TU/e, 2MBA70

Measure and Probability Theory

figures/TUe_logo.png

Pim van der Hoorn and Oliver Tse

Version 0.1 July 10 2024

Quote about measure theory?

– Someone

Disclaimer:

These are lecture notes for the course *Measure and Probability Theory*. They are by no means a replacement for the lectures or the books, nor are they intended to cover every aspect of the field of measure theory of 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. All comments and suggestions will be greatly appreciated.

Contents

1. Introduction

2. Sigma-algebras, measures and measurable functions

2.1. Recalling basic probability theory

During the first course on probability theory, Probability and Modeling (2MBS10), the concept of probabilities were 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 that represent the *probability* that the experiment will yield an outcome in this set A. This value was then denoted by $\mathbb{P}(A)$.

It turned out that in order to properly 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 we not only do we need to be able compute $\mathbb{P}(A \cup B)$, we actually 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 compute 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 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 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

In addition, the probability assignment $\mathbb{P}$ was defined as a map $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

1. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$, and
2. for any collection $A_1, A_2, \dots$ of disjoint events in $\mathcal{F}$ it holds that

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

With this setup it was possible to formally 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 of how to define a discrete and a continuous random variable. In addition, a separate definition was required to defined multivariate distribution functions. This limits the extend to which this theory can be applied. For example let U have the uniform distribution on $[0, 1]$ and Y have uniform distribution on the set $\{1, 2, \ldots, 10\}$ and define the random variable X to be equal to U with probability $1/2$ and equal to Y with probability $1/2$. How would you deal with this random variable, which is both discrete and continuous? However, the setting would becomes even more complex if we are not talking about random numbers in $\mathbb{R}$ but, say, random vectors of infinite length or random functions. Do these even exist?

The solutions to all these issues comes from a generalization of event spaces and probability measures introduced above. These go by the names sigma-algebra and measure, respectively. With this we can then define when any mapping between spaces is *measurable* and use such mappings to define random objects in that space such a function maps to. The remainder of this chapter is dedicated to introduced all these concepts.

2.2. Sigma-algebras

2.2.1. 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 2.2.1: Sigma Algebra

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

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$;
2. For every $A \in \mathcal{F}$, it holds that $A^c := \Omega \setminus A \in \mathcal{F}$;
3. For every sequence $A_1, A_2, \cdots \in \mathcal{F}$, 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 looks very similar to that of an event space. It turns out that they are actually the same, see Problem 2.1.

Before we give some examples, we first show that any σ-algebra is also closed under countable intersections. The proof is left as an exercise to the reader (see Problem 2.2).

Lemma 2.2.2

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

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

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

1. The collection $\mathcal{F} = \{\emptyset, \Omega\}$ is a σ-algebra. This is called the *trivial σ-algebra* or the *minimal* σ-algebra on Ω.
2. For any subset $A \subset \Omega$ we have that $\mathcal{F} := \{\emptyset, A, \Omega \setminus A, \Omega\}$ is a σ-algebra.
3. 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.
4. For any subset $A \subset \Omega$ such that $A \neq \emptyset, \Omega$, we have that $\mathcal{F} := \{\emptyset, A, \Omega\}$ is **not** a σ-algebra.
5. 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.
6. 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 2.3.

The idea of measure theory is that one can assign a notion of 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.*

2.2.2. Constructing σ-algebras

We now know what a σ-algebra is and have seen some example 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 2.2.3: Restriction of a σ-algebra

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

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

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

Proof. We need to check all three properties.

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

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

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

😄

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

Proposition 2.2.4: Generated σ-algebra

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

Proof. Similar to Lemma 2.2.3 we need to check all the requirements.

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

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

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

😄

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* $\mathcal{F}$.

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

Lemma 2.2.5: Inclusion property of σ-algebras

If $A \subset B \subset C$ are subset of Ω, then also $\sigma(A) \subset \sigma(B) \subset \sigma(C)$.

Using this powerful construction tool for σ-algebras, we can now construct products of measurable spaces.

Definition 2.2.6: Product σ-algebra

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

2.2.3. 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, in the introduction we noticed that the concept of random variables, as given in the course Probability and Modeling, is mainly concerned with $\mathbb{R}$. 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, who was 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 2.2.7: Borel σ-algebra

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

Remark. From the definition, it should be clear that one can 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.

Proposition 2.2.8

The Borel σ-algebra on $\mathbb{R}^d$ is the σ-algebra generated by the sets

$$(-\infty, a_1] \times \cdots \times (-\infty, a_d] \quad \text{with } a_i \in \mathbb{Q}, i = 1, \dots, d.$$

2.3. Measures

2.3.1. 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 2.3.9: 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 measure 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.*

It should be noted that, outside maybe the discrete measure, these examples 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 basis. 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, that the measure of any rectangle is simply its volume, which seems like a very natural property to ask for.

