}_1,\, B \in \mathcal{F}_2\}.$$

Prove that $\mathcal{F}_1 \times \mathcal{F}_2$ is a semi-algebra (see Definition 2.16).

Problem 5.2. The goal of this exercise it to prove Lemma 5.2.

(a) Show that $\sigma(\mathcal{A}_1 \times \mathcal{A}_2) \subset \mathcal{F}_1 \otimes \mathcal{F}_2$.

(b) Consider the following family of sets

$$\Sigma_1 := \{C \in \mathcal{F}_1 \,:\, C \times B \in \sigma(\mathcal{A}_1 \times \mathcal{A}_2) \, \forall B \in \mathcal{A}_2\}.$$

Prove that this is a σ-algebra on Ω_1.

(c) Use this to prove that $\mathcal{F}_1 \times \mathcal{A}_2 \subset \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$.

(d) Explain how you can argue in a similar way that $\mathcal{A}_1 \times \mathcal{F}_2 \subset \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$.

(e) Finish the proof.

Problem 5.3. Take $d \in \mathbb{N}$ and let $(\Omega_i, \mathcal{F}_i)_{1 \leq i \leq d}$ be σ-finite measurable spaces such that $\mathcal{F}_i = \sigma(A_i)$. Show that

$$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_d = \sigma(\mathcal{A}_1 \times \cdots \times \mathcal{A}_d).$$

Problem 5.4. Let $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ be the d-dimensional Borel space. Prove that the σ-algebra $\mathcal{B}_{\mathbb{R}^d}$ is generated by any of the following family of sets:

(a) $\{(a_1, b_1) \times \cdots \times (a_d, b_d)\}$,

(b) $\{(a_1, b_1] \times \cdots \times (a_d, b_d]\}$,

(c) $\{[a_1, b_1) \times \cdots \times [a_d, b_d)\}$,

(d) $\{(-\infty, a_1] \times \cdots \times (-\infty, a_d]\}$,

(e) $\{(-\infty, a_1) \times \cdots \times (-\infty, a_d)\}$,

(f) $\{[a_1, \infty) \times \cdots \times [a_d, \infty)\}$,

(g) $\{(a_1, \infty) \times \cdots \times (a_d, \infty)\}$,

where $a_i, b_i \in \mathbb{Q}$, or $a_i, b_i \in \mathbb{R}$ for all $i = 1, \ldots, d$.

6. Probability Theory I: The basics

6.1. Random variables and general stochastic objects

We have arrived at the stage where we know enough about measure theory to start our first journey into the realm of probability theory. In this chapter, we will cover the basic definitions of random variables and expectations and show how we can formally prove formulas related to them which you encountered in the course Probability and Modeling.

6.1.1. Definition

In the course Probability and Modeling two types of random variables were defined: discrete and continuous. Recall that a random variable was defined as a function $X : \Omega \to \mathbb{R}$ for some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\{\omega \in \Omega \,:\, X(\omega) \leq x\} \in \mathcal{F} \quad \text{for all } x \in \mathbb{R}.$$

Let us make two observations here. The first is that the set above is simply the preimage $X^{-1}((-\infty, x])$. Secondly, the sets $(-\infty, x]$ generate the Borel σ-algebra. Thus it follows from Lemma 3.4 that X is a measurable function. This is actual the proper way to define a random variable.

Definition 6.1: Random variable

A *random variable* is a measurable function from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the Borel space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

It is important to observe that the definition of a random variable does not make any specific claims on what the probability space should be.

In fact we can actually define, at a much more general level, random elements in any measurable space and put an associated probability measure on this space by a push-forward.

Definition 6.2: Random elements

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{G})$ some measurable space. A *random element* in $(E, \mathcal{G})$ is a measurable map $X : \Omega \to E$. It associated *probability measure* is defined as the push forward $X_{\#}\mathbb{P}$ of $\mathbb{P}$ under X, i.e.

$$\mathbb{P}(X \in A) := \mathbb{P}(X^{-1}(A)) \quad \text{for every } A \in \mathcal{G}.$$

Sometimes we use the term *stochastic* instead of *random*.

With this general definition we can now easily define random vectors, random matrices, random functions and so on. The only thing we need is to start with the appropriate space (vectors, matrices, functions) and turn it into a measurable space by endowing it with a suitable σ-algebra. Here are some examples:

Example 6.1 (Random elements).

(a) A random vector in $\mathbb{R}^d$ is a random element in $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

(b) A random $n \times m$ matrix is a random element in $(\mathbb{R}^n \times \mathbb{R}^m, \mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}^m})$.

While these are somewhat straightforward examples, there are also more involved ones that are important in probability theory.

Example 6.2 (Stochastic processes). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(S, \mathcal{S})$ a measurable space and T some index set. Then we denote by S^T the set of all functions $f : T \to S$. For any $t \in T$, denote by $\pi_t : S^T \to S$ the *evaluation function* $\pi_t(f) = f(t)$. Then we endow the space S^T with the σ-algebra $\mathcal{S}^T := \sigma(\pi_t : t \in T)$. A *stochastic process* on $(S, \mathcal{S})$ is then defined as a measurable function $X : (\Omega, \mathcal{F}) \to (S^T, \mathcal{S}^T)$.

The space $(S, \mathcal{S})$ is often called the *state space* of the stochastic process. Often, the index set is taken to be $\mathbb{N}$ or $\mathbb{R}_{\geq 0}$. However, the construction above allows for more exotic index sets (although this might impact the properties of the associated stochastic processes).

6.1.2. Constructing random variables

Now that we know what random variables are, there is one problem. In order to define an random variable we need to formally define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable function $X : \Omega \to \mathbb{R}$. This is different from how we are used to work with random variables.

Let X be a random variable. Then its *cumulative distribution function* was given by

$$F_X(t) = \mathbb{P}(X \leq t).$$

The idea behind the cdf $F_X(t)$ is that it denotes the "probability" that $X \in (-\infty, t]$. From the perspective of measure theory, this means we assign a measure to the preimage of $(-\infty, t]$ under the measurable function X. Here we use the shorthand notation $X \leq t$ for the set $\{X \in (-\infty, t]\}$, or equivalently the pre-image $X^{-1}((-\infty, t])$. Another way to view the cdf is via the *push-forward measure* (see Proposition 3.8)

$$F_X(t) := X_{\#}\mathbb{P}((-\infty, t]) = \mathbb{P}(X^{-1}((-\infty, t])).$$

The difference between Definition 6.1 and the normal way you are used to deal with random variables is that before this course, you simply would give a cumulative distribution function (cdf) F and say that X is a random variable with

$$\mathbb{P}(X \leq t) := \mathbb{P}(X \in (-\infty, t]) = F(t),$$

without worrying about a probability space or the measurability of X as a function. However, we now know that we need to construct a probability space and have X be a measurable functions. So how do we reconcile these two approaches?

It turns out that the way of working with random variables you are used to is still valid, as for any cdf F we can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable function X such that $\mathbb{P}(X \le t) = F$. We will provide two ways to do this. But first, we need to formally define what we mean with a cdf.

Definition 6.3: Cumulative distribution function

A function $F : \mathbb{R} \to [0, 1]$ is called a *cumulative distribution function* (or cdf for short) if it is right-continuous[a], non-decreasing and it holds that

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

[a] A function f is right-continuous at a if $\lim_{x\downarrow a} f(x) = f(a)$.

One can show that for a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable function $X : \Omega \to \mathbb{R}$, the function $F(t) := \mathbb{P}(X \in (-\infty, t])$ is a cdf as defined in Definition 6.3 (see Problem **??**). The key result of this section, which we eluded to, is that the converse is also true. That is, any cdf F defines a random variable as defined in Definition 6.1.

Theorem 6.4: Constructing random variables

Let $F : \mathbb{R} \to [0, 1]$ be a cdf. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable X, such that

$$\mathbb{P}\left(X \in (-\infty, t]\right) := \mathbb{P}\left(X^{-1}((-\infty, t])\right) = F(t).$$

In other words, X is a random variable with cdf F.

Proof. Let us first discuss the idea of the proof. Suppose we have a probability measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu((-\infty, t]) = F(t)$. Then we can simply take $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)$ as our probability space and take $X : \mathbb{R} \to \mathbb{R}$ to be $X(\omega) = \omega$. The difficulty in applying this philosophy to our setting is that we do not have this probability measure μ. So the main goal of the proof is to construct this using our cdf F.

Note that the function F is only defined on sets of the form $(-\infty, x]$ and not on general measurable sets $A \in \mathcal{B}_{\mathbb{R}}$. However, if we can extend F to a proper measure μ on $\mathcal{B}_{\mathbb{R}}$ then we would be done. To achieve this, first recall the collection of right-closed intervals (see (2.1))

$$\mathcal{S} = \{\emptyset\} \cup \{(a, b] \,:\, a \le b, a, b \in \mathbb{R}\} \cup \{(a, \infty) \,:\, a \in \mathbb{R}\} \cup \{\mathbb{R}\},$$

which is a semi-algebra that generates the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ (see Problem 2.7). We will construct a pre-measure μ_F on $\mathcal{S}$ such that $\mu_F((t, \infty)) = 1 - F(t)$. Then we use Theorem 2.18

(Carathéodory Extension) to obtain a measure μ on $\sigma(\mathcal{S}) = \mathcal{B}_{\mathbb{R}}$ that extents μ_F. This then implies that

$$\mu((-\infty, t]) = 1 - \mu((t,\infty)) = 1 - \mu_F((t,\infty)) = F(t),$$

as required.

Having explained the outline of the proof lets get started. We first note that any cdf F defines a set function μ_F on $\mathcal{S}$ by

$$\mu_F(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{if } A = \mathbb{R}, \\ F(b) - F(a) & \text{if } A = (a,b], \\ 1 - F(a) & \text{if } A = (a,\infty). \end{cases}$$

In particular, by definition $\mu_F((t,\infty)) = 1 - F(t)$ as we needed.

The only thing left to do is show that μ_F is a pre-measure on $\mathcal{S}$. Since $\mu_F(\emptyset) = 0$ by definition we only need to show that it is σ-additive.

To this end, let $(A_i)_{i\geq 1}$ be a family of disjoint sets in $\mathcal{S}$ such that $\bigcup_{i\geq 1} A_i \in \mathcal{S}$. We must show that

$$\mu_F(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} \mu_F(A_i). \tag{6.1}$$

Before we continue, we observe that if $A_i = \mathbb{R}$ for some $i \geq 1$ then $A_j = \emptyset$ for all other j in which case (6.1) holds. So without loss of generality we can assume that all A_i are non-empty and $A_i \neq \mathbb{R}$.

In addition, suppose that $\bigcup_{i\geq 1} A_i = (a,\infty)$ then there must exist an index j such that $A_j = (b,\infty)$ with $a < b$ and $\bigcup_{i\neq j} A_i = (a,b]$. In this case we get

$$\mu_F(\bigcup_{i\geq 1} A_i) = 1 - F(a) = 1 - F(b) + F(b) - F(a) = \mu_F(A_j) + \mu_F(\bigcup_{i\neq j} A_i).$$

From this we see that it is enough to prove (6.1) for the case where all sets are of the form $(a,b]$.

Consider first a countable families of disjoint sets of the form $A_i = (a_i, a_{i-1}]$ for some non-increasing sequence $(a_i)_{i\geq 1}$ with $a_0 = b$ and $\lim_{i\to\infty} a_i = a$. This implies that $\bigcup_{i\geq 1} A_i = (a,b]$. Using that F is right-continuous we get

$$\begin{aligned}\mu_F(\bigcup_{i\geq 1} A_i) &= F(b) - F(a) = F(b) - \lim_{N\to\infty} F(a_N) \\ &= \lim_{N\to\infty} \sum_{i=1}^{N} F(a_{i-1}) - F(a_i) = \lim_{N\to\infty} \sum_{i=1}^{N} \mu_F(A_i) = \sum_{i\geq 1} \mu_F(A_i).\end{aligned}$$

Now let $(A_i)_{i\geq 1}$ be a general family of disjoint sets of the form $(c,d]$ with $\bigcup_{i\geq 1} A_i = (a,b]$. The main idea is that we can order these sets into groups of sets that are of the previous form. We can then chain the result we obtained to prove that (6.1) holds.

To this end we observe that the sets $(A_i)_{i\geq 1}$ can be ordered into at most countable groups $(C_n)_{n\geq 1}$, such that each group C_n consists of intervals $A_{n_i} = (a_{n_i}, b_{n_i}]$ with $b_{n_1} \leq b$, $a_{n_i} =$

$b_{n_{i+1}}$ for all $i \geq 1$ and $\lim_{i\to\infty} a_{n_i} := a_n \geq a$. With this notation we have that $\bigcup_{i\geq 1} A_{n_i} = (a_n, b_n]$. Moreover, we can choose the ordering such that $b_1 = b$, $a_n = b_{n-1}$ for all $n \geq 2$ and $\lim_{n\to\infty} a_n = a$.

Putting this together we have

$$\begin{aligned}
\sum_{i=1}^{\infty} \mu_F(A_i) &= \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} \mu_F(A_{n_i}) \\
&= \sum_{n=1}^{\infty} \mu_F(\bigcup_{i\geq 1} A_{n_i}) \\
&= \sum_{n=1}^{\infty} F(b_n) - F(a_n) \\
&= F(b) - \lim_{n\to\infty} F(a_n) \\
&= F(b) - F(a) = \mu_F((a,b]) = \mu_F(\bigcup_{i\geq 1} A_i),
\end{aligned}$$

where we used the previous result in the third line and again in the forth one.

We thus conclude that (6.1) holds and thus μ_F is σ-additive as required.

😄

The power of Theorem 6.4 is that is guarantees the existence of random variables for any cdf F. However, the construction of the actual probability space (specifically the probability measure $\mathbb{P}$) is remains somewhat abstract. In some cases though, we can explicitly construct the probability space. This is true for one of the first random variables you encounter in any course in probability theory: *standard uniform random variable.*

Recall that this is a random variable U that takes values in $[0,1]$ and whose cdf satisfies

$$F(t) = \begin{cases} 0 & \text{if } t < 0, \\ t & \text{if } 0 \leq t \leq 1, \\ 1 & \text{if } t > 1. \end{cases} \tag{6.2}$$

While Theorem 6.4 tells us that defining this F is enough, we will construct an explicit probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable function $U : \Omega \to \mathbb{R}$ such that $\mathbb{P}\left(U^{-1}((-\infty, t])\right) = F(t)$.

The following result shows that this is indeed possible. Moreover, in its proof we see a first nice usage of the Lebesgue measure.

Lemma 6.5: Uniform random variable

Let $\Omega = [0,1]$, denote the Borel σ-algebra restricted to $[0,1]$ by $\mathcal{F} = \mathcal{B}_{[0,1]}$ and let $\mathbb{P} := \lambda|_{[0,1]}$ the Lebesgue measure restricted to $[0,1]$. Then the function $U(t) = \mathbf{1}_{(0,1]}\, t$ has a cdf that satisfies (6.2).

