TU/e, 2MBA70

Measure and Probability Theory

Pim van der Hoorn and Oliver Tse

Version 0.3 September 9 2024

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 2.1.1: 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 2.1.2

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 2.1.3: 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 $C_1, C_2, \dots$ be sets in $\mathcal{F}_A$. Then there are $B_1, B_2, \dots \in \mathcal{F}$ 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 2.1.4: 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 $\mathcal{F}$ 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 $\mathcal{F}$ is a σ-algebra.

Similar to Lemma 2.1.3 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 2.1.5: 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 2.1.6: 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 2.1.7: 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 actually define a *Borel σ-algebra* on any metric space. Actually, we can define it on any topological space. However, this requires the notion of a topology which is beyond the scope of this course. [ADD REFERENCES]

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 2.1.8

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 2.9. 😄

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 $\mathcal{C}$

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

Definition 2.2.9: 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.

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*. If the $\mu(\Omega) < \infty$ we say that μ is σ-finite and call the associated measure space a *σ-finite measure space*. If $\mu(\Omega) = 1$ we call μ a *probability measure* and the associated measure space a *probability space*.

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*.

However, there is a measure, not included above, that plays a fundamental role in measure theory and especially probability theory.

2.2.2 Null sets, complete measure spaces and the Lebesgue measure

It should be noted that, outside maybe the discrete measure, the examples listed above do not include any interesting measure. More specifically, consider the Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Then 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 is simply its volume, which seems like a very natural property to ask for. In particular, we want the measure of the rectangle, say in $\mathbb{R}^2$, to be 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.

However, to formally state the theorem we need to introduce the concept of *null sets* and *complete measure spaces*.

Definition 2.2.10: 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 2.10.

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

Theorem 2.2.11: 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 $M \in \mathbb{R}^{d\times d}$, $\lambda(M^{-1}(B)) = |\mathrm{det} M|^{-1}\lambda(B)$.

The proof of this theorem is involved and relies on a more abstract approach to constructing measures. The interested student is referred to the Appendix, where we provide the full details.

It follows from Theorem 2.2.11 that the Lebesgue measure 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).

We end this section by looking at some important general properties of measures.

2.2.3 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 2.2.12: 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 2.6.
5. Property 4 implies that $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B) \geq \mu(A \cup B)$. ☺

The subadditive property can actually be extended to any countable family of sets.

Lemma 2.2.13: 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

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

and the measure μ is said to be *σ-subadditive*.

Proof. See Problem 2.7 ☺

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? It turns out that this can be expressed as the measure of the union of all sets.

Proposition 2.2.14: Continuity from below

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

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

Proof. Define the sets $E_1 = A_1$ and $E_i = A_{i+1} \setminus A_i$, for all $i \geq 2$. Then $(E_i)_{i\in\mathbb{N}}$ is a family of mutually disjoint measurable sets with the following properties:

$$A = \bigcup_{i=1}^{\infty} E_i \quad \text{and} \quad A_k = \bigcup_{i=1}^{k} E_i.$$

Therefore, using σ-additivity we get

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

😃

A similar property holds for any *decreasing* family of sets. That is, a family $(A_i)_{i\in\mathbb{N}}$ of measurable sets such that $A_i \supset A_{i+1}$ holds for all $i \in \mathbb{N}$. Here we do have to make an assumption on the measure of the biggest set A_1.

Proposition 2.2.15: Continuity from above

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(A_i)_{i\in\mathbb{N}}$ be an decreasing family of measurable sets such that $\mu(A_1) < \infty$. Then

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

Proof. See Problem 2.8. 😃

In addition to being useful in determining the limits of the measure of families of sets, these continuity properties are actually powerful enough to characterize a measure.

Theorem 2.2.16: 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
$$\mu(A_\infty) = \lim_{i\to\infty} \mu(A_i) \quad (= \sup_{i\in\mathbb{N}} \mu(A_i)).$$

Proof. The fact that any measure satisfies these three properties follows from the definition and Propositions 2.2.12 and 2.2.14. 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 2.2.17: 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 covered in the Appendix, as it is requires another more technical result. What is important is the implication of Theorem 2.2.17: to study a measure on $\sigma(\mathcal{A})$ it suffices to look at what it does on the generator $\mathcal{A}$.

2.3 Problems

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

1. $\mathcal{F}$ is non-empty;
2. If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$;
3. 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 2.1.1.

Problem 2.2. Prove Lemma 2.1.2. [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 2.1.5.

Problem 2.6. Prove the *strongly additive* property of Proposition 2.2.12. [Hint: write $A \cup B$ as the union of three disjoint sets.]

