THEORY OF MEASURE AND INTEGRATION

0 Introduction

Probability theory looks back to a history of almost 300 years. Indeed, J. Bernoulli's law of large numbers, which was published post mortem in 1713 can be considered the mother of all probabilistic theorems. In those days, and even almost 200 years later, mathematicians had a fairly heuristic idea about probability. So the probability of an event usually was understood as the limit of the relative frequencies of a series of independent trials "under usual circumstances". This apparently coincides with both, the naive intuition of what probability is, as well as with the prediction of the law of large numbers. On the other hand, it is not at all easy to work with such a definition of probability, nor is it simple to make it mathematically rigorous.
In 1900 the German mathematician D. Hilbert was invited to give a plenary lecture on the World Congress Of Mathematics, that was being held in Paris. There he introduced his famous 23 problems in mathematics. Those problems triggered the development of the mathematics in the next 50 years. Even today some of those questions are still wide open. His sixth question was:

"Give an axiomatic approach to probability theory and physics".

Of course the axiomatization of physics is tremendously difficult and still unsolved. The question of giving an axiomatic foundation of probability theory was approached in the 1930's by the Russian mathematician A.N. Kolmogorov. He linked probability to the then relatively new theory of measures by defining a probability to be a measure with mass 1 on the set of outcomes of a random experiment.
This theory of measure and integration on the other hand had started to develop in the middle of the 19th century. Until then the only integrable functions that were known were the continuous mappings from $\mathbb{R}$ to $\mathbb{R}$. It was not until B. Riemann's Habilitation-Thesis in 1854 that the corresponding definition of an integral (which went back to A. Cauchy) was extended to

certain non-continuous functions. Yet the Riemann integral has two decisive drawbacks :

1. Certain non - continuous functions, which we would like to equip with an integral, are not Riemann-integrable. One of the most famous examples was given by P. G. L. Dirichlet:

$$\delta_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases}$$

   Considering $\delta_{\mathbb{Q}}$ as a function from $[0,1]$ to $[0,1]$, its integral would give the "size" of $\mathbb{Q}$ in $[0,1]$, and therefore is interesting.

2. The rules for interchanging limits of sequences of functions with the integral are rather strict. Recall that if $f_n, f : \mathbb{R} \to \mathbb{R}$ are Riemann - integrable functions and

$$f_n(x) \to f(x) \text{ as } n \to \infty$$

   for all x, we know that

$$\int f_n(x)dx \to \int f(x)dx$$

   only if

$$\sup_{x \in \mathbb{R}} |f_n(x) - f(x)| \to 0.$$

   This obstacle was overcome by E. Borel and H. Lebesgue at the beginning of the 20th century. They found a system of subsets of $\mathbb{R}$ (the so -called Borel $\sigma-$algebra) which they could assign a "measure" to, that agrees on intervals with their length. The corresponding integral integrates more functions than the Riemann-integral and is more liberal concerning interchanging limits of functions with integral-signs.

In the following 30 years the concepts of σ-algebra, measure, and integral were generalized to arbitrary sets. Thus A. N. Kolmogorov could rely on solid foundations, when he linked probability theory to measure theory in the early 1930's.

In this course we will give the basic concepts of measure theory. We will show how to extend a measure from some system of subsets of a given set to a much

larger family of subsets. The idea here is that for a small system of sets, such as the intervals in $\mathbb{R}$, we have an intuitive idea what their measure is supposed to be (namely their length in the example). But if we know the measure of such sets, we also know it for disjoint unions, complements, intersections, etc. This will lead to a whole class of measurable sets. After that we will construct an integral that is based on this new concept of measure.
In the case that the underlying set is $\mathbb{R}$ the new integral (which is then also called the Lebesgue - integral) will be seen to be "more powerful" than the Riemann - integral. The new measures and integrals on arbitrary sets give to new concepts for the convergence of a sequence of functions to a limit. These concepts will be discussed and compared to each other.

Already in a first course in probability one learns that measure ν on $\mathbb{R}$ are particularly nice, if there is a function

$$h : \mathbb{R} \to \mathbb{R}^+_{\{0\}}$$

such that

$$\nu(A) = \int_A h(x)dx, \qquad A \subseteq \mathbb{R}.$$

h then is called a density. We will see in a more general context, when such densities exist.

Also in probability one learns that the most relevant case is not the case of just one experiment but that of a sequence of experiments that do not influence each other and have the same probability mechanism.

This gives rise to several questions:

- How do we extend a measure ν on a set S to a measure $\nu^{\otimes n}$ on S^n?
- How can we integrate with respect to such a measure? (Intuitively we would like to first integrate the first variable, then the second etc. Fubini's theorem say that this is the right tactics).
- Are there infinite sequences of independent trials of a random experiment? Can we play "heads and tails" infinitely often, i.e. can we give a meaning to $\nu^{\otimes\infty}$?

These questions will be answered in the last section. As can be seen the interest in measure theory can be driven by different forces. First of all the theory of measure and integration is an important step in the development of modern analysis. Concepts as Lebesgue - measure or Lebesgue - integral belong to the tool box of every modern mathematician. Moreover measure theory is intrinsically linked to probability theory. This in turn is the root of many other areas, such as statistical mechanism, statistics, or mathematical finance.

