TU/e, 2MBA70

Measure and Probability Theory

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. Measurable spaces (sigma-algebras and measures)

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 $\{1, 2, \dots, 10\}$ 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 $A_1, A_2, \dots \in \mathcal{F}$, 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 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 $A_1, A_2, \dots \in \mathcal{F}$. 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 any of the following family of sets,

1. $(-\infty, a_1] \times \cdots \times (-\infty, a_d]$,
2. $(-\infty, a_1) \times \cdots \times (-\infty, b_d)$,
3. $[a_1, \infty) \times \cdots \times [a_d, \infty)$,
4. $(a_1, \infty) \times \cdots \times (a_d, \infty)$,

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

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*.

2.3.2. 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 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.

It turns out that such a measure exists. This fundamental measure is called the *Lebesgue measure* and, in addition to the property mentioned above, has several other strong features.

Theorem 2.3.10: Lebesgue measure

There exists a unique measure λ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ with 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 $M \in \mathbb{R}^{d \times d}$, $\lambda(M^{-1}(B)) = |\det M|^{-1} \lambda(B)$.

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.

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

2.3.3. 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 2.3.11: 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 2.3.12: 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 2.3.13: 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 2.3.14: 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
$$\lim_{i\to\infty} \mu(A_i) = \mu\left(\bigcap_{i\in\mathbb{N}} A_i\right).$$

The proof of this proposition is similar to that of Proposition 2.3.13 and is left as an exercise, see Problem 2.4.

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 2.3.15: 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 2.3.11 and 2.3.13. 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

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

😄

In Section 2.2.2 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 2.3.16: 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 2.3.16: to study a measure on $\sigma(\mathcal{A})$ it suffices to look at what it does on the generator $\mathcal{A}$.

2.4. 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'.

Problem 2.4. Prove Proposition 2.3.14.

3. Measurable functions and stochastic objects

3.1. Measurable functions

Now that we have defined measure spaces $(\Omega, \mathcal{F}, \mu)$, through σ-algebras and measures and studied properties of both these objects, it is time to look at functions between such spaces. We will focus on functions that preserve the measurable structure of the spaces.

The main object in analysis were *continuous* function $f : \mathbb{R}^d \to \mathbb{R}^m$. This property was important, as it allowed us to differentiate the function and perform integration.

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. The same holds for extensions.

Proposition 3.1.3: Restriction and extension 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.

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

$$g(\omega) := \begin{cases} g_A(\omega) & \text{if } \omega \in A, \\ p & \text{if } \omega \notin A, \end{cases}$$

is $(\mathcal{F}, \mathcal{G})$-measurable.

Proof. TODO 😄

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 2.2.5 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 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 2.2.4, $\sigma(f_i \,:\, i \in I)$ is a σ-algebra. We will show that any σ-algebra that makes each f_i measurable much contain $\sigma(f_i \,:\, i \in I)$. So let $\mathcal{F}$ be such a σ-algebra. Then in particular, for any $i \in I$ and $B \in \mathcal{F}_i$ we have that $f_i^{-1}(B) \in \mathcal{F}$. This implies that

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

Now since $\sigma(f_i \,:\, i \in I)$ is generated by the collection on the left hand side, Lemma 2.2.5 implies that

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

😄

Similar to Lemma 3.1.4, 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 2.2.6. There is an alternative way to construct it using functions. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces and consider the functions $\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i$, defined by

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

These are called the *canonical projections*. Following Proposition 3.1.6 we can construct the σ-algebra $\sigma(\pi_1, \pi_2)$ on $\Omega_1 \times \Omega_2$, which makes both canonical projections measurable. Then it follows that, see Problem 3.1,

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

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 3.2.

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 (see [REF]). 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 ∞ be a part of the real line (which it is not). So we first extend the real line to include both ∞ and $-\infty$.

3.2.1. Extended real line

We define $\bar{\mathbb{R}} := [-\infty, \infty]$ as the *extended real line*. We impose the natural ordering on $\bar{\mathbb{R}}$, 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, with are extend 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 $x \in (0, \infty]$, $\pm x(\infty) = (\infty) \pm x = \pm\infty$, $\pm x(-\infty) = (-\infty) \pm x = \mp\infty$,
4. $0(\pm\infty) = (\pm\infty)0 = 0$ and $1/\pm\infty = 0$.

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

Definition 3.2.9: Extended real line

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

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

The following results, whose proof is left as an exercise, relates $\bar{\mathcal{B}}$ to the original Borel σ-algebra.

Lemma 3.2.10

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

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

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

Proof. See Problem [REF] 😄

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

$$\{f < a\}, \{f \le a\}, \{f > a\}, \{f \ge a\}, \{f = a\} \text{ and } \{f \neq a\},$$

are also measurable.

Proof. Since f is measurable, it follows immediately from Proposition 2.2.8 and Lemma 3.1.4 that $\{f < a\}, \{f \le a\}, \{f > a\}, \{f \ge a\} \in \mathcal{F}$. This then implies the other two claims since $\{f = a\} = \{f \le a\} \setminus \{f < a\}$ 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 [REF].

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

$$\{f \ge g\} = \bigcup_{q \in \mathbb{Q}} \{f \ge q\} \cap \{g \le q\},$$

while

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

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

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

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

Lemma 3.2.10 $\bar{\mathcal{B}}$ 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 bare 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:

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

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

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

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

3.2.3. Limit operations

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

Lemma 3.2.13

Let $(f_n)_{n \geq 1}$ be a family of measurable functions from $(\Omega, \mathcal{F})$ to $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$. 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 [REF].

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

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

Note that if this holds then $\{\sup_{n\geq 1} f_n > x\} \in \mathcal{F}$ since each set $\{f_n > a\}$ is measurable by Lemma 3.2.11 and hence $\{\sup_{n\geq 1} f_n > x\}$ (check this yourself, see Problem [REF]).

Since $a < f_n(\omega) \leq \sup_{n\geq 1} f_n(\omega)$ 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

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

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 actual very important. In particular is 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 we 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

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

Let us make two observations here. The first is that the set above is simply the preimage $X^{-1}((-\infty, x])$. Secondly, the sets $(-\infty, x]$ generate the Borel σ-algebra. Thus it follows from Lemma 3.1.4 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

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

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).

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

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 $\mathbb{P}(X \le t) = F(t)$, 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 $0 \le t \le 1$. 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

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

Since by Theorem 2.3.10

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

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

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

😄

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

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

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 is clearly 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 monotonic increasing 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 (Equivalence of product σ-algebra)**.** Prove equation (3.1).

Problem 3.2 (Push-forward measure)**.** Prove Proposition 3.1.8.

4. The Lebesgue Integral

4.1. The integral of a simple function

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

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 $a_1, \dots, a_N \in \mathbb{R}$. 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

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

4.2. The Lebesgue integral of nonnegative functions

Definition 4.2.2

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 4.2.3: 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)
$$\lambda \int_\Omega f \,\mathrm{d}\mu = \int_\Omega (\lambda f) \,\mathrm{d}\mu.$$

4.3. The monotone convergence theorem

Theorem 4.3.4: 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

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

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

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

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

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

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

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

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

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

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 4.1. 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

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

4.5. Additivity of the Lebesgue integral of nonnegative functions

Proposition 4.5.5: 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 4.1, 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,

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

😄

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 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 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 $f \le g$ 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 4.6.8

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

4.7. Change of variables formula

Proposition 4.7.9

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 4.7.10

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 4.1. 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

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

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

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

4.8. 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 4.8.11: 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 4.8.12: 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