Problem 2.7. Prove Lemma 2.2.13. [Hint: Construct a new family of sets and use properties σ-additive and monotone]

Problem 2.8. Prove Proposition 2.2.15. [Hint: The proof is very similar to that of Proposition 2.2.14.]

Problem 2.9. The goal of this problem is to prove Proposition 2.1.8. 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 eitherh the family $\mathcal{A}_1 := \{(a,b), a,b \in \mathbb{Q}\}$ or $\mathcal{A}_1' := \{(a,b), a,b \in \mathbb{R}\}$.

1. 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?]

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

3. 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.]

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

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

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

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

7. 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 2.1.8. 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.

8. Explain what changes in the proof of point 3 of Proposition 2.1.8 from the proof of point 2 of this proposition outlined above.

9. Describe the proof strategy to get points 4-8 of Proposition 2.1.8 using 1-4.

Problem 2.10. 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

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

1. 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

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

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

3. Show that $\bar{\mu}|_{\mathcal{F}} = \mu$.

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

3 Measurable functions and stochastic objects

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 between such spaces, called *measurable functions*. We will focus on functions that preserve the measurable structure of the spaces. We do this first in a general setting and then move to functions that map to the real line $\mathbb{R}$. We end this chapter by using the notion of measurable function to properly define (general) random objects and prove a general theorem to construct any random variable from a given distribution function.

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 it the best way to do this it to look at the preimage of measurable sets in E.

Definition 3.1.1: 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* is $f^{-1}(B) \in \mathcal{F}$ 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. This is different from the notion of continuity of functions on $\mathbb{R}^d$.

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.

We will provide an important example of measurable functions to $\mathbb{R}$, 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 mean 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 3.1.2: 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 3.1.3: 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 $f|_A : \Omega \to E$ 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 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 3.1.1 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 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 3.1.4

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. Consider the following collection of subsets:

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

We claim that $\mathcal{H}$ is a σ-algebra on E. Suppose this is indeed true. Then, since by construction $\mathcal{A} \subseteq \mathcal{H}$, it follows from Lemma 2.1.5 that $\mathcal{G} = \sigma(\mathcal{A}) \subseteq \mathcal{H}$. But this then implies that $f^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{G}$ which means that f is $(\mathcal{F}, \mathcal{G})$-measurable.

So let's prove that $\mathcal{H}$ is a σ-algebra. First we note that $f^{-1}(\emptyset) = \emptyset \in \mathcal{F}$ and $f^{-1}(E) = \Omega \in \mathcal{F}$. So $\emptyset, E \in \mathcal{H}$.

Next, let $B \in \mathcal{H}$. Then

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

since by definition $f^{-1}(B) \in \mathcal{F}$. So $E \setminus B \in \mathcal{H}$.

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

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

which shows that $\bigcup_{i=1}^{\infty} B_i \in \mathcal{H}$, completing the proof that $\mathcal{H}$ is a σ-algebra. 😄

We thus see that at least. But that still requires us to check if any given function is measurable. 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 does exactly this.

Proposition 3.1.5

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 3.1.4 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 $B_x(r) \in f^{-1}(O)$.

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 have a vast world of measurable functions $f : \mathbb{R}^d \to \mathbb{R}^m$ at our disposal. 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.

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

Proposition 3.1.6

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

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

Proof. First note that by Proposition 2.1.4, $\sigma(f_i \,:\, i \in I)$ 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 2.1.5 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 3.1.4, 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 3.1.7

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 3.1.4. But by Proposition 3.1.6 $\mathcal{G}_1$ 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 2.1.6. 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 3.1.6 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 3.2,

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

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 3.1.8: 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 $f_{\#}\mu$.

The proof of this result is elementary and left as an exercise, see Problem 3.3.

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 3.3 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 thes properties will be useful when we introduce the concept of integration of measurable functions in Chapter [REF] and develop limit theorems for integrals in Chapter [REF].

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 $x \in (0, +\infty)$, $x \pm (+\infty) = \pm\infty, \; x \pm (-\infty) = \mp\infty$,
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 3.2.9: 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 3.2.10

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

$$\mathcal{B}_{\mathbb{R}} = \overline{\mathcal{B}} \cap \mathbb{R}.$$

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 3.4 ☺

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 3.2.11

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 2.1.8 and Lemma 3.1.4 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 $\{f \neq a\} = \Omega \setminus \{f = a\}$. 😊

Lemma 3.2.12

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 3.5.

(2) We first note that the sets $\{f \geq g\}$ and $\{g > f\}$ are measurable. This follows from Lemma 3.2.11 and the fact that

$$\{f \geq g\} = \bigcup_{q \in \mathbb{Q}} \{f \geq q\} \cap \{g \leq q\},$$

while

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

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

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

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

Lemma 3.2.10 $\overline{\mathcal{B}}$ is generated by the sets $[a, \infty]$, for $a \in \mathbb{Q}$. Hence, by Lemma 3.1.4 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 3.2.11.