1 σ-Algebras and their Generators, Systems of Sets

In this section we are going to discuss the form of the systems of subsets of a given set Ω on which we want to define a measure. Since we would like this system of sets as large as possible (we want to measure as many sets as possible) the most natural choice would be the power set $\mathcal{P}(\Omega)$. We will later see that this choice is not always possible. Hence we ask for the minimum requirements a system of sets $\mathcal{A} \subset \mathcal{P}(\Omega)$ is supposed to fulfill:
Of course, we want to measure the whole set Ω. Moreover, if we can measure $A \subset \Omega$, we also want to measure its complement A^c. Finally, if we can determine the size of a sequence of sets $(A_n)_{n\in\mathbb{N}}$, $A_n \subset \Omega$, we also want to know the size of $\bigcup_{n\in\mathbb{N}} A_n$.

This leads to

Definition 1.1 *A system $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called a σ - algebra over Ω, if*

$$\Omega \in \mathcal{A} \tag{1.1}$$

$$A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \tag{1.2}$$

$$\textit{If } A_n \in \mathcal{A} \textit{for } n = 1, 2, \ldots \textit{ then also } \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}. \tag{1.3}$$

Example 1.2 *1. $\mathcal{P}(\Omega)$ is a σ-algebra.*

2. *Let $\mathcal{A}$ be σ-algebra over Ω and $\Omega' \subseteq \Omega$, then*

$$\mathcal{A}' := \{\Omega' \cap A : A \in \mathcal{A}\}$$

is a σ - algebra over Ω'.

3. *Let Ω, Ω' be sets and $\mathcal{A}$ a σ-algebra over Ω. Let*

$$T : \Omega \to \Omega'$$

be a mapping. Then

$$\mathcal{A}' := \left\{A' \subset \Omega' : T^{-1}[A'] \in \mathcal{A}\right\}$$

is a σ-algebra over Ω'.

Exercise 1.3 *Prove Example 1.2.3.*

Exercise 1.4 *In the situation of Example 1.2.3. consider the system*

$$T\left[\mathcal{A}\right] := \{T(A) : A \in \mathcal{A}\}.$$

Is this also a σ - algebra over Ω'?

Exercise 1.5 *Let I be an index set and $\mathcal{A}_i, i \in I$ be σ - algebras over the same set Ω. Show that*

$$\bigcap_{i \in I} \mathcal{A}_i$$

is also a σ-algebra.

Exercise 1.6 *Show that in general the union of two σ-algebras over the same set Ω, i.e.*

$$\mathcal{A}_1 \cup \mathcal{A}_2 := \{A \in \mathcal{A}_1 \text{ or } A \in \mathcal{A}_2\}$$

is not a σ-algebra.

Corollary 1.7 *Let $\mathcal{E} \subset \mathcal{P}(\Omega)$ be a set system. Then there exists a smallest σ-algebra $\sigma(\mathcal{E})$, that contains $\mathcal{E}$.*

Proof. Consider
$$\mathcal{S} := \{\mathcal{A} \text{ is a } \sigma\text{-algebra}, \mathcal{E} \subset \mathcal{A}\}$$
Then
$$\sigma(\mathcal{E}) = \bigcap_{\mathcal{A} \in \mathcal{S}} \mathcal{A}$$
is a σ - algebra. Obviously $\mathcal{E} \subset \sigma(\mathcal{E})$ and $\sigma(\mathcal{E})$ is smallest possible. ∎

If $\mathcal{A}$ is a σ - algebra and $\mathcal{A} = \sigma(\mathcal{E})$ for some $\mathcal{E} \subset \mathcal{P}(\Omega)$, $\mathcal{E}$ is called **a generator** of $\mathcal{A}$. Often we will consider situations where $\mathcal{E}$ already possesses some of the structure of a σ - algebra. We will give those separate names:

Definition 1.8 *A system of sets $\mathcal{R} \subset \mathcal{P}(\Omega)$ is called a ring, if it satisfies*
$$\emptyset \in \mathcal{R} \tag{1.4}$$
$$A, B \in \mathcal{R} \Longrightarrow A \backslash B \in \mathcal{R} \tag{1.5}$$
$$A, B \in \mathcal{R} \Longrightarrow A \cup B \in \mathcal{R} \tag{1.6}$$
If additionally
$$\Omega \in \mathcal{R} \tag{1.7}$$
then $\mathcal{R}$ is called an algebra.

Note that for every $\mathcal{R}$ that is a ring and $A, B \in \mathcal{R}$
$$A \cap B = A \backslash (A \backslash B) \in \mathcal{R}$$

Theorem 1.9 *$\mathcal{R} \subset \mathcal{P}(\Omega)$ is an algebra, if and only if (1.1),(1.2) and (1.6) are fulfilled.*

Proof. By definition an algebra has properties (1.1) and (1.6). (1.2) follows from (1.5).
The converse follows from
$$A \backslash B = A \cap B^c = (A^c \cup B)^c,$$
and $\emptyset = \Omega^c$. ∎

Exercise 1.10 *Consider for a set* Ω
$\mathcal{A} = \{A \subset \Omega, A \text{ is finite or } A^c \text{is finite}\}$.
Show that $\mathcal{A}$ *is an algebra, but not a* σ*-algebra for infinite* Ω.

Sometimes it is difficult to determine whether a given system of sets is a σ- algebra or not. The following notion goes back to Dynkin and helps to resolve these problems.