Proof. First observe that the function $U(t)$ is measurable, as it is a product of an indicator and a continuous function. By definition of U it follows that

$$U^{-1}((-\infty,t]) = \begin{cases} \emptyset & \text{if } t \le 0, \\ (0,t] & \text{if } 0 < t \le 1, \\ [0,1] & \text{if } t > 1. \end{cases}$$

Since by Theorem 2.19

$$\lambda|_{[0,1]}((0,t]) = \lambda((0,t]) = t,$$

for any $0 < t \le 1$ we have

$$\mathbb{P}\left(U^{-1}((-\infty,t])\right) := \lambda|_{[0,1]}\left(U^{-1}((-\infty,t])\right) = \begin{cases} 0 & \text{if } t < 0, \\ t & \text{if } 0 \le t \le 1, \\ 1 & \text{if } t > 1. \end{cases}$$ 😃

The construction of a standard uniform random variable is extremely important. This is because it can serve as the base from which we can construct any other random variable. To illustrate this we consider the case of an *exponential random variable* with rate $\lambda > 0$. This is a random variable X with cdf

$$F_X(t) = \begin{cases} 0 & \text{if } t \le 0, \\ 1 - e^{-\lambda t} & \text{if } t > 0. \end{cases}$$

For $u \in (0,1)$, write $H(u) := F_X^{-1}(u)$ and note that

$$H(u) = \frac{1}{\lambda} \log\left(\frac{1}{1-u}\right).$$

Now let U be a standard uniform normal random variable and consider the composition $H \circ U : [0,1] \to \mathbb{R}$. First, we note that since cdf $F_X(x)$ is strictly monotonic increase, so is H. In particular, it follows that for any $t > 0$,

$$H^{-1}((-\infty,t]) = (-\infty, H^{-1}(t)] = (-\infty, F_X(t)].$$

While $H^{-1}((-\infty,t]) = \emptyset$ if $t \le 0$.

Hence we get

$$(H \circ U)^{-1}((-\infty,t]) = U^{-1}(H^{-1}((-\infty,t])) = \begin{cases} U^{-1}(\emptyset) & \text{if } t \le 0, \\ U^{-1}((-\infty, F_X(t)]) & \text{if } t > 0. \end{cases}$$

From this it follows that

$$\mathbb{P}\left((H \circ U)^{-1}((-\infty,t])\right) = \begin{cases} 0 & \text{if } t \le 0, \\ 1 - e^{-\lambda t} & \text{if } t > 0, \end{cases}$$

from which we conclude that $H \circ U$ is a way to construct an exponential random variable with rate λ.

The main point of the construction above is to consider the inverse of the cdf F^{-1} and evaluate this on a standard uniform random variable. However, when applying this method in a general case we run into the issue that not every cdf has an inverse. For example, consider a Bernoulli random variable with success probability $0 < p < 1$. Then

$$F(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 - p & \text{if } 0 \leq t < 1, \\ 1 & \text{if } t \geq 1, \end{cases}$$

which does not have an inverse as for any $y \in (0, 1-p)$ there is no t such that $F(t) = y$. Still the idea of inverting the cdf can be extended to this case, using the so-called *generalized inverse.*

Proposition 6.6: Constructing random variables II

Let $F : \mathbb{R} \to [0,1]$ be a cdf and define

$$\overleftarrow{F}(u) := \inf\{x \in \mathbb{R} \, : \, F(x) \geq y\}.$$

Consider the probability space $([0,1], \mathcal{B}_{[0,1]}, \lambda|_{[0,1]})$ and define $X = \overleftarrow{F} \circ U$, where $U(t) = \mathbf{1}_{(0,1]}\, t$ is the standard uniform random variable. Then X is a random variable with $\mathbb{P}(X \leq t) = F(t)$.

Proof. See Problem 6.2 😄

The definition of X as given in Proposition 6.6 is a procedure that is often referred to as the *distribution inversion method.* Due to its explicit and general nature, it can be used to generate random variables in on computers, based on a good pseudo-random number generator (which takes the role of the uniform random variable).

We end this section with an important remark for working with random variables, and random objects in general.

Remark (Probability spaces are implicit!). It is important to note that even though we sometimes have a very explicit probability space to definee a random variable X, in general the probability space will often be *implicit.* That is, if we say that X is a random variable, we assume there is some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that makes X into a measurable function with the right cdf.

Actually, when considering general random objects in $(E, \mathcal{G})$ we often also do not explicitly state or define the probability space. Since the relevant measure is defined through the push-forward, we often only have to worry about taking the right measurable space $(E, \mathcal{G})$.

There are, however, some cases where one should be cautious about the probability space that is used. For example when considering the notion of *convergence in probability* or *almost sure convergence* (see Section [??]). Or when constructing joint distributions of random variables as we will see in Section 6.4.

6.2. Discrete and continuous random variables

Most basic courses in probability theory make a distinction between two classes of random variables: *discrete* and *continuous*. In this section we will put these concepts into the framework of measure theory and provide definitions for the probability mass function and the probability density function.

6.2.1. Discrete random variables

We say that a random variable X is discrete if there exists a countable (possibly infinite) set $N \subset \mathbb{R}$ such that $X(\omega) \in N$ for all $\omega \in \Omega$.

In practice, and throughout these lecture notes, we only consider discrete random variables X whose outcomes are in $\mathbb{Z}$. In Problem 6.4 you can show that this can be done without loss of generality.

A key concept for discrete random variables is the *probability mass function* (pmf), which is often defined as $f(j) = \mathbb{P}(X = j)$ for $j \in \mathbb{Z}$. The next result shows that any discrete random variable has a pmf.

Lemma 6.7: Probability mass function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a discrete random variable, i.e. $X(\omega) \in \mathbb{Z}$ for all $\omega \in \Omega$. Then there exists a sequence $(p_j)_{j\in\mathbb{Z}}$ with $\sum_{j\in\mathbb{Z}} p_j = 1$ such that for any measurable set A,

$$\mathbb{P}(X \in A) = \sum_{j\in\mathbb{Z}} \delta_j(A) p_j.$$

Proof. See Problem 6.5. 😄

With this result, we can now formally define the *probability mass function* (pmf) of X as a function $f : \mathbb{Z} \to [0, 1]$ given by $f(j) = p_j$.

Note that with this definition we indeed have

$$F(t) = \sum_{j=-\infty}^{t} p_j$$

and that

$$f(j) = \mathbb{P}(X = j).$$

Below are some examples of some classical discrete random variables.

Example 6.3.

(a) *Bernoulli*: Let $p \in [0, 1]$. A random variable $X : \Omega \to \{0, 1\}$ with $\mathbb{P}(X = 1) = p$ is called a *Bernoulli random variable* with success probability p.

(b) *Binomial*: Let $n \in \mathbb{N}$ and $p \in [0, 1]$. A random variable $X : \Omega \to \mathbb{N}_0$ with $\mathbb{P}(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$ is called a *Binomial random variable* with n trials and success probability p

(c) *Poisson*: Let $\lambda > 0$. A random variable $X : \Omega \to \mathbb{N}_0$ with $\mathbb{P}(X = j) = \frac{\lambda^j}{j!} e^{-\lambda}$ is called a *Poisson random variable* with mean λ.

6.2.2. Continuous random variables

In contrast to discrete random variables, we say a random variable X is *continuous* if there exist a family $(\mathcal{I}_i)_{i \in I}$ of pairwise disjoint intervals such that $\operatorname{im}(X) = \bigcup_{i \in I} \mathcal{I}_i$. A concept for continuous random variables related to that of the probability mass function is the probability density function.

Definition 6.8: Probability density function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a continuous random variable. We say that X has a *probability density function* (pdf) $\rho : \mathbb{R} \to [0, \infty)$, if for every $t \in \mathbb{R}$,

$$X_{\#}\mathbb{P}((-\infty, t]) \, (= \mathbb{P}(X \leq t)) = \int_{(-\infty, t]} \rho \, \mathrm{d}\lambda.$$

In particular, a probability density function must be integrable.

Note that unlike a probability mass function, a probability density function can yield values larger than 1. An example of this is the pdf for a continuous uniform random variable on $[0, 1/2]$, which is given by $\rho(x) = 2\mathbf{1}_{x \in [0,1/2]}$.

Technically, the notion of a probability density can be defined for any random variable X. However, unlike the pmf for discrete random variables, not all random variables, including continuous ones, have a pdf (see Problem 6.6).

The following classical examples of continuous random variables do all have a pdf.

Example 6.4.

(a) *Exponential*: Let $\lambda > 0$. A random variable $X : \Omega \to \mathbb{R}_+$ with pdf $\rho(x) = \lambda e^{-\lambda x}$ is called an *Exponential random variable* with rate λ.

(b) *Normal*: Let $\mu \in \mathbb{R}$ and $\sigma > 0$. A random variable $X : \Omega \to \mathbb{R}_+$ with pdf $\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$ is called a *Normal random variable* with mean μ and variance σ^2.

(c) *Pareto*: Let $\xi, \alpha > 0$. A random variable $X : \Omega \to [\xi, \infty)$ with pdf $\rho(x) = \frac{\alpha \xi^\alpha}{x^{\alpha+1}}$ is called a *Pareto random variable* with scale ξ and shape α.

6.3. Expected value of random variables

One of the first things you learn to compute for a random variable is its *expected value*, often denoted as $\mathbb{E}[X]$. Armed with the definition of random variables and the notion of integration we can formally define what the expected value of a random variable is.

Definition 6.9: Expected value

Let $(\Omega.\mathcal{F}, \mathbb{P})$ be a probability space and X and random variable. Then

$$\mathbb{E}[X] := \int_\Omega X \, \mathrm{d}\mathbb{P}.$$

In the course Probability and Modeling you have seen the following definition of the expected value of $h(X)$, where h is a function and X a discrete random variable:

$$\mathbb{E}[h(X)] = \sum_{j \in \mathbb{Z}} j \mathbb{P}(h(X) = j).$$

The following result, called the law of the unconscious statistician, expressed this in terms of the pmf of X:

$$\mathbb{E}[h(X)] = \sum_{j \in \mathbb{Z}} h(j) p(j).$$

We will now use the change of variables proposition to prove this result, given the general definition for the expected value in Definition 6.9.

Lemma 6.10

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X be a discrete random variable and consider a function $h : \mathbb{R} \to \mathbb{R}$ such that $h \circ X$ is $\mathbb{P}$-integrable. Then

$$\mathbb{E}[h(X)] = \sum_{j \in \mathbb{Z}} h(j) p(j),$$

where p is the pmf of X.

Proof. Recall the definition of the positive and negative part of a measurable function f, denoted by respectively f^+ and f^-. Further, recall that

$$\int_\Omega f \, \mathrm{d}\mu := \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu$$

Now, for any $n \in \mathbb{N}$ define the functions

$$g_n^\pm = \sum_{j=-n}^{n} (h \circ X)^\pm \mathbf{1}_{X^{-1}(j)}.$$

Then $g_n^\pm \leq g_{n+1}^\pm$ and
$$\lim_{n\to\infty} g_n^\pm = (h\circ X)^\pm.$$
Then, using the monotone convergence theorem we get
$$\begin{aligned}
\int_\Omega (h\circ X)^\pm \,\mathrm{d}\mathbb{P} &= \int_\Omega \lim_{n\to\infty} g_n^\pm \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_\Omega g_n^\pm \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \sum_{j=-n}^{n} \int_\Omega (h\circ X)^\pm \mathbf{1}_{X^{-1}(j)} \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \sum_{j=-n}^{n} \int_{X^{-1}(j)} h^\pm(j) \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \sum_{j=-n}^{n} h^\pm(j)\mathbb{P}(X^{-1}(j)) \\
&= \lim_{n\to\infty} \sum_{j=-n}^{n} h^\pm(j)p(j) = \sum_{j\in\mathbb{Z}} h^\pm(j)p(j).
\end{aligned}$$
Since $h\circ X$ is $\mathbb{P}$-integrable if and only if its positive and negative part are, we conclude that
$$\int_\Omega (h\circ X)\,\mathrm{d}\mathbb{P} = \int_\Omega (h\circ X)^+ \,\mathrm{d}\mathbb{P} - \int_\Omega (h\circ X)^- \,\mathrm{d}\mathbb{P} = \sum_{j\in\mathbb{Z}} h(j)p(j).$$

😃

Let us now turn to the other class of random variables: continuous random variables. Here we introduced the notion of a *probability density function* ρ so that $F(t)$ was equal to the integral of ρ on $(-\infty, t]$.

Now recall that in the case of a continuous random variable Y with a probability density ρ, there was also a formula for the expected value,
$$\mathbb{E}[h(Y)] = \int_\mathbb{R} h(x)\rho(x)\,\mathrm{d}x.$$
Again, this formula is correct (under some mild obvious conditions) and follows from Definition 6.9 after applying a change of variables. The proof of this result is left as an exercise.

Lemma 6.11

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a continuous random variable with probability density ρ and let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $h\rho$ is Lebesgue integrable.

Then

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h\rho \, \mathrm{d}\lambda.$$

Proof. See problem 6.7 😄

6.4. Multi-variate random variables

Up until now we have mainly focused on single random variables, i.e. just one measurable function $X : \Omega \to \mathbb{R}$. However, within the context of probability theory one also often encounters several random variables at once. An example of this are random vectors in $\mathbb{R}^d$ (see Example 6.1). To keep the exposition simple we will cover the case of two random variables ($d = 2$) in this section. Extension to general d dimensions is straightforward.

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and X, Y are two random variables defined on this space. We can then consider the random vector (X, Y), define as the map $\omega \mapsto (X(\omega), Y(\omega))$, which yields a random element in $\mathbb{R}^2$. We can then consider the function $H : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$H(x, y) = \mathbb{P}\left((X, Y) \in (-\infty, x] \times (-\infty, y]\right) = \mathbb{P}(X \leq x, Y \leq y).$$

This function is called the *joint cumulative distribution function* of X and Y.

Similar to the case of a single random variables, we are used to the fact that a joint cdf is enough to jointly define the random variables X and Y. However, as was the case in one dimension, we have to construct a probability space on which both X and Y are defined such that $H = (X, Y)\#\mathbb{P}$.

The next theorem, which is similar in spirit to Theorem 6.4, shows that this is indeed true.

Theorem 6.12: Constructing multivariate random variables

Let F_1, F_2, be two cdfs and let $H : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that

1. For any $y \in \mathbb{R}$, $\lim_{x\to\infty} H(x, y) = F_1(x)$;
2. For any $x \in \mathbb{R}$, $\lim_{y\to\infty} H(x, y) = F_2(y)$;
3. $\lim_{(x,y)\to(\infty,\infty)} H(x, y) = 1$;
4. For any $x, y \in \mathbb{R}$, $\lim_{t\to-\infty} H(x, t) = 0 = \lim_{t\to-\infty} H(t, y)$.

Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables X and Y with cdfs, F_1 and F_2 respectively, such that

$$H(x, y) = \mathbb{P}\left((X, Y) \in (-\infty, x] \times (-\infty, y]\right).$$

Proof. The proof is done by extending the ideas in the proof of Theorem 6.4 to the two-dimensional case and is left as an exercise (see Problem 6.9). 😄

6.5. Problems

Problem 6.1. Let X, Y be two random variables with cdfs F_X and F_Y, respectively.

(a) Prove that if $F_X(t) = F_Y(t)$ for every $t \in \mathbb{R}$, then $X_{\#}\mathbb{P} = Y_{\#}\mathbb{P}$.