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:

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

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 3.2.13

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 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 3.6.

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 3.2.11 and hence $\sup_{n\geq 1} f_n$ is a measurable function (check this yourself, see Problem 3.6).

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 3.2.13 look rather trivial, it is actually very important. In particular, it shows the power of the class of measurable functions. In contrast, the class of continuous functions is not stable under point-wise limit operations.

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 3.2.13 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.

For now we will move from the general setting of measurable functions to their application in the field of probability theory, in particular the concept of random variables.

3.3 Random variables and general stochastic objects

3.3.1 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

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

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 3.1.4 that X is a measurable function. This is actual the proper way to define a random variable.

Definition 3.3.14: Random variable

A *random variable* is a measurable function from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the (extended) real line.

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.

Let X be a random variable and recall that its *cumulative distribution function* $F_X : \mathbb{R} \to [0, 1]$ is defined as

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

The fact that we use $\mathbb{P}$ here, which is the probability measure on the space $(\Omega, \mathcal{F})$ is not a coincidence.

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 need to assign a measure to the preimage of $(-\infty, t]$ under the measurable function X. For this, the only things we have at our disposal is the probability measure $\mathbb{P}$ and the measurable function X. Now recall from Proposition 3.1.8 that we can always construct a measure from this, the push-forward measure. That is exactly what the cummulative distribution is,

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

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 3.3.15: 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 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 one. 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.

Example 3.3 (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 3.4 (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).

3.3.2 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. Here we simply present a cdf F and say that X is a random variable with $\mathbb{P}(X \leq t) = F(t)$, without worrying about a probability space or the measurability if X as a function. It turns out that this way of working with random variables is valid, as for any cdf F we can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable function X such that $X_{\#}\mathbb{P} = F$. We will start this construction for a specific random variable and then use it to construct a random variable with any cumulative distribution function.

One of the first random variables you encounter in any course in probability theory is the *standard uniform random variable*. This is a random variable U that takes values in $[0, 1]$ such that its cdf satisfies $F(t) = t$ for all $0 \leq t \leq 1$. In the course Probability and Modeling this description would be enough to work with. But now that we know what a random variable actually is, we need a bit more. More precisely, we have to construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

and a measurable function $U : \Omega \to \mathbb{R}$ such that

$$\mathbb{P}\left(U^{-1}((-\infty, t])\right) = \begin{cases} 0 & \text{if } t < 0, \\ t & \text{if } 0 \le t \le 1, \\ 1 & \text{if } t > 1. \end{cases} \tag{3.3}$$

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

Proposition 3.3.16: Uniform random variable

There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable U, such that $\mathbb{P}\left(U^{-1}((-\infty, t])\right)$ satisfies (3.3).

Proof. Consider the space $\Omega = [0,1]$ together with the restricted Borel σ-algebra $\mathcal{F} = \mathcal{B}_{\mathbb{R}}|_{[0,1]}$ and as probability measure the restricted Lebesgue measure $\mathbb{P} := \lambda|_{[0,1]}$. Now consider the function $U(t) = \mathbf{1}_{(0,1]}\, t$. Then, 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 2.2.11

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

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 standard uniform random variable is extremely important, as it is the base from which we can construct any other random variable. To illustrate this let us first 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 the standard 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 $t \leq 0$.

Hence we get

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

From this it follows that

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

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 extending this to the more general case we have to deal with the fact 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 $F(t) = y$.

Nevertheless, if does hold than any cdf F is non-decreasing and right continuous. For these type of functions there exists the notion of a *generalized inverse*, defined as

$$\overleftarrow{F}(u) := \inf\{x \in \mathbb{R} \, : \, F(x) \geq y\}. \tag{3.4}$$

The construction we used for the exponential random variable can now be generalize by using $\overleftarrow{F}$ instead of F^{-1}. This results in the following theorem on the existence of random variables with a given cdf.

Theorem 3.3.17: Constructing random variables

Let $F : \mathbb{R} \to [0,1]$ be a right continuous, non-decreasing function with

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

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.

Moreover, $(\Omega, \mathcal{F}, \mathbb{P})$ can be chosen as $([0,1], \mathcal{B}_{[0,1]}, \lambda|_{[0,1]})$ and $X = \overleftarrow{F} \circ U$, where U is the standard uniform random variable.

Proof. We start with the following important observation:

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

The implication from right to left is by definition of $\overleftarrow{F}$ and the fact that F is non-decreasing. The implication from left to right is because F is right continuous.