Definition 1.11 *A system* $\mathcal{D} \subset \mathcal{P}(\Omega)$ *is called a Dynkin - system, if it satisfies*

$$\Omega \in \mathcal{D} \tag{1.8}$$

$$D \in \mathcal{D} \quad \Longrightarrow D^c \in \mathcal{D} \tag{1.9}$$

For every sequence $(D_n)_{n\in\mathbb{N}}$ *of pairwise disjoint sets in* $\mathcal{D}$*, their union* $\bigcup_{n\in\mathbb{N}} \mathcal{D}_n$ *is also in* $\mathcal{D}$. (1.10)

Example 1.12 1. *Every* σ *- algebra is a Dynkin - system.*

2. *Let* $|\Omega|$ *be finite and* $|\Omega| = 2n, n \in \mathbb{N}$.

   Then

$$\mathcal{D} = \{D \subset \Omega, |D| \text{ is even}\}$$

   is a Dynkin - system. If $n > 1$, $\mathcal{D}$ *is not an algebra, hence also not a* σ *- algebra.*

We will now try to work out the connection between Dynkin - system and σ - algebras.

Lemma 1.13 *If* $\mathcal{D}$ *is a Dynkin-system then*

$$D, E \in \mathcal{D}, D \subset E \Longrightarrow E \setminus D \in \mathcal{D} \tag{1.11}$$

Proof. Note that $D \cap E^c = \emptyset$. Thus $(D \cup E^c)^c = E \cap D^c = E \backslash D \in \mathcal{D}$. ■
We are now ready to prove

Theorem 1.14 *A Dynkin - system $\mathcal{D}$ is a σ - algebra if and only if for any two $A, B \in \mathcal{D}$ we have*

$$A \cap B \in \mathcal{D} \tag{1.12}$$

Proof. First note that if $\mathcal{D}$ is a σ-algebra and $A, B \in \mathcal{D}$, then

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{D}.$$

On the other hand any Dynkin-system satisfies (1.1), and (1.2). Suppose that it moreover satisfies (1.12), and that $D_1, D_2, D_3, \ldots \in \mathcal{D}$. Write

$$D_n' := \bigcup_{i=1}^{n} D_i.$$

The sequence $(D_n')_n$ is increasing. According to (1.11) the sets $D_n' \setminus D_{n-1}' = D_n' \setminus (D_n' \cap D_{n-1}')$ belong to $\mathcal{D}$. But setting $D_0' = \emptyset$ we obtain

$$\bigcup_{n=1}^{\infty} D_n = \bigcup_{n=1}^{\infty} \left(D_n' \backslash D_{n-1}'\right) \in \mathcal{D}$$

the latter because the sets $\left(D_n' \backslash D_{n-1}'\right)$ are pairwise disjoint. ■
Similar to the case of σ-algebras for every system of sets $\mathcal{E} \subset \mathcal{P}(\Omega)$ there is a smallest Dynkin-system $\mathcal{D}(\mathcal{E})$ generated by (and containing) $\mathcal{E}$. The importance of Dynkin system mainly is due to the following

Theorem 1.15 *For every $\mathcal{E}$, with*

$$A, B \in \mathcal{E} \Longrightarrow A \cap B \in \mathcal{E}$$

we have

$$\mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E}).$$

Proof. Since every σ-algebra is a Dynkin-system and $\sigma(\mathcal{E})$ contains $\mathcal{E}$, we see that

$$\mathcal{D}(\mathcal{E}) \subseteq \sigma(\mathcal{E}).$$

On the other hand, if we knew that $\mathcal{D}(\mathcal{E})$ was a σ-algebra, we would have that also

$$\sigma(\mathcal{E}) \subseteq \mathcal{D}(\mathcal{E}).$$

Following Theorem 1.14 we only need to prove that $\mathcal{D}(\mathcal{E})$ is $\cap$-stable, i.e. that with any two sets it contains its intersection. To show this for $D \in \mathcal{D}(\mathcal{E})$ put

$$\mathcal{D}_D = \{Q \in \mathcal{P}(\Omega) : Q \cap D \in \mathcal{D}(\mathcal{E})\} \tag{1.13}$$

One easily verifies that $\mathcal{D}_{\mathrm{D}}$ is a Dynkin - system (see the exercise below). For each $E \in \mathcal{E}$ we know from the conditions on $\mathcal{E}$ that $\mathcal{E} \subset \mathcal{D}_E$ and hence $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_E$. But this shows that for each $D \in \mathcal{D}(\mathcal{E})$ and each $E \in \mathcal{E}$ we have that $E \cap D \in \mathcal{D}(\mathcal{E})$. This means that $\mathcal{E} \subseteq \mathcal{D}_D$ and therefore also $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_D$ for all D $\in \mathcal{D}_D$. Translating this back this means that

$$E \cap D \in \mathcal{D}(\mathcal{E}) \qquad \text{for all E,D} \in \mathcal{D}(\mathcal{E})$$

which is exactly what is required in Theorem 1.14 ■

Exercise 1.16 *Show that $\mathcal{D}_D$ as defined in (1.13) is a Dynkin-system.*

Exercise 1.17 *Let Ω be a set and $A, B \subseteq \Omega$. Determine $\mathcal{D}(\{A, B\})$. Show that*

$$\sigma(\{A, B\}) = \mathcal{D}(\{A, B\})$$

if and only if one of the sets $A \cap B$, $A^c \cap B$, $A \cap B^c$, $A^c \cap B^c$ is empty.