The above relation is often denoted as $X \overset{d}{=} Y$ (X is equal to Y in distribution). Basically, this definition says that two random variables are considered equal if their distribution functions are the same, which is implied by the equality of their cdfs. However, the use of the equality sign can be slightly misleading.

(b) Let X be a random variable. Construct a random variable Y such that $X_{\#}\mathbb{P} = Y_{\#}\mathbb{P}$, but $X \neq Y$ as functions $\Omega \to \mathbb{R}$.

Problem 6.2. Here we prove Proposition 6.6.

(a) Show that $\overleftarrow{F}(u) \leq x \iff F(x) \geq u$.

(b) Prove that $\mathbb{P}\left(X^{-1}((-\infty, t])\right) = F(t)$.

Problem 6.3. In this problem we use Proposition 6.6 to explicitly construct two random variables X and Y such that X is *Poisson* distributed with parameter $\lambda > 0$, and Y is *Cauchy* distributed with parameter $\gamma > 0$.

For this we define the Poisson probability mass function (pmf)

$$f_\lambda(n) := \frac{e^{-\lambda}\lambda^n}{n!}, \qquad \forall\, n \in \mathbb{N}, \tag{6.3}$$

and the Cauchy cumulative distribution function (cdf)

$$H_\gamma(z) := \frac{1}{\pi}\arctan\left(\frac{z}{\gamma}\right) + \frac{1}{2}, \qquad \forall\, z \in \mathbb{R}. \tag{6.4}$$

We will first construct the Cauchy random variable.

(a) Define an explicit probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and provide an explicit formula for the function $Y : \Omega \to \mathbb{R}$ such that $Y_{\#}\mathbb{P} = H_\gamma$.

(b) Show that the function from 1 is measurable.

For the Poisson random variable we first need to go from the pmf to its cdf.

(c) Express the cumulative distribution function F_λ for a Poisson random variable in terms of the pmf f_λ.

(d) Define an explicit probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and provide a formula for the function $X : \Omega \to \mathbb{R}$ such that $X_{\#}\mathbb{P} = F_\lambda$.

(e) Show that the function from 4 is measurable.

(f) Show that for any $n \in \mathbb{N}$ $X_{\#}\mathbb{P}(\{n\}) = f_\lambda(n)$.

Problem 6.4. Here you will show that without loss of generality we can consider discrete random variables as measurable functions $X : \Omega \to \mathbb{Z}$.

(a) Let $N_1 \subset [0, \infty)$ be a countable set. Construct a subset $Z_1 \subseteq \mathbb{Z}_{\geq 0}$ and a bijection $\phi_1 : N_1 \to \mathbb{Z}_{\geq 0}$.

(b) Let $N_2 \subset (-\infty, 0)$ be a countable set. Construct a subset $Z_2 \subseteq \mathbb{Z}_{<0}$ and a bijection $\phi_2 : N_2 \to \mathbb{Z}_{<0}$.

(c) Conclude that for any countable set $N \subset \mathbb{R}$ there exists a $Z \subseteq \mathbb{Z}$ and a bijection $\phi : N \to Z$.

(d) Prove that this bijection is measurable.

(e) Now let $Y : \Omega \to \mathbb{Z}$ be defined as $Y(\omega) = \phi(X(\omega))$. Show that Y is a discrete random variable and that its cdf completely determines the cdf of X.

Problem 6.5. Prove Lemma 6.7

Problem 6.6. In this exercise we will see that not all random variables have probability density functions. We start with a very simple case.

(a) Let $X : \Omega \to \mathbb{Z}$ be a discrete random variable. Prove that X does not have a pdf.

You might be inclined to think that the problem with existence of pdfs lies with the discrete nature. Suppose the following two properties are satisfied:

1. There exists a uncountable set $C \subset [0, 1]$ that has Lebesgue measure zero,
2. There exists a continuous increasing function $F : [0, 1] \to [0, 1]$ that is constant outside C, $f(c) = 0$ for all $c \in C$, $F(0) = 0$ and $F(1) = 1$.

(b) Let $X : \Omega \to [0, 1]$ be a random variable with cdf given by the function F described above. Show that X does not have a pdf.

The setting we described above does occur and is a classical example for a continuous random variable in $\mathbb{R}$ that does not have a pdf. The set C is called the Cantor set and the function F is constructed using the so-called the Cantor function.

Simple example of continuous stochastic objects without pdfs exist for higher dimensions, for example $\mathbb{R}^2$. And we will come back to this in a later chapter [REF].

Problem 6.7. The goal of this problem is to prove Lemma 6.11.

Consider the set function

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

which is a measure on the Borel σ-algebra by Problem 4.8.

(a) Prove that $\nu = X_{\#}\mathbb{P}$.

(b) Let $g : \mathbb{R} \to \mathbb{R}$ be a simple function. Show that

$$\int_{\mathbb{R}} g \, \mathrm{d}\nu = \int_{\mathbb{R}} g\rho \, \mathrm{d}\lambda.$$

Now consider the general case, with $h : \mathbb{R} \to \mathbb{R}$ a measurable function such that $h\rho$ is Lebesgue integrable. Consider now the approximation of h by the simple functions $([h]_n)_{n\geq 1}$ defined in Section 4.2.

(c) Prove that

$$\int_{\mathbb{R}} h \, \mathrm{d}\nu = \int_{\mathbb{R}} h\rho \, \mathrm{d}\lambda.$$

[Hint: use monotone convergence]

(d) Prove Lemma 6.11.

Problem 6.8 (Markov's and Chebyshev's inequality)**.** A key result in Probability theory is *Markov's inequality*, which for any $t > 0$ is stated as follows

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

Another classical result is *Chebyshev's inequality*[1]

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Here, you will use the tools from measure theory to prove both statements.

(a) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \overline{\mathbb{R}}$ be an $(\mathcal{F}, \overline{\mathcal{B}})$-measurable function. Show that for any real number $t > 0$ and $p \in (0, +\infty)$, it holds that

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

[Hint: Consider the integral of $|f|^p$ on the subset $\{|f| \geq t\}$]

(b) Prove Markov's inequality. [Hint: use $p = 1$]

(c) Prove Chebyshev's inequality. [Hint: use $p = 2$]

Problem 6.9.

[1] Recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

7. Convergence of integrals and measures

7.1. Convergence of integrals

One of the motivations for developing a new theory of integration using measurable functions instead of continuous ones was that we would be able to change limits and integrals more often. We have already seen an example of such a result in the monotone convergence theorem, Theorem 4.9. However, this required that the sequence f_n of functions was monotone (i.e. non-decreasing) everywhere, which sounds a bit restrictive. That is why in this section we will use the monotone convergence theorem to derive other convergence results with less restrictive conditions.

7.1.1. Monotone convergence (continued)

Theorem 4.9 states that if we have a sequence of measurable functions $(f_n)_{n\in\mathbb{N}}$ from some measure space $(\Omega, \mathcal{F}, \mu)$ to $[0, +\infty]$ such that $f_n \leq f_{n+1}$, then we could interchange limits and integration so that

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

It should be noted that the monotone properties requires that $f_n(\omega) \leq f_{n+1}(\omega)$ for all $\omega \in \Omega$. However, from the definition of the Lebesgue integral we see that it is not affected by sets measure zero. Hence, we would expect that we can relax the monotone property to hold μ-almost everywhere, i.e. the set where it does not hold has measure zero. This turns out to be the case, providing a slightly more general version of the monotone convergence theorem.

Theorem 7.1: Monotone convergence II

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\geq 1}$ be a sequence of non-negative, measurable functions and let f be a non-negative measurable functions such that the following holds μ-almost everywhere

(1) $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$, and

(2) $\lim_{n\to\infty} f_n = f$.

Then

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

Proof. As you might have expected, the proof will use the first monotone convergence theorem. For this we first note that by assumption there exists a $N \in \mathcal{F}$ with $\mu(N) = 0$ such that properties 1 and 2 from theorem statement hold for all $\omega \in \Omega \backslash N$. Now define the function $g_n(\omega) = \max_{1 \le k \le n} f_k(\omega)$. Then $g_n(\omega) \le g_{n+1}(\omega)$ holds for *all* $\omega \in \Omega$. We further define $g(\omega) := \lim_{n\to\infty} g_n(\omega)$. Here comes the key observation. For every $\omega \in \Omega \backslash N$ it holds that $g_n(\omega) = f_n(\omega)$ and $g(\omega) = f(\omega)$. Moreover, since $\mu(N) = 0$ we have that

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

The result then follows by applying Theorem 4.9 to the functions g_n and g. 😄

7.1.2. Fatou's Lemma

Theorem 7.2: Fatou's lemma

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\ge1}$ be a sequence of non-negative, measurable functions and define

$$f := \liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{k\ge n} f_k.$$

Then

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

Proof. Our proof will use the monotone convergence theorem. There are however other proofs, based on first principles.

Define the function $g_n(\omega) := \inf_{k\ge n} f_k(\omega)$ and note that by Lemma 3.13 g_n are measurable. Moreover, $g_n(\omega) \le g_{n+1}(\omega)$ for all $\omega \in \Omega$ and $\lim_{n\to\infty} g_n(\omega) = f(\omega)$. Hence, Theorem 7.1 implies that

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

Finally we observe that by definition $g_n \le f_n$ holds for all $n \in \mathbb{N}$ so that

$$\begin{aligned}\int f \, d\mu = \lim_{n\to\infty} \int_\Omega g_n \, d\mu &= \lim_{n\to\infty} \int_\Omega \inf_{k\ge n} f_k \, d\mu \\ &\le \lim_{n\to\infty} \inf_{\ell\ge n} \int_\Omega f_\ell \, d\mu = \liminf_{n\to\infty} \int_\Omega f_\ell \, d\mu.\end{aligned}$$

Here, we used that $\inf_{k\ge n} f_k \le f_\ell$ for all $\ell \ge n$ and the monotonicity property of the integral (see Proposition 4.12). 😄

7.1.3. Dominated Convergence

Armed with Fatou's lemma we can now prove one of the most useful convergence results for Lebesgue integrals.

Theorem 7.3: Dominated convergence

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\geq 1}$ be a sequence of measurable functions and let f be a measurable function such that $\lim_{n\to\infty} f_n = f$ μ-almost everywhere. Moreover, assume there exists a non-negative μ-integrable function $g : \Omega \to [0, \infty]$ such that $|f_n| \leq g$ μ-almost everywhere. Then

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

Proof. We will first prove the result for the case that both $|f_n| \leq g$ and $\lim_{n\to\infty} f_n = f$ hold everywhere.

Consider the functions $f_n + g$ and note that $|f_n| \leq g$ implies that these are non-negative. Fatou's lemma (Theorem 7.2) now implies that

$$\int_\Omega f + g \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n + g \, \mathrm{d}\mu.$$

Using the additive property of the integral we get

$$\int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu.$$

Since $\int_\Omega g \, \mathrm{d}\mu < \infty$ this implies that

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

On the other hand, the condition $|f_n| \leq g$ also implies that the functions $g - f_n$ are non-negative. Applying Fatou's lemma here yields

$$\int_\Omega g - f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega g - f_n \, \mathrm{d}\mu.$$

The additive property of integral now yields

$$\int_\Omega g \mathrm{d}\mu - \int_\Omega f \, \mathrm{d}\mu \leq \int_\Omega g \, \mathrm{d}\mu + \liminf_{n\to\infty} \int_\Omega -f_n \, \mathrm{d}\mu,$$

which implies that

$$\int_\Omega f \, \mathrm{d}\mu \geq -\liminf_{n\to\infty} \int_\Omega -f_n \, \mathrm{d}\mu = \limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu.$$

We thus conclude that

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

Now let us consider the general case. Then there exists a $N \in \mathcal{F}$ such that $\mu(N) = 0$ and both $|f_n| \leq g$ and $f_n \to f$ hold for every $\omega \in \Omega \backslash N$. Let us now define the following functions

$$\hat{f}_n(\omega) = \begin{cases} f_n(\omega) & \text{if } \omega \in \Omega\backslash N, \\ 0 & \text{else,} \end{cases} \qquad \hat{f}(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in \Omega\backslash N, \\ 0 & \text{else,} \end{cases}$$

and

$$\hat{g}(\omega) = \begin{cases} g(\omega) & \text{if } \omega \in \Omega\backslash N, \\ 0 & \text{else.} \end{cases}$$

Then

$$\int_\Omega \hat{f}_n \,\mathrm{d}\mu = \int_\Omega f_n \,\mathrm{d}\mu \quad \text{and} \quad \int_\Omega \hat{f} \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu$$

Moreover, $\hat{f}_n \leq \hat{g}$ and $\hat{f}_n \to \hat{f}$ hold *everywhere*. So using the first part of the proof we have that

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

Example 7.1. Consider the sequence of functions $f_n(x) = \frac{n \sin(x/n)}{x(x^2+1)}$. We will use dominated convergence to determine $\lim_{n\to\infty} \int_\mathbb{R} f_n \,\mathrm{d}\lambda$. Define $g(x) = \frac{1}{x^2+1}$ and note that

$$f_n(x) = \frac{\sin(x/n)}{x/n} g(x).$$

Note that $|\sin(y)| \leq |y|$ holds for all $y > 0$ and that for every x we have that $\lim_{n\to\infty} \frac{\sin(x/n)}{x/n} = 1$. We thus conclude that $|f_n(x)| \leq g(x)$ and $f_n \to g(x)$ holds for all $x \in \mathbb{R}\backslash\{0\}$. Since the set $\{0\}$ has Lebesgue measure zero, all the conditions of Theorem 7.3 are satisfied. Hence (see Example 4.1)

$$\lim_{n\to\infty} \int_\mathbb{R} \frac{n\sin(x/n)}{x(x^2+1)} \lambda(\mathrm{d}x) = \int_\mathbb{R} \frac{1}{x^2+1} \lambda(\mathrm{d}x) = \pi.$$

7.2. Convergence of finite measures

Until now we have been mainly concerned with convergence of integrals for a sequence of functions $(f_n)_{n\geq 1}$ and fixed measure μ. But what if instead we have a sequence of measures $(\mu_n)_{n\geq 1}$ on a given measurable space. When does this sequence convergence to a limit measure μ? And what does that actually mean?

These are the questions we will address in this section. To properly address them we need to restrict ourselves to finite measures, and thus we will without loss of generality consider probability measures. Moreover, while the concepts we will introduce can be generalized to any topological space with the corresponding Borel σ-algebra, for the sake of clarity we will restrict our attention to $(\mathbb{R}, \mathcal{B}_\mathbb{R})$.

We start by defining what convergence of probability measures means. We will use $C_b(\mathbb{R})$ to denote the class of continuous bounded functions on $\mathbb{R}$.

Definition 7.4

Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. We say that μ_n *converges weakly* (or *narrowly*) to μ if for every continuous bounded function $h \in C_b(\mathbb{R})$ it holds that

$$\int_{\mathbb{R}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} h \,\mathrm{d}\mu.$$

If this is the case we write $\mu_n \Rightarrow \mu$.