Ideally, we would like to be able to

1. construct a measure by defining it only on the basis of a σ-algebra, and
2. extend any *reasonable* set function with a given property on a set to a measure on the generated σ-algebra.

Luckily, it turns out that both are possible. However, for this we need to consider how unique measures are when defined only on a basis of a σ-algebra and develop a tool that allows us to extend functions on sets to measures on the generated σ-algebra. These two topics will be covered in Section 2.3.3.

Before we go there, let us first study some important properties of measures.

2.3.2. Important properties

Although the number of properties a measure needs to satisfy are 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 2.3.10: Basic properties of measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $A, B \in \mathcal{F}$. 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. First note that if $\mu(A \cap B) = \infty$ then by property 2 we have that also $\mu(A), \mu(B)$ and $\mu(A \cup B) = \infty$ and hence the result holds trivially. So assume now that $\mu(A \cap B) < \infty$. Since
$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B),$$
it follows from property 1 that
$$\mu(A \cup B) = \mu(A \setminus (A \cap B))) + \mu(A \cap B) + \mu(B \setminus (A \cap B)).$$
Adding $\mu(A \cap B) < \infty$ to both side we get
$$\begin{aligned} \mu(A \cup B) + \mu(A \cap B) &= \mu(A \setminus (A \cap B))) + \mu(A \cap B) + \mu(B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A) + \mu(B), \end{aligned}$$
where the last line follows from applying property 3 twice.

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 2.3.11: 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*.

The proof of this lemma is left as an exercise, see Problem [REF].

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 2.3.12: 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. TODO 😄

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 2.3.13: 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. TODO 😄

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 2.3.14: Alternative definition of a measure

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

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

Proof. TODO 😄

2.3.3. Uniqueness and existence of measures

In Section 2.2.2 we discussed that in order to define measures on the Borel space we need to be able to construct measure on a σ-algebra $\sigma(\mathcal{A})$ by defining them on the generator $\mathcal{A}$. In particular, we need that if to 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 they should agree on the entire σ-algebra. The purpose of this section is to provide two key theorems that show that both these things are possible. We state them here without proof, as this would require the introduction of several different concepts. The full proof is provided in the Appendix. What we will do is utilize both results to construct a canonical measure on the Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, called the *Lebesgue measure*.

We start with showing that measures that agree on the generators of σ-algebras agree on the entire σ-algebra, under some small conditions on the generator set.

Theorem 2.3.15: Uniqueness of measures

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

1. for any $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, and

2. there exists a sequence $(A_i)_{i\in\mathbb{N}}$ with $\Omega = \bigcup_{i\in\mathbb{N}} A_i$.

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

The main implication of Theorem 2.3.15 is that it makes sense to construct a measure on $\sigma(\mathcal{A})$ by defining it on the generator $\mathcal{A}$. But how can we construct a measure on this set? In particular, is it possible to start with a set function that does not satisfy all the properties of a measure? We will address these questions next. But in order to do so we need to introduce the notion of an *algebra*.

Definition 2.3.16: Algebra's of sets

A collection $\mathcal{A}$ of subsets of Ω is called an *algebra* if

1. $\emptyset \in \mathcal{A}$,
2. $\Omega \setminus A \in \mathcal{A}$ for all $A \in \mathcal{A}$, and
3. $A \cup B \in \mathcal{A}$ for every $A, B \in \mathcal{A}$.

Note that every σ-algebra is an algebra. The idea is that is we start with a set function on an algebra, we can extend this to all the way to a measure on σ-algebra. To ensure this extension is possible, we need to start with set functions that have some structure, suspiciously called premeasures.

Definition 2.3.17: Premeasures

Let $\mathcal{A}$ be an algebra on Ω. A set function $\mu_o : \mathcal{A} \to [0, \infty]$ is called a *premeasure* if

1. $\mu_o(\emptyset) = 0$, and
2. μ_o is σ-additive.

If we start with a premeasure μ_o on an algebra $\mathcal{A}$ we can construct a new set function on the entire collection of subsets of Ω.

Definition 2.3.18: Outer measure

Let μ_o be a premeasure on an algebra $\mathcal{A}$ on Ω. Then the set function μ^* defined by

$$\mu^*(A) := \inf\left\{\sum_{i=1}^{\infty} \mu_o(A) \,:\, A \subset \bigcup_{i\in\mathbb{N}} A_i,\ A_i \in \mathcal{A}\right\},$$

is called the *outer measure induced by* μ_o.

The idea is that the outer measure μ^* is almost a measure. This is captured by the following set of properties it has.

Proposition 2.3.19

Let μ_o be a premeasure on an algebra $\mathcal{A}$ on Ω and μ^* be the outer measure induced by μ_o. Then μ^* satisfies the following properties:

1. $\mu^*(A) = \mu_o(A)$ for all $A \in \mathcal{A}$,
2. $\mu^*(\emptyset) = 0$ and $\mu^*(A) \geq 0$ for all $A \subset \Omega$,
3. μ^* is monotone, and
4. μ^* is σ-subadditive.