2 Volume, Pre-measure, measure

In this section we will meet again the ideas that were already sketched in the introduction: Often, when we want to construct the measure of certain sets, we already have an idea how it should act on certain elementary sets. For example, in $\mathbb{R}^d$, we have the intuitive (and correct) feeling that a measure that assigns to a rectangle $[a, b[= [a_1, b_1[\times \ldots [a_d, b_d[$ its geometric volume, i.e. $\prod_{i=1}^{d} (b_i - a_i)$ may be interesting to study. The question, whether we can also measure sets other than rectangles, then arises naturally. Can we e.g. measure the size of a circle? Already since Archimedes we know that one possibility is to approximate the circle by a sequence of (smaller and smaller) rectangles. Of course, this heavily relies on the fact that the class of rectangles is rich enough. In principle, there is nothing special about the case $\Omega = \mathbb{R}^d$ and the volume being defined on the rectangles, even though we will treat this case in some detail in the following section. In this section we will develop the concepts of volume, measure and pre-measure and discuss its properties. Then we will see that a volume may be extended (basically by applying the idea of tighter and tighter coverings) to a σ-algebra of sets.

Definition 2.1 *Let $\mathcal{R}$ be a ring. A set function*

$$\mu : \mathcal{R} \to [0, \infty] \tag{2.1}$$

is called a volume, if it satisfies

$$\mu(\emptyset) = 0 \tag{2.2}$$

and

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) \tag{2.3}$$

for all pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{R}$ and all $n \in \mathbb{N}$. A volume is called a pre - measure if

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{2.4}$$

for all pairwise disjoint sequence $(A_i)_{i\in\mathbb{N}} \in \mathcal{R}$.
We will call (2.3) finite additivity and (2.4) σ-additivity.

Example 2.2 *Let $\mathcal{R}$ be a ring over the set Ω and for $\omega \in \Omega$ define*

$$\delta_\omega(A) = \begin{cases} 1 & \omega \in A \\ 0 & \textit{otherwise} \end{cases}$$

for $A \in \mathcal{R}$. Then $\delta_\omega(\cdot)$ is a pre-measure.

Exercise 2.3 *Let Ω be a countably infinite set and $\mathcal{A}$ be the algebra*

$$\mathcal{A} := \{A \subseteq \Omega : A \textit{ finite or } A^c \textit{ finite}\}.$$

Define

$$\mu(A) = \begin{cases} 0 & \textit{if } A \textit{ is finite} \\ 1 & \textit{if } A^c \textit{ is finite} \end{cases}$$

for $A \in \mathcal{A}$. Show that μ is a volume, but not a pre - measure.

We will now discuss further properties of a volume function.

Lemma 2.4 *$\mathcal{R}$ be a ring and $A, B, A_1, A_2, \ldots \in \mathcal{R}$. Let μ be a volume on $\mathcal{R}$. Then:*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B) \tag{2.5}$$

$$A \subseteq B \Longrightarrow \mu(A) \leq \mu(B) \tag{2.6}$$

$$A \subseteq B, \mu(A) < \infty \Longrightarrow \mu(B \backslash A) = \mu(B) - \mu(A) \tag{2.7}$$

$$\mu(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mu(A_i) \tag{2.8}$$

and if the $(A_i)_{i \in \mathbb{N}}$ are pairwise disjoint and $\bigcup_{n=1}^{\infty} A_i \in \mathcal{R}$

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{i=1}^{\infty} A_n\right). \tag{2.9}$$

Proof. : Note that

$$A \cup B = A \cup (B \backslash A)$$

and

$$B = (A \cap B) \cup (B \backslash A)$$

and that these unions are disjoint implying that

$$\mu(A \cup B) = \mu(A) + \mu(B \backslash A) \tag{2.10}$$

and

$$\mu(B) = \mu(A \cap B) + \mu(B \backslash A). \tag{2.11}$$

By adding the right hand side of (2.10) and the left hand side of (2.11) this yields

$$\mu(A \cup B) + \mu(A \cap B) + \mu(B \backslash A) = \mu(A) + \mu(B) + \mu(B \backslash A).$$

If $\mu(B \backslash A) < \infty$ this is equivalent with (2.5). Otherwise $\mu(A \cup B) = \mu(B) = \infty$ and (2.5) is obvious.
For $A \subseteq B$ equation (2.11) becomes

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

which readily implies (2.6) and (2.7). Defining now $B_1 := A_1$, $B_k := A_k \backslash \left(\bigcup_{i=1}^{k-1} A_i \right)$, we see that the $B_1, \ldots, B_n$ are pairwise disjoint and $B_k \subseteq A_k$. Thus

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

Eventually for (2.9) remark that for $A = \bigcup_{i=1}^{\infty} A_i$ we have that

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

and (2.9) follows by taking the limit $n \to \infty$. ■

Note that, if μ is a pre-measure one obtains for $A_1, A_2, .. \in \mathcal{R}$ with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$ by setting

$$B_1 := A_1 \ldots \quad B_k := A_k \backslash \left(\bigcup_{i=1}^{k-1} A_i\right), \ldots$$

that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_k). \tag{2.12}$$

The following theorem relates σ - additivity to certain continuity properties of pre - measures. To facilitate notation write $E_n \uparrow E$ if $E_1 \subset E_2 \subset \ldots$and $E = \bigcup_{n=1}^{\infty} E_n$ and write $E_n \downarrow E$ if $E_1 \supset E_2 \supset \ldots$ and $E = \bigcap_{n=1}^{\infty} E_n$.

Theorem 2.5 *Let $\mathcal{R}$ be a ring and μ be a volume on $\mathcal{R}$. Consider*

(a) μ is a pre - measure.