Now let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and U a standard normal random variable. We will show that $X = \overleftarrow{F} \circ U$ is a random variable with the right probability measure. Since we can construct a standard uniform random variable on the probability $([0,1], \mathcal{B}_{[0,1]}, \lambda|_{[0,1]})$ this also implies the last part.

Consider the preimage of $(-\infty, t]$ under X. Then, using the above observation, we have

$$\begin{aligned} X^{-1}((-\infty, t]) &= \{\omega \in \Omega \, : \, \overleftarrow{F}(U(\omega)) \in (-\infty, t]\} \\ &= \{\omega \in \Omega \, : , U(\omega) \in (-\infty, F(t)]\} = U^{-1}((-\infty, F(t)]) \in \mathcal{B}_{[0,1]}. \end{aligned}$$

Hence, X is measurable. Finally, the above computation, together with Proposition 3.3.16, also implies that

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

which finished the proof. 😄

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 used a very explicit probability space to construct a standard uniform random variable and the random variable X in the proof of Theorem 3.3.17, 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. Theorem 3.3.17 actually says that this is okay as we can always construct an appropriate probability space and measurable function to achieve the needed 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*. Or when constructing joint distributions of random variables.

3.4 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 $A \cap B = \emptyset$.

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

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

Problem 3.3 (Push-forward measure)**.** Prove Proposition 3.1.8.

Problem 3.4. Prove Lemma 3.2.10.

Problem 3.5. The goal of this problem is to prove points 1, 3, and 5 of Lemma 3.2.12.

(a) Based on the proofs of points 2 and 4 of Lemma 3.2.12, explain the general idea behind the proof.

(b) Prove point 1 of Lemma 3.2.12.

(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 3.2.12 (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 3.2.12.

Problem 3.6.

(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 3.2.13. [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 3.7 (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.

Problem 3.8. 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 $X \stackrel{d}{=} Y$ (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.

2. 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 3.9. In this problem we will 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)

$$f_\lambda(n) := \frac{e^{-\lambda}\lambda^n}{n!}, \qquad \forall\, n \in \mathbb{N}, \tag{3.5}$$

and the Cauchy cumulative distribution function (cdf)

$$H_\gamma(z) := \frac{1}{\pi}\arctan\left(\frac{z}{\gamma}\right) + \frac{1}{2}, \qquad \forall\, z \in \mathbb{R}. \tag{3.6}$$

We will first construct the Cauchy random variable.

1. 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 $Y_\#\mathbb{P} = H_\gamma$.

2. Show that the function from 1 is measurable.

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

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

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

5. Show that the function from 4 is measurable.

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

Problem 3.10. Let X be a random variable that only takes values in $\mathbb{Z}$, i.e. $X(\omega) \in \mathbb{Z}$ for all ω. Show that there exists a sequence $(p_j)_{j\in\mathbb{Z}}$ with $0 \leq p_j \leq 1$ and $\sum_{j\in\mathbb{Z}} p_j = 1$ such that for any Borel measurable set A

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

4 The Lebesgue Integral

We now arrive at one of our main characters in Measure Theory: The Lebesgue integral. Unlike the Riemann integral, the Lebesgue integral can be constructed on any measure space $(\Omega, \mathcal{F}, \mu)$. The construction will be done in multiple steps, starting with simple functions.

4.1 The integral of a simple function

Definition 4.1.1

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

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

for some positive integer $N \in \mathbb{N}$, disjoint measurable sets $A_1, \dots, A_N \in \mathcal{F}$ and constants $a_1, \dots, a_N \in \mathbb{R}$. We define the *$\mu$-integral* of a simple function f by

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

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

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

By the finite additivity of the measure μ,

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

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

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

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

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

4.2 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 4.2.2

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

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

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

Remark. If X is a nonnegative random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we call the integral the expectation value of X and often write instead

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

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

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

Lemma 4.2.3: Properties of the Lebesgue integral of nonnegative functions

Let f, g be two nonnegative, measurable functions and $\alpha \ge 0$ 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 $f \leq g$, then
$$\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu.$$
3. (homogeneity)
$$\alpha \int_\Omega f \,\mathrm{d}\mu = \int_\Omega (\alpha f) \,\mathrm{d}\mu.$$

4.3 The monotone convergence theorem

We now arrive at the first convergence result telling us that the point-wise limit of monotone sequences of μ-integrable functions is again μ-integrable, highlighting the difference with Riemann integration.

Theorem 4.3.4: 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

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

😄

4.4 Intermezzo: Approximation by simple functions

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

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

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

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

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

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

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

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

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

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

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

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

4.5 Additivity of the Lebesgue integral of nonnegative functions

A fundamental property that we need for a good notion of an integral is linearity.