The definition of weak convergence asks us to verify the convergence of the μ_n integral of h for any $h \in C_b(\mathbb{R})$. In some cases that can be a cumbersome task. Hence it would be helpful if we would have some equivalent conditions for weak convergence. The beauty here is that there are many equivalent definitions. They are often summarized in what is known as the Portmanteau theorem (or lemma). We provide one version of it below.

For a set $A \subset \mathbb{R}$ denote by $\bar{A}$ the smallest closed set that contains A and by A° the largest open set that is contained in A. The sets $\bar{A}$ and A° are called the *closure* and *interior* of A, respectively. We now define the *boundary* of A as $\partial A := \bar{A} \backslash A^\circ$. Given a measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, a set A is called a *μ-continuity set* if $\mu(\partial A) = 0$.

Definition 7.5: μ-continuity set

Let $A \in \mathcal{B}_{\mathbb{R}}$. Given a measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, a set A is called a *μ-continuity set* if $\mu(\partial A) = 0$.

We can now state a list of equivalent definitions for weak convergence of probability measures.

Theorem 7.6: Portmanteau Theorem

Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then the following statements are equivalent:

(1) $\mu_n \Rightarrow \mu$.

(2) $\int_{\mathbb{R}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} h \,\mathrm{d}\mu$ for all bounded measurable functions h with $\mu(\mathcal{C}_h) = 1$.

(3) $\int_{\mathbb{R}} g \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} g \,\mathrm{d}\mu$ for all continuous functions g with compact support, i.e., functions $g \in C_b(\mathbb{R})$ that are zero outside an interval $[-K, K]$ for some $K > 0$.

(4) $\limsup_{n\to\infty} \mu_n(B) \leq \mu(B)$ for all closed sets $B \subset \mathbb{R}$.

(5) $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$ for all open sets $A \subset \mathbb{R}$.

(6) $\lim_{n\to\infty} \mu(C) = \mu(C)$ for all μ-continuity sets C.

Proof. The proof of this theorem is mostly technical and for the most part does not provide any interesting insights. The implication **1** $\iff$ **3** is dealt with in Problem 7.6. For the full proof we refer to Appendix C. 😄

7.3. Problems

Problem 7.1. Prove the reverse Fatou lemma:

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $(f_n)_{n\geq 1}$ and f be non-negative, measurable functions such that $f_n \leq f$ and $\int_\Omega f \,\mathrm{d}\mu < \infty$. Then

$$\limsup_{n\to\infty} \int_\Omega f_n \mathrm{d}\mu \leq \int_\Omega \limsup_{n\to\infty} f_n \mathrm{d}\mu.$$

Problem 7.2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that $f : \Omega \times (a,b) \to \mathbb{R}$ and that $\omega \mapsto f(\omega, t)$ is integrable with respect to μ for all fixed $t \in (a,b)$. Suppose there exists a non-negative μ-integrable function g such that $|f(\omega,t)| \leq g(\omega)$ for all $t \in (a,b)$ and all $\omega \in \Omega$.

(a) Fix $t_0 \in (a,b)$. Show that if $\lim_{t\to t_0} f(\omega,t) = f(\omega,t_0)$ for all $\omega \in \Omega$, then

$$\lim_{t\to t_0} \int_\Omega f(\omega,t)\,\mu(d\omega) = \int_\Omega f(\omega,t_0)\,\mu(d\omega).$$

(b) Deduce from (a) that if $f(\omega,\cdot)$ is continuous for all ω, then the map

$$(a,b) \ni t \mapsto F(t) := \int_\Omega f(\omega,t)\,\mu(d\omega) \quad \text{is continuous.}$$

Problem 7.3. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that $f : \Omega \times (a,b) \to \mathbb{R}$ and that $\omega \mapsto f(\omega, t)$ is integrable with respect to μ for all fixed $t \in (a,b)$. Suppose $\partial f/\partial t$ exists on (a,b) for all $\omega \in \Omega$, i.e. for every fixed $\omega \in \Omega$,

$$\frac{\partial f}{\partial t}(\omega,t_0) := \lim_{t\to t_0} \frac{f(\omega,t) - f(\omega,t_0)}{t - t_0} \quad \text{exists for all } t_0 \in (a,b).$$

Furthermore, suppose that there is a non-negative μ-integrable function g such that $|\partial f/\partial t|(\omega,t) \leq g(\omega)$ for all $t \in (a,b)$ and all $\omega \in \Omega$. Show that the map

$$(a,b) \ni t \mapsto F(t) = \int_\Omega f(\omega,t)\,\mu(d\omega) \quad \text{is differentiable,}$$

and the following equality holds:

$$\frac{d}{dt}\int_\Omega f(\omega,t)\,\mu(d\omega) = \int_\Omega \frac{\partial f}{\partial t}(\omega,t)\,\mu(d\omega).$$

Hint: Make the following steps:

(1) Show that $(\partial f/\partial t)(\cdot, t)$ is measurable and integrable for all $t \in (a, b)$.

(2) Show that for $t_0 \in (a, b)$ arbitrary

$$\left| \frac{f(\omega, t) - f(\omega, t_0)}{t - t_0} \right| \leq g(\omega) \quad \text{for any } t \in (a, b),\, t \neq t_0 \text{ and all } \omega \in \Omega.$$

(3) Conclude with the help of Problem 7.2.

Problem 7.4. Compute the following limits and justify the computation

$$\lim_{n\to\infty} \int_0^1 \frac{1 + nx^2}{(1 + x^2)^n}\, \mathrm{d}x\,, \qquad \lim_{n\to\infty} \int_{(0,+\infty)} \frac{x^{n-2}}{1 + x^n} \cos\left(\frac{\pi x}{n}\right) \lambda(\mathrm{d}x)\,.$$

Problem 7.5 (Generalized DCT)**.** Prove the following generalization of DCT:

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that f_n, g_n, f and g are μ-integrable functions satisfying

(i) $f_n \to f$ and $g_n \to g$ μ-almost everywhere,

(ii) $|f_n| \leq g_n$ for all $n \in \mathbb{N}$ and $\int_\Omega g_n \,\mathrm{d}\mu \to \int_\Omega g \,\mathrm{d}\mu$ as $n \to \infty$.

Then also $\int_\Omega f_n \,\mathrm{d}\mu \to \int_\Omega f \,\mathrm{d}\mu$ as $n \to \infty$.

Problem 7.6. Here we will prove the implication **1** $\iff$ **3** from Theorem 7.6.

(a) Prove that **1** implies **3**.

The remainder of this problem is dedicate to proving **3** $\Rightarrow$ **1**. So we can assume that $\int_\mathbb{R} g \,\mathrm{d}\mu_n \to \int_\mathbb{R} g \,\mathrm{d}\mu$ holds for all continuous bounded functions with compact support.

Now let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous bounded function with $|f(x)| \leq M$ for all $x \in \mathbb{R}$.

(b) Suppose that for any such f and for any $\varepsilon > 0$ the following holds

$$\left| \int_\mathbb{R} f \,\mathrm{d}\mu_n - \int_\mathbb{R} f \,\mathrm{d}\mu \right| \leq \varepsilon. \tag{7.1}$$

Prove the **1** holds.

It is clear that we have to prove that (7.1) holds for any continuous bounded function f. So from now on let $\varepsilon > 0$ be fixed.

(c) Prove that there exists an $\alpha > 0$ such that $\mu(\mathbb{R}\backslash[-\alpha, \alpha]) < \varepsilon/(2M)$.

(d) Show that we can define a non-negative continuous function g such that $g = 1$ on $[-\alpha, \alpha]$ and $g = 0$ on $\mathbb{R}\backslash(-(\alpha + 1), \alpha + 1)$.

We now write

$$\begin{aligned} \left| \int_\mathbb{R} f \,\mathrm{d}\mu_n - \int_\mathbb{R} f \,\mathrm{d}\mu \right| \leq \left| \int_\mathbb{R} f \,\mathrm{d}\mu_n - \int_\mathbb{R} fg \,\mathrm{d}\mu_n \right| + \left| \int_\mathbb{R} f \,\mathrm{d}\mu - \int_\mathbb{R} fg \,\mathrm{d}\mu \right| \\ + \left| \int_\mathbb{R} fg \,\mathrm{d}\mu_n - \int_\mathbb{R} fg \,\mathrm{d}\mu \right| \end{aligned}$$

(e) Show that
$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mu - \int_{\mathbb{R}} fg \, \mathrm{d}\mu \right| < \frac{\varepsilon}{2}.$$

(f) Show that
$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}} fg \, \mathrm{d}\mu_n \right| < \frac{\varepsilon}{2}.$$

(g) Prove that
$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} fg \, \mathrm{d}\mu_n - \int_{\mathbb{R}} fg \, \mathrm{d}\mu \right| = 0.$$

(h) Conclude that (7.1) holds.

8. Probability II: Convergence of random variables

8.1. Convergence of random variables

Consider a sequence $(X_n)_{n\geq 1}$ of random variables. Similar to the setting of real numbers, we would like to have a notion of convergence of this sequence. In other words, we would like to say that $X_n \to X$ where X is a different random variable. It turns out that there are different ways to define the concept of convergence of random variables. In this section, we will discuss three of them: convergence in distribution, convergence in probability, and almost sure convergence. While the last one has a more straightforward definition (see Definition 8.7) the other two rely the notion of weak convergence of finite measures, which discussed in Section7.2.

8.1.1. Convergence in distribution

Convergence in distribution of a sequence $(X_n)_{n\geq 1}$ is defined as weak convergence of the corresponding probability measures.

Definition 8.1: Convergence in distribution

Let $(X_n)_{n\geq 1}$ and X be random variables, possibly defined on different probability spaces with probability measures $\mathbb{P}_n$ and $\mathbb{P}$, respectively. We say that X_n *converges in distribution* to X if

$$(X_n)_{\#}\mathbb{P}_n \Rightarrow X_{\#}\mathbb{P}.$$

If this is the case write we $X_n \xrightarrow{d} X$.

Note that convergence in distribution of random variables is simply defined as weak convergence of their push-forward measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. This might seem strange to those who have encountered the *more standard* definition used in courses on Probability Theory. There $X_n \xrightarrow{d} X$ if and only if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

holds for all continuity points t of F, with F_n and F denoting the cdfs of X_n and X respectively.

But fear not; this definition is simply an equivalent statement for Definition 8.1.

Lemma 8.2

Let $(X_n)_{n\geq 1}$ and X be random variables and denote by, respectively, F_n and F their associated cdfs. Then $X_n \xrightarrow{d} X$ if and only if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

holds for all continuity points t of F.

Proof. See Problem 8.2. ☺

8.1.2. Convergence in probability

Definition 8.3: Convergence in probability

Let $(X_n)_{n\geq 1}$ and X be random variables that are defined on the *same* probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and define the random variable $Y_n := |X_n - X|$. We say that X_n *converges in probability* to X if

$$(Y_n)_{\#}\mathbb{P} \Rightarrow 0_{\#}\mathbb{P},$$

where 0 denotes the constant function $\omega \mapsto 0$. If this is the case, we write $X_n \xrightarrow{\mathbb{P}} X$.

Remark. Note that, unlike convergence in distribution, the concept of convergence in probability requires all random variables to be defined on the same probability space. This requirement can be relaxed a bit by having each X_n be defined on a different probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ but have X be defined on each of these spaces. From this perspective, we see that if we talk about convergence in probability to a constant $X_n \xrightarrow{\mathbb{P}} a \in \mathbb{R}$, then this is always true as constant random variables can be defined on any probability space.

Recall that for a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we wrote $\mathbb{P}(X \leq t)$ as a short hand notation for $X_{\#}\mathbb{P}((-\infty, t])$, i.e., the cdf of X at t, and $\mathbb{P}(X > t)$ for $X_{\#}\mathbb{P}((t, \infty))$, i.e., 1 minus the cdf of X at t.

The following result relates the definition of convergence in probability to a version that is presented in most probability courses.

Lemma 8.4

Let $(X_n)_{n\geq 1}$ and X be random variables defined on the same probability space. Then $X_n \xrightarrow{\mathbb{P}} X$ if and only if

$$\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0 \quad \text{for every } \varepsilon > 0.$$

Proof. See Problem 8.2 ☺

The notion of convergence in probability is a stronger condition than convergence in distribution. In particular, the first statement implies the second.

Lemma 8.5

Let $(X_n)_{n\geq 1}$ and X be random variables such that $X_n \xrightarrow{\mathbb{P}} X$. Then $X_n \xrightarrow{d} X$.

Proof. We will use the equivalent definition given by Lemma 8.4. Denote by F_n and F the cdfs of X_n and X, respectively. We will also write $\mathbb{P}(X > t)$. Let t be a continuity point of F and fix some $\varepsilon > 0$. First we note that if $X_n \leq t$ and $|X - X_n| \leq \varepsilon$ then $X \leq t + \varepsilon$. We thus obtain

$$\begin{aligned}\mathbb{P}(X_n \leq t) &= \mathbb{P}(\{X_n \leq t\} \cap \{|X_n - X| \leq \varepsilon\}) + \mathbb{P}(\{X_n \leq t\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq \mathbb{P}(X \leq t + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).\end{aligned}$$

Taking the limsup on both sides yields

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leq t) \leq \mathbb{P}(X \leq t + \varepsilon) + \limsup_{n\to\infty} \mathbb{P}(||X_n - X| > \varepsilon) = \mathbb{P}(X \leq t + \varepsilon),$$

since $X_n \xrightarrow{\mathbb{P}} X$ implies that

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Since t is a continuity point of F it follows that

$$\lim_{\varepsilon\downarrow 0} \mathbb{P}(X \leq t + \varepsilon) = \mathbb{P}(X \leq t),$$

which implies that $\limsup_{n\to\infty} \mathbb{P}(X_n \leq t) \leq \mathbb{P}(X \leq t)$.

To prove the result, it now suffices to show that $\mathbb{P}(X \leq t) \leq \liminf_{n\to\infty} \mathbb{P}(X_n \leq t)$. We shall do this in a way that is similar to the case with the limsup. First observe that $X \leq t - \varepsilon$ and $|X_n - X| \leq \varepsilon$ implies that $X_n \leq t$. This way we get

$$\begin{aligned}\mathbb{P}(X \leq t - \varepsilon) &= \mathbb{P}(\{X \leq t - \varepsilon\} \cap \{|X_n - X| \leq \varepsilon\}) + \mathbb{P}(\{X \leq t - \varepsilon\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq \mathbb{P}(X_n \leq t) + \mathbb{P}(|X_n - X| > \varepsilon).\end{aligned}$$

Taking the liminf on both sides gives

$$\mathbb{P}(X \leq t - \varepsilon) \leq \liminf_{n\to\infty} \mathbb{P}(X_n \leq t),$$

and using that t is a continuity point of F_X we conclude that

$$\mathbb{P}(X \leq t) \leq \liminf_{n\to\infty} \mathbb{P}(X_n \leq t).$$

😃

While convergence in probability implies convergence in distribution, the other implication is not true in general (see Problem 8.1). However, if X is constant (deterministic instead of random) then both notions of convergence are equivalent.

Lemma 8.6

Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that $X_n \xrightarrow{d} a$ for some constant $a \in \mathbb{R}$. Then we also have that $X_n \xrightarrow{\mathbb{P}} a$.