(b) For $(A_n)_n, A_n \in \mathcal{R}, A_n \uparrow A \in \mathcal{R}$ it holds

$$\lim_{n\to\infty} \mu(A_n) = \mu(A)$$

(c) For $(A_n)_n, A_n \in \mathcal{R}, A_n \downarrow A \in \mathcal{R}$ and $\mu(A_n) < \infty$ it holds

$$\lim_{n\to\infty} \mu_n(A_n) = \mu(A)$$

(d) *For all* $(A_n)_n, A_n \in \mathcal{R}$ *with* $\mu(A_n) < \infty$ *and* $A_n \downarrow \emptyset$ *it holds*

$$\lim_{n\to\infty} \mu(A_n) = 0.$$

Then

$$(a) \Leftrightarrow (b) \Rightarrow (c) \Leftrightarrow (d)$$

If μ *is finite* $(a) - (d)$ *are even equivalent.*

Proof. $a \Longrightarrow b$: Define $A_0 := \emptyset$ and $B_n := A_n \backslash A_{n-1}$. Then the B_n are pairwise disjoint and $\bigcup_{n=1}^{\infty} B_n = A$. Thus

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \mu(A_n).$$

$b \Longrightarrow a$: Let (A_n) be pairwise disjoint in $\mathcal{R}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$. By putting $B_n = \bigcup_{i=1}^{n} A_i$, we obtain $B_n \uparrow A$ and thus

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

Thus μ is σ - additive.

$b \Longrightarrow c$: Consider $B_n := A_1 \backslash A_n$. Then $A_n \downarrow A$ implies $B_n \uparrow B := A_1 \backslash A$. Thus from (b) we get

$$\mu(B) = \mu(A_1 \backslash A) = \lim_{n\to\infty} \mu(A_1 \backslash A_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$$

If $\mu(A_n) < \infty$ we know that also $\mu(A) < \infty$ (since $A_n \supseteq A$) and therefore

$$\begin{aligned} \mu(A_1 \backslash A) &= \mu(A_1) - \mu(A) \qquad && \text{and} \\ \mu(A_1 \backslash A_n) &= \mu(A_1) - \mu(A_n) && \text{for all } n \in \mathbb{N}. \end{aligned}$$

This implies c.

$c \Longrightarrow d$: is obvious.

$d \Longrightarrow c$: If $A_n \downarrow A$, then $A_n \backslash A \downarrow \emptyset$. Since $\mu(A) \leq \mu(A_n) < \infty$ we obtain

$$\mu(A_n) - \mu(A) = \mu(A_n \backslash A) \to 0$$

which implies (c).

Eventually, if μ, ∞ on $\mathcal{R}$, also

$c \Longrightarrow b$: If $A_n \uparrow A$, then $A \backslash A_n \downarrow \emptyset$. This together with the finiteness of μ implies

$$0 = \lim_{n\to\infty} \mu\left(A \backslash A_n\right) = \lim_{n\to\infty} [\mu\left(A\right) - \mu\left(A_n\right)],$$

which in turn implies (b). ∎

Now we are ready to define the central object of this course:

Definition 2.6 *A pre-measure μ on a σ-algebra $\mathcal{A}$ is called a measure. If $\mu(\Omega) < \infty$ the measure μ is called finite; if there is a sequence of $\Omega_n \in \mathcal{A}$, $\Omega_n \uparrow \Omega$,$\mu\left(\Omega_n\right) < \infty$, μ is called σ-finite.*

Example 2.7 1. *If $\mathcal{R}$ in Example 2.2 is a σ-algebra the δ_ω defined there is a measure. δ_ω is called the Dirac measure concentrated in ω.*

2. *Let Ω be an arbitrary set and $\mathcal{A}$ be a σ - algebra on Ω. Then*

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

for $A \in \mathcal{R}$ defines a measure on $\mathcal{R}$. μ is called the counting measure.

Exercise 2.8 *Let μ be a volume over a ring $\mathcal{R}$. Show that for $A_1, \ldots, A_n \in \mathcal{R}$*

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \sum_{1\leq i_1<\ldots<i_k\leq n} (-1)^{k+1}\mu\left(A_{i_n} \cap .. \cap A_{i_k}\right). \tag{2.13}$$

We will now discuss the key problem in this section: Under which condition can a volume μ on a ring $\mathcal{R}$ be extended to a larger σ-algebra, i.e. under which condition does there exist a σ-algebra $\mathcal{A} \supseteq \mathcal{R}$ and a measure μ on $\mathcal{A}$, such that $\tilde{\mu} \mid \mathcal{R} = \mu$. Apparently we already have met a necessary condition : μ needs to be a pre-measure (because a measure has the corresponding σ-additivity property). We will now see that condition is also sufficient (which justifies the name pre-measure).