Lemma 4.5.5: Additivity of the Lebesgue integral of nonnegative functions

Let f, g be two nonnegative $(\mathcal{F}, \mathcal{B}_{[0,+\infty]})$-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 is easy to check. 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 pointwise to $(f+g)$. By the monotone convergence theorem,

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

On the other hand, by a limit theorem and Problem 4.2, we know that

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

Therefore,

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

😄

4.6 Integrable functions

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

Definition 4.6.6

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

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

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

Proposition 4.6.7

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

Definition 4.6.8

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 $(A, \mathcal{F}_A, \mu|_A)$.

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 **??**). 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 4.7.9: 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, $g_n \to g$ point-wise monotonically. By the MCT, 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 4.7.9. 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.

4.8 Change of variables formula

Proposition 4.8.10

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,

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

Proof. We first show 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 $h \circ f$. By the monotone convergence theorem,

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

😄

As a direct consequence, we have the following proposition.

Proposition 4.8.11

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 $h\colon E \to \mathbb{R}$ be $(\mathcal{F}, \mathcal{G})$- and $(\mathcal{G}, \mathcal{B}_\mathbb{R})$-measurable maps respectively. Then $h \circ f$ is integrable with respect to μ if and only if h is integrable with respect to $f_\# \mu$, in which case,

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

4.9 Expectation of random variables

Now that we have a notion of integration we can formally define what the expected value of a random variables is.

Definition 4.9.12

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

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

We say that a random variable is discrete if $X(\omega) \in \mathbb{Z}$ for all $\omega \in \Omega$. It then follows (see Problem 3.10) that

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

for some sequence $(p_j)_{j\in\mathbb{Z}}$ with $\sum_{j\in\mathbb{Z}} p_j = 1$. We can now define the *probability mass function* (pmf) of X as $p(j) = p_j$.

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

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

This was actually the definition of the expectation for a discrete random variable. The following result shows that this is correct, given the general definition for expectations in Definition 4.9.12.

Lemma 4.9.13

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 it 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 the notion of a *probability density function* ρ was introduce so that $F(t)$ was equal to the integral of ρ on $(-\infty, t]$. Expressed formally in the language of measure theory we have the following

Definition 4.9.14: 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* $\rho : \mathbb{R} \to [0, \infty)$, if for every $t \in \mathbb{R}$,

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

In particular, a probability density function much be integrable.

Now recall that in the case of a continuous random variable Y with a probability density ρ, there was also a formula for its expectation,

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

Again, this formula is correct and follows from Definition 4.9.12 after applying a change of variables. The proof of this result is left as an exercise.

Lemma 4.9.15

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 measurable function. Then

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

Proof. See problem [??]

4.10 The Markov inequality

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

Lemma 4.10.16: The Markov inequality

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

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

Proof. The result follows easily from

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

😄

In probability language, the Markov inequality looks as follows.

Lemma 4.10.17: The Markov inequality

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

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

4.11 Problems

Problem 4.1. Consider the measure space $(\mathbb{N}, 2^{bbN}, \mu)$, 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 4.2. 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 4.3. 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 4.4. 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 4.5 (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.

Problem 4.6 (Chebyshev's inequality)**.** Prove the following statement:

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

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

5 Convergence of integrals and functions

5.1 Convergence Theorems

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 4.3.4. However, this required that the sequence f_n of function 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 that have less restrictive conditions.

5.1.1 Monotone convergence (continued)

Theorem 4.3.4 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 5.1.1: 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 $\omega \in \Omega \setminus N$. 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 $\omega \in \Omega \setminus N$ it holds that $g_n(\omega) = f_n(\omega)$ and $g(\omega) = f(\omega)$. Moreover, since $\mu(N) = 0$ we have that

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

The result then follows by applying Theorem 4.3.4 to the functions g_n and g. 😄

5.1.2 Fatou's Lemma

Theorem 5.1.2: 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

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

Proof. Our proof will use the monotone convergence theorem. There are however other proofs, based on first principles. See for example [REF].

Define the function $g_n(\omega) := \inf_{k\ge n} f_k(\omega)$ and note that by Lemma 3.2.13 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 5.1.1 implies that

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

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

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

Here we used that $\inf_{k\ge n} f_k \le f_\ell$ for all $\ell \ge k$ and monotonicity of the integral (see Proposition 4.6.7). 😄

5.1.3 Dominated Convergence

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

Theorem 5.1.3: Dominated convergence

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 $f_n \to f$ point-wise 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 proof the result for the case that both $|f_n| \leq g$ and $f_n \to f$ 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 5.1.2) 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 ot 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 $\omega \in \Omega \setminus N$. Let us now define the following functions

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

and

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

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 5.1. 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 $x \in \mathbb{R} \setminus \{0\}$. Since the set $\{0\}$ has Lebesgue measure zero, all the conditions of Theorem 5.1.3 are satisfied. Hence (see Example 4.1)