Proof. We again use the equivalent definition given by Lemma 8.4. Fix some $\varepsilon > 0$. We have to show that

$$\lim_{n\to\infty} \mathbb{P}(|X_n - a| > \varepsilon) = 0.$$

Let $B_\varepsilon(a)$ denote the open ball of radius ε around a and consider the compliment $B_\varepsilon^c(a) := \mathbb{R}\backslash B_\varepsilon(a)$, which is a closed set. We then have

$$\mathbb{P}(|X_n - a| > \varepsilon) \leq \mathbb{P}(|X_n - a| \geq \varepsilon) = \mathbb{P}(X_n \in B_\varepsilon^c(a)).$$

In particular we have

$$\lim_{n\to\infty} \mathbb{P}(|X_n - a| > \varepsilon) \leq \limsup_{n\to\infty} \mathbb{P}(|X_n - a| > \varepsilon) \leq \limsup_{n\to\infty} \mathbb{P}(X_n \in B_\varepsilon^c(a)).$$

Since $X_n \xrightarrow{d} a$, statement 3 in Theorem 7.6 implies that

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in B_\varepsilon^c(a)) \leq \limsup_{n\to\infty} \mathbb{P}(a \in B_\varepsilon^c(a)) = 0,$$

because obviously $a \notin B_\varepsilon(a)^c$. Therefore we conclude that

$$\lim_{n\to\infty} \mathbb{P}(|X_n - a| > \varepsilon) \leq \limsup_{n\to\infty} \mathbb{P}(a \in B_\varepsilon^c(a)) = 0$$

which implies that $\lim_{n\to\infty} \mathbb{P}(|X_n - a| > \varepsilon) = 0$. 😄

8.1.3. Almost-sure convergence

The final notion of convergence we will discuss is *almost-sure convergence*, which looks much more natural than the previous two notions and requires less notation.

Definition 8.7: Almost-sure convergence

Let $(X_n)_{n\geq 1}$ and X be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X_n *converges almost-surely* to X if

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

In this case we write $X_n \xrightarrow{\text{a.s.}} X$.

The definition of almost-sure convergence says that the set for which X is *not* the limit of X_n must have probability zero. This is why it is also often referred to as *convergence with probability* 1.

There is a different way to characterize *almost-sure* convergence which is often useful. This requires the concept of *infinitely often*.

Definition 8.8: Infinitely often

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and consider a sequence $(A_n)_{n\geq 1}$ of measurable sets. We then define the event $\{A_n \text{ i.o.}\}$ (A_n happens infinitely often) as

$$\{A_n \text{ i.o.}\} := \bigcap_{k=1}^{\infty} \bigcup_{n\geq k} A_n.$$

Lemma 8.9

Let $(X_n)_{n\geq 1}$ and X be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$X_n \overset{\text{a.s.}}{\to} X \iff \mathbb{P}(\|X_n - X\| > \varepsilon \text{ i.o.}) = 0 \quad \text{for all } \varepsilon > 0.$$

Proof. Write $A_n(\varepsilon) := \{\|X_n - X\| > \varepsilon\}$ and $A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$. We first observe that

$$\Omega \backslash A = \bigcup_{m\in\mathbb{N}} \{A_n(1/m) \text{ i.o.}\}.$$

Now suppose that $X_n \overset{\text{a.s.}}{\to} X$ and let $\varepsilon > 0$. Then $\mathbb{P}(A) = 1$ and there exist a $m \in \mathbb{N}$ such that $\{A_n(\varepsilon) \text{ i.o.}\} \subset \{A_n(1/m) \text{ i.o}\}$. Thus

$$\mathbb{P}(\{A_n(\varepsilon) \text{ i.o.}\}) \leq \mathbb{P}(\{A_n(1/m) \text{ i.o.}\}) \leq \mathbb{P}(\bigcup_{m\in\mathbb{N}} \{A_n(1/m) \text{ i.o.}\}) = \mathbb{P}(\Omega \backslash A) = 0.$$

For the other implication suppose that $\mathbb{P}(\{A_n(\varepsilon) \text{ i.o.}\}) = 0$ for all $\varepsilon > 0$. Then clearly the same holds for all $\varepsilon = 1/m$ with $m \in \mathbb{N}$. Hence

$$\mathbb{P}(\Omega \backslash A) = \mathbb{P}(\bigcup_{m\in\mathbb{N}} \{A_n(1/m) \text{ i.o.}\}) \leq \sum_{m\in\mathbb{N}} \mathbb{P}(\{A_n(1/m) \text{ i.o.}\}) = 0,$$

which implies that $\mathbb{P}(A) = 1$. 😄

While this notion of convergence looks very natural, it turn out it is the strongest of the three notions. In practice proving almost-sure convergence is often much harder than proving convergence in probability or distribution.

Lemma 8.10

Let $(X_n)_{n\geq 1}$ and X be random variables such that $X_n \xrightarrow{\text{a.s.}} X$. Then $X_n \xrightarrow{\mathbb{P}} X$.

Proof. See Problem 8.4. 😄

A useful tool for proving almost everywhere convergence is the following result.

Lemma 8.11: Borel-Cantelli

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $(A_n)_{n\in\mathbb{N}}$ be a family of measurable sets. If

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

then for μ-almost every $\omega \in \Omega$, there are only finitely many $n \in \mathbb{N}$ such that $\omega \in A_n$.

Proof. Define the sets

$$B_j := \bigcup_{i\geq j} A_i, \qquad j \in \mathbb{N}.$$

Clearly the sequence $(B_j)_{j\in\mathbb{N}}$ is decreasing and $\{A_n \text{ i.o.}\} \subset B_j$ for every $j \in \mathbb{N}$.

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

$$\mu(\{A_n \text{ i.o.}\}) \leq \mu\left(\bigcup_{j=1}^{\infty} B_j\right) = \lim_{j\to\infty} \mu(B_j) = 0,$$

i.e., $\{A_n \text{ i.o.}\}$ is a null set. In other words, μ-almost every ω is in only finitely many A_n. 😄

Theorem 8.12: Borel-Cantelli Lemma II

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A_i \in \mathcal{F}$, $i \in \mathbb{N}$ be a sequence of independent events such that

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = +\infty.$$

Then

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{i=m}^{\infty} A_i\right) = 1,$$

i.e., infinitely many of the events occur almost surely.

Proof. Let $m \in \mathbb{N}$. Note that

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=m}^{\infty} (\Omega \backslash A_i)\right) &= \prod_{i=m}^{\infty} (1 - \mathbb{P}(A_i)) \\ &\leq \prod_{i=m}^{\infty} \exp(-\mathbb{P}(A_i)) = \exp\left(-\sum_{i=m}^{\infty} \mathbb{P}(A_i)\right) = 0. \end{aligned}$$

😄

8.2. Problems

Problem 8.1. Give an example (with proof) of a sequence of random variables that converge in distribution but not in probability.

Problem 8.2. Prove Lemma 8.4.

Problem 8.3. The goal of this problem is to prove Lemma 8.2. That is

$$X_n \xrightarrow{d} X \iff F_n(t) \to F(t) \quad \text{for all } t \in \mathcal{C}_F,$$

where F_n and F denote the cdfs of the random variables X_n and X, respectively.

Write $\mu_n = (X_n)_\# \mathbb{P}_n$ and $\mu = X_\# \mathbb{P}$ and note that for any measurable function f, $\mathbb{E}[f(X_n)] = \int f \, \mathrm{d}\mu_n$ and $\mathbb{E}[f(X)] = \int f \, \mathrm{d}\mu$.

We will first prove the $\Rightarrow$ implication.

(a) Let $t \in \mathbb{R}$. Find a measurable function h_t, such that $F_n(t) = \int_{\mathbb{R}} h_t \, \mathrm{d}\mu_n$ and $F(t) = \int_{\mathbb{R}} h_t \, \mathrm{d}\mu$.

(b) Show that $\mu(\mathcal{C}_{h_t}) = 1$.

(c) Prove the $\Rightarrow$ implication.

For the other implication $\Leftarrow$ let g be a continuous function with compact support. Then there is some $K > 0$ such that g is zero outside $[-K, K]$. You may use the fact that any continuous function with compact support is uniformly continuous, i.e. for every $\varepsilon > 0$ there exist a $\delta > 0$ such that $\|x - y\| < \delta$ implies that $\|g(x) - g(y)\| < \varepsilon$.

(d) Let $\delta > 0$. Construct a partition of $[-K, K]$ into L intervals I_ℓ such that for each $\ell \leq L$ and $x, y \in I_i$ it holds that $\|x - y\| < \delta$.

We will now define an approximate function

$$\hat{g}(x) = \sum_{\ell=1}^{L} g(\max_{x \in I_i} x)\mathbf{1}_{I_i}.$$

(e) Show that there exists an M and sequences $(\beta_m)_{1\le m\le M}$ and $(t_m)_{1\le m\le M}$ such that

$$\hat{g}(x) = \sum_{m=1}^{M} \beta_m \mathbf{1}_{(-\infty,t_m]}.$$

(f) Prove that $\lim_{n\to\infty} \mathbb{E}[\hat{g}(X_n)] = \mathbb{E}[\hat{g}(X)]$. [Hint: Use the assumption $F_n(t) \to F(t)$ for all $t \in \mathcal{C}_F$.]

(g) Prove that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$. [Hint: First use the previous result to show that $\| \int g \,\mathrm{d}\mu_n - \int g \,\mathrm{d}\mu\| \to 2\varepsilon$ by adding and subtracting $\int \hat{g} \,\mathrm{d}\mu_n$ and $\int \hat{g} \,\mathrm{d}\mu$.]

(h) Conclude that $X_n \xrightarrow{d} X$.

Problem 8.4. Prove Lemma 8.10. [Hint: use the alternative definition of Lemma 8.9 and reverse Fatou from Problem 7.1.]

9. L^p-Spaces

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

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

For $p = +\infty$, we set

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

9.1. The Hölder inequality

Proposition 9.1: 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,

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

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

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

By Young's inequality 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 $\omega \in \Omega$, that

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

Hence

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

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

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

😄

9.2. The Minkowski inequality

Proposition 9.2

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

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

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

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

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

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

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. 😄

9.3. 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

1. $\|v\| = 0 \Leftrightarrow v = 0$ for all $v \in V$
2. $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$
3. $\|v + w\| \leq \|v\| + \|w\|$ 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.

We have seen in Section 4 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 9.3 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

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

9.4. Completeness of L^p-spaces

Theorem 9.3: 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

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

In particular, by the Borel-Cantelli Lemma (cf. Lemma 8.11), for μ-a.e. $\omega \in \Omega$, there is an $N_\omega \in \mathbb{N}$ such that

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

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

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 $N \in \mathbb{N}$ such that for all $m, n \geq N$,

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

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

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

Then, for $n \geq n_K$, 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

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

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

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

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. 😄

9.5. Littlewood's principles

In this section, we will discuss 3 principles—called Littlewood's principles—that provide practical ways of seeing measurable sets, almost everywhere convergence, and measurable functions. These principles hold for measures that are both inner and outer regular as defined in the following.

Definition 9.4: Inner/Outer regularity

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

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

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

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

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

Theorem 9.5

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

Littlwood's first principle 1:

Every measurable set is "practically open".

This principle allows us to approximate arbitrary Borel measurable sets in $\mathbb{R}^d$ with a finite union of open rectangles.

Theorem 9.6

Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Let $A \in \mathcal{B}_{\mathbb{R}^d}$ be a Borel measurable set. Then for any $\epsilon > 0$, 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 9.5 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

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

Littlewood's second principle:

Pointwise almost everywhere convergence is "practically uniform convergence".

In other words, one can think of almost everywhere convergence as uniform convergence on a 'smaller' set that can be chosen arbitrarily 'close' to the full set.

Theorem 9.7: Egorov's Theorem

Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and $(f_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{B}_{\mathbb{R}^d}$-measurable functions that converges μ-almost everywhere to a function f. Then for all $\epsilon > 0$ there exists a set $E \in \mathcal{B}_{\mathbb{R}^d}$ such that $\mu(\mathbb{R}^d\backslash E) \leq \epsilon$ and $f_n \to f$ uniformly on E.

Proof. Since $f_n \to f$ μ-almost everywhere, where exists a μ-null set $N \subset \Omega$, i.e., $\mu(N) = 0$, for which $f_n(x) \to f(x)$ for every $x \in \Omega := \mathbb{R}^d \setminus N$. Consider the sets

$$E_{\ell,n} := \bigcap_{k\geq n} \left\{ \omega \in \Omega \,:\, |f_k(\omega) - f(\omega)| < \frac{1}{\ell} \right\}, \qquad n, \ell \geq 1$$

It is not difficult to check that $E_{\ell,n}$ is measurable for every $\ell, n \geq 1$ and that $E_{\ell,n} \subset E_{\ell,m}$ for $n \leq m$. Moreover, for each $\ell \geq 1$, we have that $\Omega = \cup_{n\geq 1} E_{\ell,n}$. Hence, by the continuity-from-below property of μ, we obtain

$$\mu(\Omega) = \mu\Big(\bigcup_{n\geq 1} E_{\ell,n}\Big) = \lim_{n\to\infty} \mu(E_{\ell,n}).$$

Now choose $n_\ell \geq 1$ such that $\mu(\Omega \backslash E_{\ell,n_\ell}) \leq \varepsilon\, 2^{-\ell}$ and define the measurable set

$$E := \bigcap_{\ell \geq 1} E_{\ell,n_\ell}.$$

Then, by the subadditivity of μ, we obtain

$$\mu(\Omega \backslash E) = \mu\left(\bigcup_{\ell \geq 1} (\Omega \backslash E_{\ell,n_\ell})\right) \leq \sum_{\ell \geq 1} \mu(\Omega \backslash E_{\ell,n_\ell}) = \epsilon.$$

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

$$|f_k(\omega) - f(\omega)| \leq \frac{1}{\ell} \qquad \text{for all } k \geq n_k,$$

thus implying that $f_n \to f$ uniformly on E. ☺

Remark. Given the inner regularity of μ, one may choose E compact in Egorov's Theorem.

Littlewood's third principle:

Every Borel measurable function is "practically continuous".

Theorem 9.8: Lusin's Theorem

Let μ be a finite measure on the measurable space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and $f : \mathbb{R}^d \to \mathbb{R}$ 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 measurable sets

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

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

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

We define the continuous functions $\varphi_k^n : \mathbb{R}^d \to \mathbb{R}$ such that φ_k^n is compactly supported in U_k^n, satisfying $0 \leq \varphi_k^n \leq 1$ and $\varphi_k^n(\omega) = 1$ for $\omega \in K_k^n$. We set

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

which is continuous for all $n \geq 1$. Since the functions $\varphi^n \to f$ μ-almost everywhere, by Egorov's Theorem, there is a compact set K such that φ^n converge to f uniformly on K. Since uniform convergence preserves continuity, $f|_K$ is uniformly continuous on K.