Theorem 2.9 *(Carathéodory) For every pre-measure μ on a ring $\mathcal{R}$ over Ω there is at least one way to extend μ to a measure on $\sigma\left(\mathcal{R}\right)$.*

Proof. The proof in the first step follows the geometric idea of covering a given set as neatly as possible. So for $Q \subseteq \Omega$ denote by $\mathcal{C}(Q)$ the set of all sequences $(A_n)_n; A_n \in \mathcal{R}$ with $Q \subseteq \bigcup_{n=1}^{\infty} A_n$. Define μ on $\mathcal{P}(\Omega)$ by

$$\mu^*(Q) := \begin{cases} \inf\left\{\sum_{n=1}^{\infty} \mu(A_n), (A_n)_n \in \mathcal{C}(Q)\right\} & \text{,if } \mathcal{C}(Q) \neq \emptyset \\ \infty & \text{otherwise} \end{cases} \tag{2.14}$$

This function has the following properties

$$\mu^*(\emptyset) = 0 \tag{2.15}$$

$$\mu^*(Q_1) \leq \mu^*(Q_2) \quad \text{ if } Q_1 \subseteq Q_2 \tag{2.16}$$

$$\mu^*\left(\bigcup_{n=1}^{\infty} Q_n\right) \leq \sum_{n=1}^{\infty} \mu^*(Q_n) \tag{2.17}$$

for all sequence $(Q_n)_n$,$Q_n \in \mathcal{P}(\Omega)$. This has to be shown in Exercise 2.10 below. Now note that moreover for all $A \in \mathcal{R}$ and $Q \in \mathcal{P}(\Omega)$.

$$\mu^*(Q) \geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \tag{2.18}$$

and

$$\mu^*(A) = \mu(A) \tag{2.19}$$

For the proof of (2.18) it may, of course, be assumed $\mu^*(Q) < \infty$, thus $\mathcal{C}(Q) \neq \emptyset$. Hence by finite additivity

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \cap A^c)$$

for all $(A_n)_n \in \mathcal{C}(Q)$. Moreover $(A_n \cap A)_n \in \mathcal{C}(Q \cap A)$ and $(A_n \backslash A)_n \in \mathcal{C}(Q \backslash A)$. Thus $\sum_{n=1}^{\infty} \mu(A_n) \geq \mu^*(Q \cap A) + \mu^*(Q \backslash A)$. This implies (2.18). (2.19) follows since $(A, \emptyset, \emptyset, \ldots) \in \mathcal{C}(A)$, because $\mu(A) \leq \mu^*(A)$.The importance of the observations discussed above lies in the fact that we will show the system $\mathcal{A}^*$ of all sets fulfilling (2.18) is a σ-algebra and that $\mu^* \mid \mathcal{A}^*$ is a measure. (2.18) shows that $\mathcal{R} \subset \mathcal{A}^*$, thus $\sigma(\mathcal{R}) \subseteq \mathcal{A}^*$. (2.19) eventually shows that $\mu^* \mid \mathcal{R} = \mu$, hence μ^* is a continuation of μ, which is what we have been looking for. The proof will thus be concluded by Definition 2.11 and Theorem 2.12 below. ∎

Exercise 2.10 *Prove 2.15, 2.16, 2.17. Hint for 2.17, for each* $\varepsilon > 0, n \in \mathbb{N}$, *we can take* $(A_{m,n})_m \in \mathcal{C}(Q_n)$, *such that*

$$\left|\sum_{m=1}^{\infty} \mu(A_{m,n}) - \mu^*(Q_n)\right| < \varepsilon 2^{-n}$$

Then

$$(A_{m,n})_{n,m\in\mathbb{N}} \in \mathcal{C}\left(\bigcup_{m=1}^{\infty} Q_m\right).$$

Definition 2.11 *A function* μ^* *on* $\mathcal{P}(\Omega)$ *with (2.15) - (2.17) is called an outer measure on* Ω. $A \subseteq \Omega$ *is called* μ^**- measurable, if (2.18) is satisfied for all* $Q \subseteq \Omega$.

Theorem 2.12 *Let* μ^* *be an outer measure on* Ω. *The system* $\mathcal{A}^*$ *of* μ^* *- measurable sets is a* σ *- algebra.* $\mu^* \mid \mathcal{A}^*$ *is a measure.*

Proof. Note that (2.18) is equivalent with

$$\mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \backslash A) \text{ for all } Q \in \mathcal{P}(\Omega)). \tag{2.20}$$

Indeed, applying (2.17) to the sequence

$$Q \cap A, Q \backslash A, \emptyset, \emptyset, ... \tag{2.21}$$

we immediately obtain

$$\mu^*(Q) \leq \mu^*(Q \cap A) + \mu^*(Q \backslash A) \quad \text{for all } Q \in \mathcal{P}(\Omega).$$

(2.20) implies that $\Omega \in \mathcal{A}^*$ and that with $A \in \mathcal{A}^*$ also $A^c \in \mathcal{A}^*$ holds true. Next we see that $\mathcal{A}^*$ is an algebra. So let $A, B \in \mathcal{A}^*$.The defining property (2.20) applied to B (and $Q = Q \cap A$ and $Q = Q \cap A^c$, respectively) yields

$$\mu^*(Q \cap A) = \mu^*(Q \cap A \cap B) + \mu^*(Q \cap A \cap B^c)$$

$$\mu^*(Q \cap A^c) = \mu^*(Q \cap A^c \cap B) + \mu^*(Q \cap A^c \cap B^c)$$

Since also $A \in \mathcal{A}^*$ we know that

$$\begin{aligned} \mu^*(Q) &= \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \\ &= \mu^*(Q \cap A \cap B) + \mu^*(Q \cap A \cap B^c) \\ &+ \mu^*(Q \cap A^c \cap B) + \mu^*(Q \cap A^c \cap B^c). \end{aligned} \tag{2.22}$$