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

5.2 Convergence of random variables

5.2.1 Weak convergence of probability measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be μ-integrable. Then we define the measure $\mu * f : \mathcal{F} \to [0, \infty]$ as

$$\mu * f(A) := \int_A f \mathrm{d}\mu \quad \text{for } A \in \mathcal{F}. \tag{5.1}$$

Definition 5.2.4

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* to μ if for every continuous bounded function $f : \mathbb{R} \to \mathbb{R}$ it holds that

$$\mu_n * f \to \mu * f.$$

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

The definition of weak convergence ask us to verify the convergence of the μ_n integral of h for any continuous bounded function h. In some cases that can be 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. There are often summarized in what is known as the Portmanteau lemma (or theorem). We provide one version of it below.

We will first prove an important technical lemma, needed for the proof of this theorem. For any function $h : \mathbb{R} \to \mathbb{R}$ 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 allows us to approximates measurable functions by continuous ones, with arbitrary precision in terms of the integrals.

Lemma 5.2.5

Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $h : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function with $\mu(\mathcal{C}_h) = 1$. Then for every $\varepsilon > 0$, there exist continuous bounded functions h_ε^- and h_ε^+ such that

1. $h_\varepsilon^- \leq h \leq h_\varepsilon^+$ and
2. $\mu * h_\varepsilon^+ - \mu * h_\varepsilon^- < \varepsilon$.

Proof. TODO 😄

Recall that a set $A \subset \mathbb{R}^d$ is open if for every $x \in A$ there exists an $r > 0$ such that $B_x(r) \subset A$. In addition, a set $B \subset \mathbb{R}^d$ is called *closed* if $B = \mathbb{R}^d \setminus A$ for some open set A.

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 $\partial A := \bar{A} \setminus A^\circ$. Given a measure μ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, a set A is called a *μ-continuity set* if $\mu(\delta A) = 0$.

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

Theorem 5.2.6: 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. $\mu_n * h \to \mu * h$ for all bounded measurable functions $h : \mathbb{R} \to \mathbb{R}$ with $\mu(\mathcal{C}_h) = 1$.
3. $\limsup_{n\to\infty} \mu_n(B) \leq \mu(B)$ for all closed sets $B \subset \mathbb{R}$.
4. $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$ for all open sets $A \subset \mathbb{R}$.
5. $\lim_{n\to\infty} \mu(C) = \mu(C)$ for all μ-continuity sets C.

Proof. We will prove the following implication chain: 4 $\iff$ 3 $\Rightarrow$ 1 $\Rightarrow$ 2 $\Rightarrow$ 5 $\Rightarrow$ 3

4 $\iff$ 3: This follows directly since every closed set B is the complement of an open set A, i.e. $B = \mathbb{R} \setminus A$ and thus

$$\limsup_{n\to\infty} \mu_n(B) = \limsup_{n\to\infty} 1 - \mu_n(A) = 1 - \liminf_{n\to\infty} \mu_n(A).$$

3 $\Rightarrow$ 1: Let h be a continuous bounded function. Then, without loss of generality we may assume that $0 \geq h < 1$. Now fix some $k \in \mathbb{N}$ and define the following sets:

$$B_j := \{x \in \mathbb{R} \,:\, \frac{j}{k} \leq f(x)\} \quad \text{for } j = 0, 1, \ldots, k.$$

Note that since f is continuous these are closed sets. Also note that $\mu(B_0) = 1$ and $\mu(B_k) = 0$.

We further observe that $f(x) = \sum_{j=1}^{k} f(x)\mathbf{1}_{B_{j-1}\cap B_j^c}$, where $B_j^c = \mathbb{R} \setminus B_j$. Hence we can now bound the integral $\mu * h$ from above and below as follows:

$$\sum_{j=1}^{k} \frac{j-1}{k} \mu(B_{j-1} \cap B_j^c) \leq \int_\mathbb{R} h \,\mathrm{d}\mu \leq \sum_{j=1}^{k} \frac{j}{k} \mu(B_{j-1} \cap B_j^c). \tag{5.2}$$

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 (5.2) 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 (5.2) equals

$$\frac{1}{k}\sum_{j=1}^{k}\mu(B_j).$$

We thus conclude that for any $k \geq 1$,

$$\frac{1}{k}\sum_{j=1}^{k}\mu(B_j) \leq \int_{\mathbb{R}} h \,\mathrm{d}\mu \leq \frac{1}{k} + \frac{1}{k}\sum_{j=2}^{k}\mu(B_j). \tag{5.3}$$

Moreover, the same inequalities hold for the measures μ_n.