We now construct $g : \mathbb{R}^d \to \mathbb{R}$. Since $f|_K$ is uniformly continuous on K, there is a continuous increasing function $\eta : [0, \infty) \to [0, \infty)$ with $\eta(0) = 0$ (also called the *modulus of continuity*) such that

$$|f(\omega) - f(\sigma)| \leq \eta(|\omega - \sigma|) \qquad \omega, \sigma \in K.$$

Setting

$$g(\omega) := \sup_{\sigma \in K} \Big\{ f(\sigma) - \eta(|\omega - \sigma|) \Big\}.$$

Note that g is continuous on $\mathbb{R}^d$ and coincides with f on K. 😄

The final result of this chapter is an important application of Lusin's theorem, which allows us to approximate any integrable function with continuous and bounded functions whenever μ is a finite measure. In other words, the following statement shows that the space of continuous and bounded functions $C_b(\mathbb{R}^d)$ is *dense* in $L^1(\mathbb{R}^d, \mu)$. This fact is widely used in, e.g., Approximation Theory, Functional Analysis, Partial Differential Equations, and Stochastic Analysis.

Theorem 9.9: Approximation in L^1

Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and $f \in L^1(\mathbb{R}^d, \mu)$. Then for any $\varepsilon > 0$, there is a bounded continuous function $g \in L^1(\mathbb{R}^d, \mu)$ such that $\|f - g\|_1 < \varepsilon$.

Proof. Let $E_n := \{\omega \in \mathbb{R}^d : |f(\omega)| \geq n\}$. Since $\mathbf{1}_{E_n} f \to 0$ as $n \to \infty$, and $\mathbf{1}_{E_n}|f| \leq |f|$ for every $n \geq 1$, we can apply DCT to conclude that

$$\int_{E_n} |f| \,\mathrm{d}\mu = \int_{\mathbb{R}^d} \mathbf{1}_{E_n} |f| \,\mathrm{d}\mu \longrightarrow 0 \quad \text{as } n \to \infty.$$

Now pick some $n \geq 1$ such that $\int_{E_n} |f| \,\mathrm{d}\mu < \varepsilon/3$ and define

$$f_n(\omega) := \max\{-n, \min\{f(\omega), n\}\}, \qquad \omega \in \mathbb{R}^d,$$

i.e., f_n is a truncation of f. From Lusin's theorem, we find a continuous function g such that $f_n \equiv g$ on a compact set $K \subset \mathbb{R}^d$ with $\mu(\mathbb{R}^d \backslash K) < (2\varepsilon)/(3n)$. We assume w.l.o.g. that $|g| \leq n$, since otherwise, we can consider a truncation of g. Altogether, this yields

$$\begin{aligned}
\int_{\mathbb{R}^d} |f - g| \,\mathrm{d}\mu &= \int_{\mathbb{R}^d} |f - f_n| \,\mathrm{d}\mu + \int_{\mathbb{R}^d} |f_n - g| \,\mathrm{d}\mu \\
&= \int_{E_n} |f| \,\mathrm{d}\mu + \int_{\mathbb{R}^d \backslash K} |f_n - g| \,\mathrm{d}\mu \\
&\leq \frac{\varepsilon}{3} + 2n\, \mu(\mathbb{R}^d \backslash K) \leq \varepsilon.
\end{aligned}$$

Finally, $g \in L^1(\mathbb{R}^d, \mu)$ holds simply due to the triangle inequality. 😄

Remark. All three Littlewood principles can be generalized to inner and outer regular measures μ that are *locally finite* on any measurable space $(\Omega, \mathcal{F})$, i.e., a measure for which every point $\omega \in \Omega$ has a neighborhood $N_\omega \in \mathcal{F}$ such that $\mu(N_\omega) < +\infty$.

In particular, the Littlewood principles hold also for the Lebesgue measure λ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ since $\lambda(B_\omega(r)) < +\infty$ for every $\omega \in \mathbb{R}^d$ and any $r > 0$.

9.6. Problems

Problem 9.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon \Omega \to \mathbb{R}$ be an $(\mathcal{F}, \mathcal{B}_\mathbb{R})$-measurable function. Let $p_0, p_1 \in [1, \infty)$ be such that $p_0 < p_1$ and let $\theta \in (0,1)$. Define p_θ by

$$\frac{1}{p_\theta} = \frac{\theta}{p_0} + \frac{1-\theta}{p_1}.$$

(a) Show that $\|f\|_{p_\theta} \le \|f\|_{p_0}^{\theta} \|f\|_{p_1}^{1-\theta}$.

(b) Show that for all $p \in (p_0, p_1)$, there exists $\theta \in (0,1)$ such that $p = p_\theta$. Deduce from this that if $f \in L^{p_0}(\mu) \cap L^{p_1}(\mu)$, then also $f \in L^p(\mu)$ for all $p \in (p_0, p_1)$.

Problem 9.2. Let $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ be a probability space and X be a real-valued random variable.

(a) Show that if $X \in L^\infty(\mathbb{P})$, then $X \in L^p(\mathbb{P})$ for all $p \ge 1$.

(b) Let X be a Gaussian random variable with mean 0 and variance 1. Show that $X \in L^p(\mathbb{P})$ for all $p \ge 1$, but $X \notin L^\infty(\mathbb{P})$.

Problem 9.3. Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. For any $p \in [1, \infty)$ and $(\mathcal{F}, \mathcal{B}_\mathbb{R})$-measurable function $f : \Omega \to \mathbb{R}$, let

$$\Phi_p(f) := \begin{cases} \left(\dfrac{1}{\mu(\Omega)} \displaystyle\int_\Omega |f(\omega)|^p \mu(d\omega) \right)^{1/p} & \text{if } f \in L^p(\mu), \\ +\infty & \text{otherwise.} \end{cases}$$

(a) Show that $p \mapsto \Phi_p(f)$ is monotonically nondecreasing.

(b) Show that if $f \in L^\infty(\mu)$, then

$$\lim_{p \to \infty} \Phi_p(f) = \|f\|_\infty.$$

Hint: Make use of Markov's inequality.

Problem 9.4. In this problem, we would like to refine Theorem 9.9:

Show that for any $\varepsilon > 0$, we can find a bounded continuous function $g \in L^1(\mathbb{R}^d, \mu)$ with compact support, i.e., $g \in C_c(\mathbb{R}^d)$, such that the conclusion of Theorem 9.9 remains true.

10. The Radon-Nikodym Theorem

10.1. A General Radon-Nikodym Theorem

Lemma 10.1

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

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

Proof. Let

$$H := \left\{ f \text{ measurable } : \ 0 \leq f \leq 1, \ \int_E f \, \mathrm{d}\mu \leq \nu(E) \ \text{ for all } E \in \mathcal{F} \right\}.$$

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 $A = \{x \in \Omega \ : \ f_1(x) \geq f_2(x)\}$, then

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

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

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

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

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

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

it follows that $\nu(E_0) > \int_{E_0} f_0 \,\mathrm{d}\mu$. 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

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

In particular, there exists some n_0 such that

$$\begin{aligned}
\nu(F_{n_0}) > \int_{E_0} f_0 \,\mathrm{d}\mu &= \int_{F_{n_0}} f_0 \,\mathrm{d}\mu + \int_\Omega f_0 \,\mathbf{1}_{\{f_0 \geq 1 - n_0^{-1}\}\cap E_0} \,\mathrm{d}\mu \\
&\geq \int_{F_{n_0}} f_0 \,\mathrm{d}\mu + \left(1 + n_0^{-1}\right)\mu(\{f_0 \geq 1 - n_0^{-1}\} \cap E_0) \\
&= \int_{F_{n_0}} f_0 + \varepsilon_0 \mathbf{1}_{F_{n_0}} \,\mathrm{d}\mu,
\end{aligned}$$

with $\varepsilon_0 := \left(1 + n_0^{-1}\right)\mu(F_0^c \cap E_0)/\mu(F_{n_0}) > 0$.

Based on this fact, we claim the existence of a measurable set $F \subset F_{n_0}$ with $\mu(F) > 0$ such that $f_0 + \varepsilon_0 \mathbf{1}_F \in H$. Otherwise, every measurable set $F \subset F_{n_0}$ with $\mu(F) > 0$ would contain a measurable subset $G \subset F$ with $\int_G f_0 + \varepsilon_0 \mathbf{1}_F \,\mathrm{d}\mu > \nu(G)$. By an exhaustion argument, we can construct a disjoint partition $\bigcup_{m\geq 1} G_m = F_{n_0}$ of F_{n_0} such that $\int_{G_m} f_0 + \varepsilon_0 \mathbf{1}_F \,\mathrm{d}\mu > \nu(G_m)$ for all $m \geq 1$. Consequently,

$$\nu(F_{n_0}) > \int_{F_{n_0}} f_0 + \varepsilon_0 \mathbf{1}_F \,\mathrm{d}\mu = \sum_{m\geq 1} \int_{G_m} f_0 + \varepsilon_0 \mathbf{1}_F \, d\mu > \sum_{m \geq 1} \nu(G_m) = \nu(F_{n_0}),$$

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

However, since $\int_\Omega f_0 + \varepsilon_0 \mathbf{1}_F \,\mathrm{d}\mu = M + \varepsilon_0 \mu(F) > M$, this leads to another contradiction, and hence $\nu(E) = \int_E f_0 \,\mathrm{d}\mu$ for all $E \in \mathcal{F}$ as desired. 😃

Theorem 10.2: Radon-Nikodym Theorem

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

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

Proof. Let $\lambda = \mu + \nu$. Then $0 \leq \nu \leq \lambda$, so by Lemma 10.1, there exists a measurable function g with $0 \leq g \leq 1$ such that $\nu(E) = \int_E g \,\mathrm{d}\lambda$ for all $E \in \mathcal{F}$. It follows that $\mu(E) = \int_E (1-g)\,\mathrm{d}\lambda$ for all $E \in \mathcal{F}$. Let $D = \{g = 1\}$. Then $\mu(D) = 0$.

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

$f = (1 + g + \cdots g^n)\mathbf{1}_E$, we learn that

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

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

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

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

Definition 10.3: Absolute continuity of measures

Let μ, ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say that ν is *absolutely continuous w.r.t.* 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 10.4: 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 $\mathrm{d}\nu/\mathrm{d}\mu$, called the *Radon-Nikodym derivative of* ν *w.r.t.* μ such that

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

The function $\mathrm{d}\nu/\mathrm{d}\mu$ is also often called the μ-density of ν.

Proof. We apply Theorem 10.2 to obtain a μ-measurable function f_0 such that

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

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 $\mathrm{d}\nu/\mathrm{d}\mu := f_0$, we then obtain the assertion. 😄

Theorem 10.5: Radon-Nikodym III

Let $\mathbb{P}$, $\mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{Q} \ll \mathbb{P}$. Then there exists a nonnegative $\mathbb{P}$-integrable function $\frac{d\mathbb{Q}}{d\mathbb{P}}$ such that

$$\mathbb{Q}(E) = \int_E \frac{d\mathbb{Q}}{d\mathbb{P}} \, d\mathbb{P} \qquad \text{for all } E \in \mathcal{F}.$$

10.2. Problems

Problem 10.1. Let μ, ν be two measures on $(\Omega, \mathcal{F})$ such that ν is absolutely continuous with respect to μ. Show that for every $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that for every $A \in \mathcal{F}$:

$$\mu(A) < \delta_\epsilon \implies \nu(A) < \epsilon.$$

Hint: Prove by contradiction.

Problem 10.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$. Let $f, g : \Omega \to \mathbb{R}$ be $\mathcal{H}$-measurable functions such that

$$\int_B f \, d\mathbb{P} = \int_B g \, d\mathbb{P} \qquad \text{for all } B \in \mathcal{H}.$$

Prove that $f = g$ $\mathbb{P}$-almost everywhere, i.e.

$$\int_\Omega |f - g| \, d\mathbb{P} = 0.$$

11. Probability Theory III: Independence and Condition Expectation

11.1. Independent random variables

Let us now go back to the setting of probability theory. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variable $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{G}_i)$, $i = 1, 2$. Recall that the *law* of X_i is defined as $\mu_i := (X_i)_{\#}\mathbb{P}$. In Probability and Modeling you were taught that 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 $\mu_1 \otimes \mu_2$.

The nice thing is that we can frame the notion of independence in a more general measure-theoretical setting. To this end we start by defining what we mean by the independence of σ-algebras.

Definition 11.1

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 it 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 $j \in J$,

$$\mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}(A_j).$$

We can now express the independence of random variables in this language.

Definition 11.2

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 11.3

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

$$\mathcal{F}_\alpha := \{\emptyset, A_\alpha, \Omega\backslash A_\alpha, \Omega\} \qquad \text{is independent.}$$

The following result shows that these concepts are really generalizations of concepts that you have seen in elementary probability theory.

Lemma 11.4

Let $X_1, X_2 : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be two random variables. Then X_1 and X_2 are independent according to Definition 11.2, if and only if

$$\mathbb{P}(X_1 \le a,\ X_2 \le b) = \mathbb{P}(X_1 \le a)\,\mathbb{P}(X_2 \le b)$$

for every $a \in \mathbb{R}$, $b \in \mathbb{R}$.

Proof. See Problem 11.1. ☺

11.2. 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 11.5

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

$$\int_B \mathbb{E}[X|\mathcal{H}]\,\mathrm{d}\mathbb{P} = \int_B X\,\mathrm{d}\mathbb{P} \qquad \text{for every } B \in \mathcal{H}.$$

Proof. Define the measure $\mathbb{Q}$ on the measurable space $(\Omega, \mathcal{H})$ by

$$\mathbb{Q}(B) := \int_B X\,\mathrm{d}\mathbb{P} \qquad \text{for every } B \in \mathcal{H}.$$

The measure $\mathbb{Q}$ 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 $\mathbb{E}[X|\mathcal{H}] := \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$, such that for all $B \in \mathcal{H}$,

$$\int_B X\,\mathrm{d}\mathbb{P} = \mathbb{Q}(B) = \int_B \mathbb{E}[X|\mathcal{H}]\,\mathrm{d}\mathbb{P},$$

thereby concluding the proof. ☺

Definition 11.6

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

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)].$$

It is important to note that a conditional expectation is (in general) a stochastic object, i.e. a measurable function. Nevertheless, the next lemma shows that it still satisfies many of the same properties as the regular expectation. Moreover, if X is measurable with respect to $\sigma(Y)$ then conditioning on Y does nothing, i.e. $\mathbb{E}[X|Y] = X$.

Lemma 11.7: Properties of conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$. Furthermore, let X, Y be random variables and $a \in \mathbb{R}$. Then following statements hold $\mathbb{P}$-almost everywhere:

(a) If X is $\mathcal{H}$-measurable, then $\mathbb{E}[X|\mathcal{H}] = X$.

(b) $\mathbb{E}[aX|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}]$.

(c) $\mathbb{E}[X + Y|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] + \mathbb{E}[Y|\mathcal{H}]$.

(d) If $X \leq Y$, then $\mathbb{E}[X|\mathcal{H}] \leq \mathbb{E}[Y|\mathcal{H}]$.