Since this is true for all $Q \in \mathcal{P}(\Omega)$ we may also replace Q by $Q \cap (A \cup B)$ to obtain

$$\mu^*(Q \cap (A \cup B)) = \mu^*(Q \cap A \cap B) + \mu^*(Q \cap A \cap B^c) + \mu^*(Q \cap A^c \cap B) \tag{2.23}$$

for all $Q \in \mathcal{P}(\Omega)$. (2.22) together with (2.23) gives

$$\mu^*(Q) = \mu^*(Q \cap (A \cup B)) + \mu^*(Q \backslash (A \cup B))$$

for all $Q \in \mathcal{P}(\Omega)$. This shows that $A \cup B \in \mathcal{A}^*$. In the next two steps we will see that the algebra $\mathcal{A}^*$ is a $\cap$ - stable Dynkin - system, thus a σ - algebra. So let $(A_n)_n$ be a sequence of pairwise disjoint sets in $\mathcal{A}^*$ and set $A := \bigcup_{n=1}^{\infty} A_n$. (2.23) yields by induction:

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

for all $n \in \mathbb{N}, Q \in \mathcal{P}(\Omega)$.Taking into account that from the above we know that $B_n := \bigcup_{i=1}^{n} A_i \in \mathcal{A}^*$ and that $Q \backslash B_n \supseteq Q \backslash A$ and therefore

$$\mu^*(Q \backslash B_n) \geq \mu^*(Q \backslash A)$$

we obtain

$$\mu^*(Q) = \mu^*(Q \cap B_n) + \mu^*(Q \backslash B_n) \geq \sum_{i=1}^{n} \mu^*(Q \cap A_i) + \mu^*(Q \backslash A).$$

Using (2.17) this gives

$$\mu^*(Q) \geq \sum_{i=1}^{\infty} \mu^*(Q \cap A_i) + \mu^*(Q \backslash A) \geq \mu^*(Q \cap A) + \mu^*(Q \backslash A).$$

This, according to what we said at the beginning of this proof, even yields:

$$\mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \backslash A) = \sum_{i=1}^{\infty} \mu^*(Q \cap A_i) + \mu^*(Q \backslash A). \tag{2.24}$$

This means that $A \in \mathcal{A}^*$. Therefore we have shown that $\mathcal{A}^*$is a Dynkin - system. Moreover $\mathcal{A}^*$ is an algebra. But a Dynkin - system, that is an

algebra , is $\cap$ - stable (because $A \cap B = (A^c \cup B^c)^c$. Thus we see that $\mathcal{A}^*$ is a $\cap$ - stable Dynkin - system, hence a σ - algebra.
Choosing $A = Q$ in (2.24) gives

$$\mu^*(A) = \sum_{i=1}^{\infty} \mu^*(A_i)$$

which means that μ^* restricted to $\mathcal{A}^*$ is a measure. ■

Of course, it would be nice to know, that μ continued to $\mathcal{A}^*$ not only exists, but also is unique. This in many important cases indeed is true. We bring a frequently applied technique using Dynkin-system into action.

Theorem 2.13 *Let $\mathcal{E}$ be a $\cap$ - stable generator of a σ - algebra $\mathcal{A}$ over Ω. Assume there is a sequence $(E_n)_n, E_n \in \mathcal{E}$ with $\bigcup_{i=1}^{\infty} E_i = \Omega$. Assume that μ_1, μ_2 are two measure on $\mathcal{A}$ with*

$$\mu_1(E) = \mu_2(E) \qquad \text{for all } E \in \mathcal{E} \tag{2.25}$$

and

$$\mu_1(E_n) < \infty \qquad \text{for all } n \in \mathbb{N}. \tag{2.26}$$

Then $\mu_1 = \mu_2$.

Proof. Let $\mathcal{E}_E$ be the system of all $E \in \mathcal{E}$ with $\mu_1(E) = \mu_2(E) < \infty$. For an arbitrary $E \in \mathcal{E}_E$ consider

$$\mathcal{D}_E := \{D \in \mathcal{A} : \mu_1(E \cap D) = \mu_2(E \cap D)\}.$$

In Exercise 2.14 below it has to be show that $\mathcal{D}_E$ is a Dynkin - system. Since $\mathcal{E}$ is $\cap$ - stable we have $\mathcal{E} \subset \mathcal{D}_E$, because of (2.25) and the definition of $\mathcal{D}_E$. Thus $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_E$. On the other hand the $\cap$ - stability of $\mathcal{E}$ yields $\mathcal{A} = \mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E})$ and hence (since $\mathcal{D}_E \subset \mathcal{A}$), that $\mathcal{D}_E = \mathcal{A}$. Thus

$$\mu_1(E \cap A) = \mu_2(E \cap A) \tag{2.27}$$

for all $E \in \mathcal{E}_E$ and $A \in \mathcal{A}$. Because of (2.26) this in particular means that

$$\mu_1(E_n \cap A) = \mu_2(E_n \cap A)$$

for all $A \in \mathcal{A}, n \in \mathbb{N}$. The rest of the proof consists of slicing A into pieces. Put

$$F_1 := E_1 \quad \text{ and } \quad F_n := E_n \backslash \left(\bigcup_{i=1}^{n-1} E_i \right) \quad n \in \mathbb{N}.$$