Applying 3 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 $\lim_{n\to\infty} \int_{\mathbb{R}} h \,\mathrm{d}\mu_n = \int_{\mathbb{R}} h \,\mathrm{d}\mu$.

1 ⇒ 2: Fix $\varepsilon > 0$ and let h_ε^- and h_ε^+ be the function from Lemma [REF]. Then

$$\mu * h \leq \mu * h_\varepsilon^+ = \mu * h_\varepsilon^+ - \mu * h_\varepsilon^- + \mu * h_\varepsilon^-,$$

which implies that

$$\mu * h - \varepsilon \leq \mu * h_\varepsilon^-.$$

In a similar way we obtain that

$$\mu * h_\varepsilon^+ \leq \mu * h + \varepsilon.$$

Now we employ condition 1 for the functions h_ε^- and h_ε^+ to get

$$\begin{aligned}
\mu * h - \varepsilon &\leq \mu * h_\varepsilon^- \\
&= \lim_{n\to\infty} \mu_n * h_\varepsilon^- \\
&\leq \liminf_{n\to\infty} \mu_n * h \\
&\leq \limsup_{n\to\infty} \mu_n * h \\
&\leq \mu_n * h_\varepsilon^+ \\
&= \mu * h_\varepsilon^+ \leq \mu * h + \varepsilon.
\end{aligned}$$

From this it follows that

$$\mu * h - \varepsilon \leq \liminf_{n\to\infty} \mu_n * h \leq \limsup_{n\to\infty} \mu_n * h \leq \mu * h + \varepsilon.$$

And since $\varepsilon > 0$ was arbitrary we conclude that

$$\liminf_{n\to\infty} \mu_n * h = \limsup_{n\to\infty} \mu_n * h,$$

which then implies that $\mu_n * h \to \mu * h$.

2 ⇒ 5: 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

$$\mu(\mathcal{C}_h) = \mu(\mathbb{R} \setminus \partial C) = 1 - \mu(\partial C) = 1 - 0 = 1.$$

Hence the function h satisfies the conditions of 2 and thus

$$\mu_n(C) = \int_{\mathbb{R}} h \,\mathrm{d}\mu_n = \mu_n * h \to \mu * h = \int_{\mathbb{R}} h \,\mathrm{d}\mu = \mu(C).$$

5 ⇒ 3: 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

$$\partial A_\delta = \bar{A}_\delta \setminus A_\delta \subset \{x \in \mathbb{R} : \|x - B\| \leq \delta\} \setminus A_\delta = \{x \in \mathbb{R} : \|x - B\| = \delta\}.$$

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 5, which implies that $\mu_n(B_k) \to \mu(B_k)$.

Taking $k \to \infty$ now yields 3.

😄

5.2.2 Convergence in distribution

Definition 5.2.7: 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. It turns out that this definition is yet another equivalent statement for Definition 5.2.7.

Lemma 5.2.8

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. Let $\mu_n := (X_n)_\#\mathbb{P}_n$ and $\mu := X_\#\mathbb{P}_n$.

We will first prove that $X_n \xrightarrow{d} X$ implies $\lim_{n\to\infty} F_n(t) = F(t)$ for all $t \in \mathcal{C}_F$. For this let $h(x) = \mathbf{1}_{(-\infty,t]}$ and note that

$$F_n(t) = (X_n)_\#\mathbb{P}_n((-\infty, x]) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty,t]}\,\mathrm{d}(X_n)_\#\mathbb{P}_n = \mu_n * h.$$

The function h is discontinuous only at t, i.e. $\mathcal{C}_h = \mathbb{R} \setminus \{t\}$. Moreover, for any $\varepsilon > 0$

$$\mu((t-\varepsilon, t+\varepsilon)) = \mu((t-\varepsilon, t]) + \mu((t, t+\varepsilon)) = F(t) - F(t-\varepsilon) + F(t+\varepsilon) - F(t).$$

Since F is continuous at t, the right hand side goes to zero as $\varepsilon \to 0$. Therefore

$$\mu(\{t\}) = \lim_{\varepsilon\to 0} \mu((t-\varepsilon, t+\varepsilon)) = 0,$$

which implies that $\mu(\mathcal{C}_h) = 1$. The result now follows by applying condition 2 in Theorem 5.2.6.

😄

Convergence in distribution to a constant $r \in \mathbb{R}$.

5.2.3 Convergence in probability

Definition 5.2.9: Convergence in probability

Let $(X_n)_{n\geq 1}$ and X be random variables define 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 $X_n \xrightarrow{\mathbb{P}} X$.

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. the ccdf 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 5.2.10

Let $(X_n)_{n\geq 1}$ and X be random variables define on the same probability space. Then $X_n \xrightarrow{\mathbb{P}} X$ if and only if

$$\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0 \quad \text{for every } \varepsilon > 0.$$

5.2.4 Almost-sure convergence

5.3 Problems

Problem 5.1. The goal of this problem is to prove Lemma 5.2.10. 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.