Proof. TODO ☺

Observe that indeed, μ^* is almost a measure. The only property missing is full σ-additivity. Then next fundamental result, due to the Greek mathematician Constantin Carathéodory, provides a way to construct a σ-algebra from a given algebra such that μ^* can be extended to a true measure on it. We state a partial version here, without proof.

Theorem 2.3.20: Carathéodory's extension theorem (partial)

Let $\mathcal{A}$ be an algebra on Ω. Let μ_0 be a pre-measure on $\mathcal{A}$ and denote by μ^* the outer measure induced by μ_0. Then the collection defined by

$$\mathcal{A}_{\mu^*} := \{B \subset \Omega \,:\, \mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \setminus B) \,\forall A \in \mathcal{A}\},$$

is a σ-algebra on Ω. Moreover, the restriction $\bar{\mu} := \mu^*|\mathcal{A}_{\mu^*}$ of μ^* to $\mathcal{A}_{\mu^*}$ is a measure on $\mathcal{A}_{\mu^*}$ called the *Carathéodory extension of* μ_o.

At this point we should take some time to fully appreciate what Theorem 2.3.20 gives us. In order to construct a measure all we need is an algebra on Ω and some premeasure.

Remark. The statement in Theorem 2.3.20 only covers part of the original theorem. It actually turns out that the σ-algebra constructed has some very nice properties and the measure space $(\Omega, \mathcal{A}_{\mu^*}, \bar{\mu})$ is *complete*. However, in order to properly define these notions we needed to introduce additional concepts going beyond the goal of this section. The interested reader is referred to the Appendix for the full statement and details, including the proof of this theorem.

Let us now utilize the Carathéodory extension to obtain a measure on the Borel space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

2.4. Measurable functions

2.5. Random variables and general stochastic objects

2.6. Problems

Problem 2.1. Show that the definition of an *event space* as given in Section 2.1 is equivalent to that of a σ-algebra as given in Definition 2.2.1.

Problem 2.2. Prove Lemma 2.2.2.

Problem 2.3. 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'.

3. The Lebesgue Integral

3.1. The integral of a simple function

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

Definition 3.1.1

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

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

for some positive integer $N \in \mathbb{N}$, disjoint measurable sets $A_1, \ldots, A_N \in \mathcal{F}$ and constants $a_1, \ldots, a_N \in \mathbb{R}$. For a non-negative simple function

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

we define the integral of g with respect to μ by

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

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

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

By the finite additivity of the measure μ,

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

Remark. 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

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

3.2. The Lebesgue integral of nonnegative functions

Definition 3.2.2

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

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

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

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

Proposition 3.2.3: Properties of the Lebesgue integral of nonnegative functions

Let f, g be two nonnegative, measurable functions and $\lambda \ge 0$ 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)

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

3.3. The monotone convergence theorem

Theorem 3.3.4: Monotone convergence theorem I

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

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

😄

3.4. 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.

Problem 3.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : (\Omega, \mathcal{F}) \to ([0, +\infty), \mathcal{B}_{[0,+\infty)})$ be a nonnegative measurable function. Show that

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

3.5. Additivity of the Lebesgue integral of nonnegative functions

Proposition 3.5.5: Additivity of the Lebesgue integral of nonnegative functions

Let f, g be two nonnegative 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 for every $n \in \mathbb{N}$,

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

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

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

On the other hand, by a limit theorem and Problem 3.1, 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,

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

😄

3.6. 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 3.6.6

A 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

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

Proposition 3.6.7

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 $f \le g$ μ-a.e., then
$$\int_\Omega f \,\mathrm{d}\mu \le \int_\Omega g \,\mathrm{d}\mu.$$
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 3.6.8

We say that a measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ is integrable on a set $A \in \mathcal{F}$ if 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 $(A, \mathcal{F}_A, \mu|_A)$.

3.7. Change of variables formula

Proposition 3.7.9

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ and $h : (E, \mathcal{G}) \to ([0, \infty], \mathcal{B}_{[0,\infty]})$ be measurable maps. 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 3.7.10

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f : (\Omega, \mathcal{F}) \to (E, \mathcal{G})$ and $h : (E, \mathcal{G}) \to (\mathbb{R}, \mathcal{B})$ be measurable maps. 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,

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

Example 3.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be a continuous random variable. Assume that the law of X can be represented by a *Lebesgue density function* $\varrho : (\mathbb{R}, \mathcal{B}) \to ([0, \infty), \mathcal{B}_{[0,\infty)})$, i.e. $X_\# \mathbb{P}$ is the (unique) measure given by

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

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

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

On the other hand, the set function

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

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

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

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

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

and thus

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

3.8. The Markov inequality

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

Lemma 3.8.11: 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 3.8.12: The Markov inequality

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

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

A. Appendix

Bibliography