Proof. See Problem 11.4 😃

11.3. Conditional Probability

Definition 11.8: Conditional probability of events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. Then the conditional probability of A given B is defined as:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Armed with the notion of conditional expectation we can now define conditional probabilities. The key observation is that for any probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ and random variable X we have that

$$\mathbb{P}(X \in A) = \int_\Omega \mathbf{1}_{X \in A} \, \mathrm{d}\mathbb{P} = \mathbb{E}[\mathbf{1}_{X \in A}].$$

Definition 11.9: Conditional probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{H}$ a sub-σ-algebra of $\mathcal{F}$, and X a random variable. Then conditional probability of X with respect to $\mathcal{H}$ is defined as

$$\mathbb{P}(X \in A|\mathcal{H}) := \mathbb{E}[\mathbf{1}_{X\in A}|\mathcal{H}].$$

If Y is another random variable we define

$$\mathbb{P}(X \in A|Y) := \mathbb{P}(X \in A|\sigma(Y)).$$

A common formula given to you when considering discrete random variables X and Y was the following:

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbb{P}(X \in A|Y)] = \sum_{k\in\mathbb{Z}} \mathbb{P}(X \in A|Y = k)\mathbb{P}(Y = k). \tag{11.1}$$

This was referred to as the total law of probability. However, in the above expression it is not clear what $\mathbb{P}(X \in A|Y = k)$ is. The issue here is that we have a definition for the random variable $\mathbb{P}(X \in A|Y)$, which is a measurable function $\Omega \to \mathbb{R}$. But here we would like to consider $\mathbb{P}(X \in A|Y = k)$ as a function $\mathbb{Z} \to \mathbb{R}$. How would you define this? This is what the next lemma will help us with.

Lemma 11.10

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y be two random variables with joint density $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Further, let f_Y denote the density of Y and define

$$h(y) := \int_{\mathbb{R}} \frac{x\, f(x, y)}{f_Y(y)}\, \mathrm{d}\lambda$$

Then

$$h(Y) = \mathbb{E}[X|Y].$$

Proof. We need to show that for all $B \in \sigma(Y)$

$$\int_B h(Y)\, \mathrm{d}\mathbb{P} = \int_B X\, \mathrm{d}\mathbb{P} := \mathbb{E}[\mathbf{1}_B X].$$

Note that it suffices to consider sets of the form $B = Y^{-1}(A)$ for some $A \in \mathcal{B}$. Moreover

$\mathbf{1}_B(\omega) = \mathbf{1}_B(Y(\omega))$. Hence

$$\begin{aligned}
\int_B h(Y)\,\mathrm{d}\mathbb{P} &= \int_\Omega \mathbf{1}_B h(Y)\,\mathrm{d}\mathbb{P} = \int_\Omega \mathbf{1}_{Y^{-1}(A)} h(Y)\,\mathrm{d}\mathbb{P} \\
&= \int_{\mathbb{R}} \mathbf{1}_A(y) h(y) f_Y(Y)\,\lambda(\mathrm{d}y) && \text{by change of variables} \\
&= \int_{\mathbb{R}} \mathbf{1}_A(y) \left(\int_{\mathbb{R}} \frac{x\, f(x,y)}{f_Y(y)}\,\lambda(\mathrm{d}x) \right) f_Y(y)\lambda(\mathrm{d}y) \\
&= \iint_{\mathbb{R}^2} \mathbf{1}_A(y) x f(x,y)\,\lambda(\mathrm{d}x)\lambda(\mathrm{d}y) \\
&= \mathbb{E}[\mathbf{1}_A(Y)X] = \mathbb{E}[\mathbf{1}_B X] && 😄
\end{aligned}$$

The function

$$g(x,y) := \frac{f(x,y)}{f_Y(y)}, \tag{11.2}$$

is referred to as the conditional density of X given $Y = y$. We often write $f_{X|Y}(x|y) := g(x,y)$ to emphasize the conditioning on $Y = y$.

The function $h(y)$ is the formal way to interpret the expression $\mathbb{E}[X|Y = y]$. So in the example of the total law of probability, we would have that $\mathbb{P}(X \in A|Y = y) := \mathbb{E}[\mathbf{1}_{X\in A}|Y = y]$. With this we can now formally establish (11.1), see Problem 11.5.

11.4. Problems

Problem 11.1. Prove Lemma 11.4

Problem 11.2 (Joint densities)**.** Let X, Y be two random variables with a joint density $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. That is, f as a function from $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable and integrable, and for any $A \in \mathcal{B}_{\mathbb{R}^2}$

$$\mathbb{P}((X,Y) \in A) = \int_A f\,\mathrm{d}\lambda^2,$$

where λ^2 is the 2-dimensional Lebesgue measure on $\mathbb{R}^2$

Define the marginal function $f_X(x) = \int_{\mathbb{R}} f(x,y)\,\lambda(\mathrm{d}y)$.

(a) Prove that f_X, as a function from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable and integrable.

(b) Show that

$$X_{\#}\mathbb{P}(A) = \int_A f_X\,\mathrm{d}\lambda.$$

That is, f_X is the density function of X.

Problem 11.3. Prove that the conditional expectation is unique $\mathbb{P}$-almost everywhere.

Problem 11.4 (Properties conditional expectation)**.** The goal of this problem is to prove Lemma 11.7.

Problem 11.5. Use Lemma 11.10 to prove (11.1).

A. Appendix

A.1. Uniqueness of measures

In this section will provide the proof of Theorem 2.15. For this weneed to introduce a few concepts as well as a powerful theorem, called the monotone class theorem.

We start with the definition of an algebra.

Definition 1.1: Algebra's of sets

A collection $\mathcal{A}$ of subsets of Ω is called an *algebra* if

1. $\emptyset \in \mathcal{A}$,
2. $\Omega \setminus A \in \mathcal{A}$ for all $A \in \mathcal{A}$, and
3. $A \cup B \in \mathcal{A}$ for every $A, B \in \mathcal{A}$.

Observe that, as the name suggests, every σ-algebra is indeed and algebra. However, in addition to the properties of an algebra, σ-algebras where also closed under countable unions and intersections. We will actually take these properties on their own and define any collection of subsets that have these two properties a monotone class.

Definition 1.2: Monotone classes

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

1. $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$ holds for any increasing family of sets $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{M}$, and
2. $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{M}$ holds for any decreasing family of sets $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{M}$

As we already remarked, any σ-algebra is a monotone class. However, there are monotone classes that are not algebras and vise versa, there are algebras that are not monotone classes. However, suppose we start with an algebra $\mathcal{A}$ and we want to turn this into a σ-algebra. Then we at least need to ensure it is also a monotone class. Similar to the construction of $\sigma(\mathcal{A})$ we can construct the smallest monotone class that contains $\mathcal{A}$. Moreover, it turns out, maybe not surprisingly, that the resulting collection is σ-algebra. Even better, it is exactly $\sigma(\mathcal{A})$. This is the content of the monotone class theorem.

Theorem 1.3: Monotone class theorem

Let $\mathcal{A}$ be an algebra on Ω and let $\Xi_{\mathcal{A}}$ denote the collection of all monotone classes that contain $\mathcal{A}$. Then

1. the collection defined by
$$\mathcal{M}(\mathcal{A}) = \bigcup_{\mathcal{M}\in\Xi_{\mathcal{A}}} \mathcal{M},$$
is a monotone class, and moreover
2. $\mathcal{M}(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$, i.e. $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.

Proof. TODO ☺

With the monotone class theorem at hand we can prove the uniqueness theorem for measures.

Proof of Theorem 2.15. Define the collection
$$\mathcal{M} := \{A \in \mathcal{F} \,:\, \mu_1(A) = \mu_2(A)\}\,.$$
The goal of the proof is to show that this is a monotone class. If that is true then, since $\mathcal{A} \subset \mathcal{M}$, the monotone class theorem (Theorem [REF]) implies that $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A}) \subset \mathcal{M}$ and hence $\mu_1 = \mu_2$ on $\sigma(\mathcal{A})$.

To show that $\mathcal{M}$ is a monotone class let $(A_i)_{i\in\mathbb{N}}$ be an increasing sequence in $\mathcal{M}$. Since by definition, $\mu_1(A_i) = \mu_2(A_i)$ for all $i \in \mathbb{N}$, continuity from below (Proposition [REF]) implies that
$$\mu_1\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{i\to\infty} \mu_1(A_i) = \lim_{i\to\infty} \mu_2(A_i) = \mu_2\left(\bigcup_{i\in\mathbb{N}} A_i\right),$$
which implies that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$.

Similarly, now let $(A_i)_{i\in\mathbb{N}}$ be a decreasing sequence. Again, by definition $\mu_1(A_i) = \mu_2(A_i)$ for all $i \in \mathbb{N}$ and moreover $\mu_1(A_1) = \mu(A_1) < \infty$ since both measures are finite. Hence continuity from above (Proposition [REF]) implies that
$$\mu_1\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{i\to\infty} \mu_1(A_i) = \lim_{i\to\infty} \mu_2(A_i) = \mu_2\left(\bigcap_{i\in\mathbb{N}} A_i\right).$$
It then follows that $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{M}$ which shows that $\mathcal{M}$ is indeed a monotone class. ☺

A.2. Construction of the Lebesgue measure

But how can we construct a measure on this set? In particular, is it possible to start with a set function that does not satisfy all the properties of a measure? We will address these questions next. But in order to do so we need to introduce the notion of an *algebra*.

Definition 1.4: Algebra's of sets

A collection $\mathcal{A}$ of subsets of Ω is called an *algebra* if

1. $\emptyset \in \mathcal{A}$,
2. $\Omega \setminus A \in \mathcal{A}$ for all $A \in \mathcal{A}$, and
3. $A \cup B \in \mathcal{A}$ for every $A, B \in \mathcal{A}$.

Note that every σ-algebra is an algebra. The idea is that is we start with a set function on an algebra, we can extend this to all the way to a measure on σ-algebra. To ensure this extension is possible, we need to start with set functions that have some structure, suspiciously called premeasures.

Definition 1.5: Premeasures

Let $\mathcal{A}$ be an algebra on Ω. A set function $\mu_o : \mathcal{A} \to [0, \infty]$ is called a *premeasure* if

1. $\mu_o(\emptyset) = 0$, and
2. μ_o is σ-additive.

If we start with a premeasure μ_o on an algebra $\mathcal{A}$ we can construct a new set function on the entire collection of subsets of Ω.

Definition 1.6: Outer measure

Let μ_o be a premeasure on an algebra $\mathcal{A}$ on Ω. Then the set function μ^* defined by

$$\mu^*(A) := \inf \left\{ \sum_{i=1}^{\infty} \mu_o(A) \ : \ A \subset \bigcup_{i \in \mathbb{N}} A_i, \ A_i \in \mathcal{A} \right\},$$

is called the *outer measure induced by* μ_o.

The idea is that the outer measure μ^* is almost a measure. This is captured by the following set of properties it has.

Proposition 1.7

Let μ_o be a premeasure on an algebra $\mathcal{A}$ on Ω and μ^* be the outer measure induced by μ_o. Then μ^* satisfies the following properties:

1. $\mu^*(A) = \mu_o(A)$ for all $A \in \mathcal{A}$,

2. $\mu^*(\emptyset) = 0$ and $\mu^*(A) \geq 0$ for all $A \subset \Omega$,
3. μ^* is monotone, and
4. μ^* is σ-subadditive.

Proof. TODO 😀

Observe that indeed, μ^* is almost a measure. The only property missing is full σ-additivity. Then next fundamental result, due to the Greek mathematician Constantin Carathéodory, provides a way to construct a σ-algebra from a given algebra such that μ^* can be extended to a true measure on it. We state a partial version here, without proof.

Theorem 1.8: Carathéodory's extension theorem (partial)

Let $\mathcal{A}$ be an algebra on Ω. Let μ_0 be a pre-measure on $\mathcal{A}$ and denote by μ^* the outer measure induced by μ_0. Then the collection defined by

$$\mathcal{A}_{\mu^*} := \{B \subset \Omega \,:\, \mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \setminus B)\, \forall A \in \mathcal{A}\}\,,$$

is a σ-algebra on Ω. Moreover, the restriction $\bar{\mu} := \mu^*|\mathcal{A}_{\mu^*}$ of μ^* to $\mathcal{A}_{\mu^*}$ is a measure on $\mathcal{A}_{\mu^*}$ called the *Carathéodory extension of* μ_o.

At this point we should take some time to fully appreciate what Theorem 1.8 gives us. In order to construct a measure all we need is an algebra on Ω and some premeasure.

Remark. The statement in Theorem 1.8 only covers part of the original theorem. It actually turns out that the σ-algebra constructed has some very nice properties and the measure space $(\Omega, \mathcal{A}_{\mu^*}, \bar{\mu})$ is *complete*. However, in order to properly define these notions we needed to introduce additional concepts going beyond the goal of this section. The interested reader is referred to the Appendix for the full statement and details, including the proof of this theorem.

Let us now utilize the Carathéodory extension to obtain a measure on the Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

B. Appendix: From Riemann to Lebesgue

B.1. Recalling Riemann integration

Definition 2.1

A partition $P = (x_0, \dots, x_n)$ of $[a, b]$ is an $(n+1)$-tuple of real numbers x_i such that $a = x_0 < x_1 < \cdots < x_n = b$, 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

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

and

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

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

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

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

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

and

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

Definition 2.2

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

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

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

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

B.2. Riemann vs Lebesgue integration

Theorem 2.3

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

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

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

$$\lim_{n\to\infty} U(P_n, f) = \lim_{n\to\infty} L(P_n, f) = \int_a^b f(x)\,\mathrm{d}x. \tag{B.1}$$

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

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

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

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

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

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

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

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

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

$$\begin{aligned}L(P_n, f) &\leq \int_a^b f(x)\,\mathrm{d}x < L(P_n, f) + \frac{1}{n}\\ U(P_n, f) - \frac{1}{n} &< \int_a^b f(x)\,\mathrm{d}x \leq U(P_n, f)\end{aligned}$$

from which the limits (B.1) follow.

Now define the functions

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

Because P_{n+1} is a refinement of P_n, we find that $\ell_n \leq \ell_{n+1}$, $u_{n+1} \leq u_n$, 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, $\ell \leq f \leq u$. Note that

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

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

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

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

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

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

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

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. 😆

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

Theorem 2.4

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

C. Appendix: Portmanteau Theorem

We first mention an important technical lemma needed for the proof of this theorem. For any function $h\colon \mathbb{R} \to \mathbb{R}$, we denote by $\mathcal{C}_h \subset \mathbb{R}$ the set of continuity points of h, i.e., the set of all points $x \in \mathbb{R}$ at which h is continuous. The following technical lemma, which can be deduced from Lusin's theorem (cf. Theorem 9.8), allows us to approximate measurable functions by continuous ones, with arbitrary precision in terms of the integrals.

Lemma 3.1

Consider a probability measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and bounded measurable function $h \in B_b(\mathbb{R})$ with $\mu(\mathcal{C}_h) = 1$. Then for every $\varepsilon > 0$, there exist continuous bounded functions h_ε^- and h_ε^+ such that

(1) $h_\varepsilon^- \le h \le h_\varepsilon^+$ and

(2) $\int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu - \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu < \varepsilon$.