We will first prove the $\Rightarrow$ implication. Write $\mu_n = (X_n)_\#\mathbb{P}_n$ and $\mu = X_\#\mathbb{P}$.

1. Let $t \in \mathbb{R}$. Find a measurable function h_t, such that $F_n(t) = \mu_n * h_t$ and $F(t) = \mu * h_t$.
2. Show that $\mu(\mathcal{C}_{h_t}) = 1$.
3. Prove the $\Rightarrow$ implication.

For the other implication $\Leftarrow$ we will first show that $F_n(t) \to F(t)$ for all $t \in \mathcal{C}_F$ implies that $\mu_n * g \to \mu * g$ for all continuous function that are non-zero only on a bounded and closed set C.

So let g be such a function. You may use the fact that any continuous function that is non-zero on a bounded and closed set is uniformly continuous, i.e. for every $\varepsilon > 0$ there exist a $\delta > 0$ such that $\|x - y\| < \delta$ implies that $\|g(x) - g(y)\| < \varepsilon$.

4. Construct a partition of C into T intervals $I_i = (a_i, b_i]$ such that for each $i \leq T$ and $x, y \in I_i$ it holds that $\|x - y\| < \delta$.

We will now define an approximate function

$$\hat{g}(x) = \sum_{i=1}^{T} \eta_i \mathbf{1}_{I_i},$$

where $\eta_i = h(\max\{x \in I_i\})$.

5. Show that there exists sequences $(b_i)_{1\leq i\leq m}$ and $(t_i)_{1\leq i\leq m}$ such that

$$\tilde{g}(x) = \sum_{i=1}^{m} b_i \mathbf{1}_{(-\infty, t_i]}.$$

6. Prove that $\lim_{n\to\infty} \mathbb{E}[\tilde{g}(X_n)] = \mathbb{E}[\tilde{g}(X)]$. [Hint: Use the assumption $F_n(t) \to F(t)$ for all $t \in \mathcal{C}_F$.]

7. Prove that $\|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]\| \to 0$. [Hint: First use the previous result to show that $\|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]\| \to 2\varepsilon$ by adding and subtracting $\mathbb{E}[\tilde{g}(X_n)]$ and $\mathbb{E}[\tilde{g}(X)]$.]

Now consider a general continuous bounded function h and suppose that $\|h(x)\| \le H$ for all $x \in \mathbb{R}$. Let $\varepsilon > 0$, let $\alpha = \alpha(\varepsilon) > 0$ be such that $\mathbb{P}(\|X\| > \alpha) \le \varepsilon/(2H)$ and define

$$\hat{h}(x) = \begin{cases} x + 1 + \alpha g & \text{if } -\alpha - 1 < x < -\alpha, \\ 1 & \text{if } \|x\| \le \alpha, \\ x + (1 - \alpha) & \text{if } \alpha < \alpha < \alpha + 1, \\ 0 & \text{else.} \end{cases}$$

8. Show that $\mathbb{E}[\hat{h}(X)] \ge 1 - \frac{\varepsilon}{2M}$.

9. Prove that $\left|\mathbb{E}[X_n] - \mathbb{E}[h(X_n)\hat{h}(X_n)]\right| \to \varepsilon/2$ and similarly for X.

10. Show that $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \to \varepsilon$. [Hint: add and subtract both $\mathbb{E}[h(X_n)\hat{h}(X_n)]$ and $\mathbb{E}[h(X)\hat{h}(X)]$ and use the previous result and the one show in 7.]

11. Conclude that $\mathbb{E}[h(X)_n)] \to \mathbb{E}[h(X)]$.

6 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

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

6.1 The Hölder inequality

Proposition 6.1.1: Hölder's inequality

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

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

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

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

6.2 The Minkowski inequality

Proposition 6.2.2

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. 😄

6.3 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 6.3 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

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

6.4 Completeness of L^p-spaces

Theorem 6.4.3: 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

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

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

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

For such x, 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 $N_1 \in \mathbb{N}$ such that for all $m, n \geq N_1$,

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

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

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

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

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

which gives the required convergence.

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

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

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

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

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. 😄

6.5 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 6.5.4: 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 6.5.5

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 6.5.6

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 let $\epsilon > 0$, 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 6.5.5 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 6.5.7: 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 6.5.8: 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

$$A_k^n := \left\{\omega \in \mathbb{R}^d : \ (k-1)2^{-n} < f(\omega) \leq k2^{-n}\right\}.$$

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 6.5.9: 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 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 $r > 0$.

6.6 Problems