Proof. For $k \in \mathbb{N}$ define the functions

$$g_k := \inf_{y\in\mathbb{R}} h(y) + k\|x-y\| \quad G_k := \sup_{y\in\mathbb{R}} h(y) - k\|x-y\|.$$

We then observe that for any $k \in \mathbb{N}$, $g_k \le g_{k+1}$, $G_{k+1} \le G_k$ and $g_k \le h \le G_k$. Hence

$$g_1 \le g_2 \le \cdots \le h \le \cdots \le G_2 \le G_1.$$

Moreover, since for any fixed $x \in \mathbb{R}$ the sequences $(g_k(x))_{k\ge1}$ and $(G_k(x))_{k\ge1}$ are bounded we get that their limits as $k \to \infty$ exist and

$$\lim_{k\to\infty} g_k(x) \le h(x) \le \lim_{k\to\infty} G_k(x).$$

We now claim that for every $k \ge 1$ the functions g_k and M_k are continuous and bounded. The last part follows from directly from the definitions. For the continuity we note that

$$g_k(x) = \inf_{y\in\mathbb{R}} h(y) + k\|x-y\| \le \inf_{y\in\mathbb{R}} h(y) + k\|z-y\| + k\|x-z\| = g_k(z) + k\|x-z\|,$$

which implies that $\|g_k(x) - g_k(y)\| \le k\|x-y\|$. A similar argument works for G_k.

Now, let $x \in \mathcal{C}_h$, i.e. x is a continuity point of h, and fix $\varepsilon > 0$. Then there exist a $\delta > 0$ such that $\|x-y\| < \delta$ implies that $\|h(x) - h(y)\| < \varepsilon$. If we then define

$$r = \left\lceil \frac{h(x) - \inf_{z\in\mathbb{R}} h(z)}{\delta} \right\rceil,$$

then

$$\begin{aligned}\lim_{k\to\infty} g_k(x) &\geq g_r(x)\\ &= \min\{\inf_{\|x-y\|\geq\delta} h(y) + r\|x-y\| + \inf_{\|x-y\|<\delta} h(y) + r\|x-y\|\}\\ &\geq \min\{h(x)-\varepsilon,\ \inf_{z\in\mathbb{R}} h(z) + (h(x) - \inf_{z\in\mathbb{R}} h(z))\delta/\delta\}\\ &= h(x)-\varepsilon.\end{aligned}$$

Similarly, we get that $\lim_{k\to\infty} G_K(x) \leq h(x)+\varepsilon$. Since $\mu(\mathcal{C}_h)=1$ this then implies that

$$\int_{\mathbb{R}} \lim_{k\to\infty} g_k \,\mathrm{d}\mu = \int_{\mathbb{R}} h \,\mathrm{d}\mu = \int_{\mathbb{R}} \lim_{k\to\infty} G_k \,\mathrm{d}\mu.$$

Applying Theorem 7.1 to g_k and to $-G_k$ we get that

$$\lim_{k\to\infty} \int_{\mathbb{R}} g_k \,\mathrm{d}\mu = \int_{\mathbb{R}} h \,\mathrm{d}\mu = \lim_{k\to\infty} \int_{\mathbb{R}} G_k \,\mathrm{d}\mu.$$

Finally, since g_k is non-decreasing and G_k is non-increasing, for every ε there must exist an K such that for all $k \geq K$

$$\int_{\mathbb{R}} (G_k - g_k) \,\mathrm{d}\mu = \int_{\mathbb{R}} (G_k - h) \,\mathrm{d}\mu + \int_{\mathbb{R}} (h - g_k) \,\mathrm{d}\mu \leq \varepsilon.$$

So we can take

$$h_\varepsilon^- := g_K \quad \text{and} \quad h_\varepsilon^+ := G_K.$$

😄

Theorem 3.2: Portmanteau Theorem

Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then the following statements are equivalent:

(1) $\mu_n \Rightarrow \mu$.

(2) $\int_{\mathbb{R}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} h \,\mathrm{d}\mu$ for all $h \in B_b(\mathbb{R})$ with $\mu(\mathcal{C}_h) = 1$.

(3) $\int_{\mathbb{R}} g \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} g \,\mathrm{d}\mu$ for all continuous functions with compact support $g \in C_c(\mathbb{R})$, i.e., functions $g \in C_b(\mathbb{R})$ that are zero outside an interval $[-K, K]$ for some $K > 0$.

(4) $\limsup_{n\to\infty} \mu_n(B) \leq \mu(B)$ for all closed sets $B \subset \mathbb{R}$.

(5) $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$ for all open sets $A \subset \mathbb{R}$.

(6) $\lim_{n\to\infty} \mu(C) = \mu(C)$ for all μ-continuity sets C.

Proof. We will prove the following implication chain: $5 \iff 4 \Rightarrow 1 \Rightarrow 2 \Rightarrow 6 \Rightarrow 4$ and $1 \iff 3$.

5 $\iff$ 4: This follows directly since every closed set B is the complement of an open set A, i.e., $B = \mathbb{R}\backslash A$ and thus

$$\limsup_{n\to\infty} \mu_n(B) = \limsup_{n\to\infty} 1 - \mu_n(A) = 1 - \liminf_{n\to\infty} \mu_n(A).$$

4 $\Rightarrow$ 1: Let h be a continuous bounded function. Then, without loss of generality, we may assume that $0 \le h < 1$. Now fix some $k \in \mathbb{N}$ and define the following sets:

$$B_j := \left\{ x \in \mathbb{R} \,:\, \frac{j}{k} \le h(x) \right\} \qquad \text{for } j = 0, 1, \ldots, k.$$

Note that since h is continuous these are closed sets. Also note that $\mu(B_0) = 1$ and $\mu(B_k) = 0$.

We further observe that $h(x) = \sum_{j=1}^{k} h(x)\mathbf{1}_{B_{j-1}\cap B_j^c}$, where $B_j^c = \mathbb{R}\backslash B_j$. Hence, we can obtain the following bounds:

$$\sum_{j=1}^{k} \frac{j-1}{k}\mu(B_{j-1}\cap B_j^c) \le \int_{\mathbb{R}} h \,\mathrm{d}\mu \le \sum_{j=1}^{k} \frac{j}{k}\mu(B_{j-1}\cap B_j^c). \tag{C.1}$$

Using that $B_{j-1} \supset B_j$ we get

$$\mu(B_{j-1}) = \mu(B_{j-1}\cap B_j^c) + \mu(B_{j-1}\cap B_j) = \mu(B_{j-1}\cap B_j^c) + \mu(B_j)$$

so that

$$\mu(B_{j-1}\cap B_j^c) = \mu(B_{j-1}) - \mu(B_j)$$

Plugging this into the sum on the right-hand side in Equation (C.1), we get

$$\begin{aligned}
\sum_{j=1}^{k} \frac{j}{k}\mu(B_{j-1}\cap B_j^c) &= \sum_{j=1}^{k} \frac{j}{k}(\mu(B_{j-1}) - \mu(B_j)) \\
&= \frac{1}{k}\left(\mu(B_0) + \sum_{j=1}^{k-1}(j+1)\mu(B_j) - \sum_{j=1}^{k}\mu(B_j) \right) \\
&= \frac{1}{k}\left(1 + \sum_{j=1}^{k-1}\mu(B_j) - k\mu(b_k) \right) \\
&= \frac{1}{k} + \frac{1}{k}\sum_{j=2}^{k}\mu(B_j),
\end{aligned}$$

where we used that $\mu(B_0) = 1$ and $\mu(B_k) = 0$.

In a similar fashion, the sum on the left-hand side in Equation (C.1) equals

$$\frac{1}{k}\sum_{j=1}^{k}\mu(B_j).$$

We thus conclude that for any $k \geq 1$,

$$\frac{1}{k}\sum_{j=1}^{k}\mu(B_j) \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu \leq \frac{1}{k} + \frac{1}{k}\sum_{j=2}^{k}\mu(B_j). \tag{C.2}$$

Moreover, the same inequalities hold for the measures μ_n.

Applying (4) we then get

$$\begin{aligned}
\limsup_{n\to\infty}\int_{\mathbb{R}} h \,\mathrm{d}\mu_n &\leq \limsup_{n\to\infty}\left(\frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k}\mu_n(B_j)\right)\\
&\leq \frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k}\limsup_{n\to\infty}\mu_n(B_j)\\
&\leq \frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k}\mu(B_j)\\
&\leq \frac{1}{k} + \int_{\mathbb{R}} h \,\mathrm{d}\mu.
\end{aligned}$$

So that by taking $k \to \infty$ we obtain

$$\limsup_{n\to\infty}\int_{\mathbb{R}} h \,\mathrm{d}\mu_n \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu.$$

Apply this conclusion to the function $-h$, which is also continuous and bounded, we get

$$\int_{\mathbb{R}} h \,\mathrm{d}\mu \leq \liminf_{n\to\infty}\int_{\mathbb{R}} h \,\mathrm{d}\mu_n,$$

from which it follows that $\lim_{n\to\infty}\int_{\mathbb{R}} h \,\mathrm{d}\mu_n = \int_{\mathbb{R}} h \,\mathrm{d}\mu$ for any bounded continuous function.

1 $\Rightarrow$ 2: Fix $\varepsilon > 0$ and let h_ε^- and h_ε^+ be the function from Lemma [REF]. Then

$$\int_{\mathbb{R}} h \,\mathrm{d}\mu \leq \int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu = \int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu - \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu + \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu,$$

which implies that

$$\int_{\mathbb{R}} h \,\mathrm{d}\mu - \varepsilon \leq \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu.$$

In a similar way we obtain that

$$\int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu + \varepsilon.$$

Now we employ condition 1 for the functions h_ε^- and h_ε^+ to get

$$\begin{aligned}\int_{\mathbb{R}} h \,\mathrm{d}\mu - \varepsilon &\leq \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} h_\varepsilon^- \,\mathrm{d}\mu_n \\ &\leq \liminf_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n \\ &\leq \limsup_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n \\ &\leq \int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu_n \\ &= \int_{\mathbb{R}} h_\varepsilon^+ \,\mathrm{d}\mu \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu + \varepsilon.\end{aligned}$$

From this it follows that

$$\int_{\mathbb{R}} h \,\mathrm{d}\mu - \varepsilon \leq \liminf_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n \leq \limsup_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu + \varepsilon.$$

And since $\varepsilon > 0$ was arbitrary we conclude that

$$\liminf_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n = \limsup_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n,$$

which then implies that $\int_{\mathbb{R}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} h \,\mathrm{d}\mu$.

2 ⇒ 6: Let C be a μ-continuity set and consider the function $h(x) = \mathbf{1}_C$. Then clearly h is measurable and bounded. Moreover, the function h is discontinuous precisely on the boundary ∂C and hence

$$\mu(\mathcal{C}_h) = \mu(\mathbb{R}\backslash\partial C) = 1 - \mu(\partial C) = 1 - 0 = 1.$$

Hence the function h satisfies the conditions of 2 and thus

$$\mu_n(C) = \int_{\mathbb{R}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} h \,\mathrm{d}\mu = \mu(C).$$

6 ⇒ 4: Let B be a closed set, take $\delta > 0$ and consider the sets

$$A_\delta = \{x \in \mathbb{R} \,:\, \|x - B\| < \delta\},$$

where $\|x - B\| = \inf_{y\in B} \|x - y\|$ denotes the distance from x to the set B. Note that A_δ is an open set in $\mathbb{R}$, and hence $A_\delta^\circ = A_\delta$.

Next we observe that $A_\delta \subset \{x \in \mathbb{R} \,:\, \|x - B\| \leq \delta\}$ where the latter sets are closed. It then follows that

$$\partial A_\delta = \bar{A}_\delta \backslash A_\delta \subset \{x \in \mathbb{R} \,:\, \|x - B\| \leq \delta\}\backslash A_\delta = \{x \in \mathbb{R} \,:\, \|x - B\| = \delta\}.$$

It then follows that $\partial A_\delta \cap \partial A_{\delta'} = \emptyset$ for all $\delta \neq \delta'$. Since μ is a probability measure, there can be only a countable number of disjoint sets with positive measure. From this we conclude that there exists a sequence $(\delta_k)_{k\geq 1}$ with $\delta_k \to 0$ such that $\mu(\partial A_{\delta_k}) = 0$ for all $k \geq 1$. Let us write $B_k := A_{\delta_k}$. Then each B_k is a μ-continuity set, $B_k \supset B_{k+1}$ and $B_k \downarrow B$ because B is closed.

We then have that

$$\limsup_{n\to\infty} \mu_n(B) \leq \limsup_{n\to\infty} \mu_n(B_k) = \mu(B_k),$$

where the last equality is due to 6, which implies that $\mu_n(B_k) \to \mu(B_k)$.

Taking $k \to \infty$ now yields 4.

1 $\iff$ 3: The implication $1 \Rightarrow 3$ is trivial. So assume that $\int_{\mathbb{R}} g \,\mathrm{d}\mu_n \to \int_{\mathbb{R}} g \,\mathrm{d}\mu$ holds for all continuous bounded functions with compact support and let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous bounded function with $|f(x)| \leq M$ for all $x \in \mathbb{R}$. We will show that for any $\varepsilon > 0$

$$\left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} f \,\mathrm{d}\mu \right| \leq \varepsilon,$$

which then implies the result.

So let $\varepsilon > 0$ be fixed and observe that there exists an $\alpha > 0$ such that $\mu(\mathbb{R}\backslash[-\alpha,\alpha]) < \varepsilon/(2M)$. Also observe that we can define a non-negative continuous function g such that $g = 1$ on $[-\alpha,\alpha]$ and $g = 0$ on $\mathbb{R}\backslash(-(\alpha+1),\alpha+1)$. Observe that g is a non-negative continuous bounded function that is zero outside the interval $[-(\alpha+1),\alpha+1]$, and thus we can apply (3). We now have that

$$\begin{aligned}\left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} fg \,\mathrm{d}\mu_n \right| = \left| \int_{\mathbb{R}} f(1-g) \,\mathrm{d}\mu_n \right| &\leq M \int_{\mathbb{R}} (1-g) \,\mathrm{d}\mu_n \\ &\leq M \int_{\mathbb{R}} (1-g) \,\mathrm{d}\mu_n = M \left(1 - \int_{\mathbb{R}} g \,\mathrm{d}\mu_n \right)\end{aligned}$$

Since the later term converges to $\int_{\mathbb{R}} g \,\mathrm{d}\mu$ by our assumption we get that

$$\limsup_{n\to\infty} \left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} fg \,\mathrm{d}\mu_n \right| \leq M \int_{\mathbb{R}} (1-g) \,\mathrm{d}\mu \leq M\mu(\mathbb{R}\backslash[-\alpha,\alpha]) < \frac{\varepsilon}{2}.$$

The same conclusion holds true for $\left|\int_{\mathbb{R}} f \,\mathrm{d}\mu - \int_{\mathbb{R}} fg \,\mathrm{d}\mu\right|$.

If we now write

$$\begin{aligned}\left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} f \,\mathrm{d}\mu \right| \leq \left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} fg \,\mathrm{d}\mu_n \right| &+ \left| \int_{\mathbb{R}} f \,\mathrm{d}\mu - \int_{\mathbb{R}} fg \,\mathrm{d}\mu \right| \\ &+ \left| \int_{\mathbb{R}} fg \,\mathrm{d}\mu_n - \int_{\mathbb{R}} fg \,\mathrm{d}\mu \right|\end{aligned}$$

we see that the first two terms converge to $\varepsilon/2$ (by the computation above) while the term on the second line converges to zero by our assumption since fg is also a continuous bounded function that is zero outside the interval $[-(\alpha+1),\alpha+1]$.