Then the (F_n) are pairwise disjoint with $F_n \subset E_n$ and
$\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} E_n = \Omega$. Since $F_n \cap A \in \mathcal{A}$ we obtain from(2.27):

$$\mu_1 (F_n \cap A) = \mu_1 (E_n \cap F_n \cap A) = \mu_2 (E_n \cap F_n \cap A) = \mu_2 (F_n \cap A).$$

for all $A \in \mathcal{A}$ and $n \in \mathbb{N}$. Since

$$A = \bigcup_{n=1}^{\infty} (F_n \cap A)$$

the σ -additivity of $\mu_1$and μ_2 gives

$$\mu_1 (A) = \sum_{n=1}^{\infty} \mu_1 (F_n \cap A) = \sum_{n=1}^{\infty} \mu_2 (F_n \cap A) = \mu_2 (A) \text{ for all } A \in \mathcal{A}$$

which is $\mu_1 = \mu_2$. ■

Exercise 2.14 *Show that $\mathcal{D}_E$ as defined in the proof of Theorem 2.13 is a Dynkin - system.*

Theorem 2.9, 2.12, and 2.13 can be summarized in the following

Theorem 2.15 *Every σ-finite pre-measure on a ring $\mathcal{R}$ over a set Ω can be uniquely extended to a measure $\tilde{\mu}$ on $\sigma(\mathcal{R})$.*

Proof. Only uniqueness still needs to be proven. But this is immediate from Theorem 2.13: Since μ is σ-finite, the ring $\mathcal{R}$ possesses all properties of the generator in Theorem 2.13. ■
Already the construction given in the proof of Theorem 2.9 suggests that for $A \in \mathcal{A}^*$ its measure $\tilde{\mu}(A)$ can be approximated by measures on the ring. This is formalized in

Theorem 2.16 *Let μ be a finite measure on a σ - algebra $\mathcal{A}$ over Ω, which is generated by an algebra $\mathcal{A}_0$ over Ω. Then for $A \in \mathcal{A}$ there exists a sequence $(C_n)_{n \in \mathbb{N}}, C_n \in \mathcal{A}_0$ with*

$$\mu (A \Delta C_n) \to 0 \tag{2.28}$$

as $n \to \infty$. Here for any two sets $A, B \subseteq \Omega$

$$A \Delta B := A \backslash B \cup B \backslash A.$$

Proof. Let $\varepsilon > 0, A \in \mathcal{A}$. According to (2.14) there is a sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_0$ with $\bigcup_{n=1}^{\infty} A_n \supseteq A$ and

$$0 \leq \sum_{n=1}^{\infty} \mu(A_n) - \mu(A) < \frac{\varepsilon}{2} \tag{2.29}$$

Set $C_n := \bigcup_{i=1}^{n} A_i$ and $A' := \bigcup_{n=1}^{\infty} A_n$. Then $C_n \uparrow A'$ and $A' \backslash C_n \downarrow \emptyset$. μ is finite and thus $\emptyset$ - continuous, therefore

$$\mu(A' \backslash C_{n_0}) < \frac{\varepsilon}{2}$$

for some n_0. Now

$$A \triangle C_{n_0} = (A \backslash C_{n_0}) \cup (C_{n_0} \backslash A) \subset (A' \backslash C_{n_0}) \cup (A' \backslash A)$$

and hence

$$\begin{aligned} & \mu(A \triangle C_{n_0}) \leq \mu(A' \backslash C_{n_0}) + \mu(A' \backslash A) \\ \leq \quad & \mu(A' \backslash C_{n_0}) + \sum_{n=1}^{\infty} \mu(A_n) - \mu(A) < \varepsilon \end{aligned}$$

because of (2.29) and (2). This proves the theorem. ∎

Exercise 2.17 *Let* $\mu = \delta_\omega$ *be the Dirac - measure on a ring* $\mathcal{R}$ *over* Ω. *Assume* $\{\omega\} = \bigcap_{n=1}^{\infty} A_n$ *and* $\Omega = \bigcup_{n=1}^{\infty} B_n$, *for two sequences* $(A_n)_n, (B_n)_n$ *in* $\mathcal{R}$. *Prove that:*

a) The outer measure μ^* *generated by* μ *assigns* 1 *or* 0 *to* $A \in \mathcal{P}(\Omega)$, *depending on whether* $\omega \in A$ *or not.*

b) $\mathcal{A}^* = \mathcal{P}(\Omega)$.

c) $\mu^* = \delta_\omega$ *on* $\mathcal{P}(\Omega)$.

Exercise 2.18 *A measure* μ *over a* σ *- algebra* $\mathcal{A}$ *is called complete, if* $N \in \mathcal{A}, \mu(N) = 0, N' \subset N$ *implies* $N' \in \mathcal{A}$. *Show that:*

a) $\mu^*|A^*$ *as defined in Theorem 2.12 is complete.*

b) Let $\mathcal{A}$ *be a* σ *- algebra over* Ω *and* $\{\omega\} \in \mathcal{A}$.

δ_ω (the Dirac measure) is complete, if and only if $\mathcal{A} = \mathcal{P}(\Omega)$.

3 Lebésgue-measure

From a technical point of view this section starts by applying the concepts developed in Section 2 to a particular, yet important case, the case of $\mathbb{R}^d$. As already mentioned here we have an intuitive idea what the measure for fairly easy geometric objects, say e.g. rectangles should be. We want to extend this measure to more subtle sets.

Definition 3.1 *Let $a, b \in \mathbb{R}^d$.By the rectangle $[a, b[$ we mean the set*

$$[a, b[:= \left\{x \in \mathbb{R}^d : a_i \leq x_i < b_i, i = 1...d\right\}$$

Similarly, we define $]a, b[,]a, b]$, and $[a, b]$
Let moreover

$$\mathcal{J}^d := \left\{[a, b[: \ a, b \in \mathbb{R}^d\right\}$$

and

$$\mathcal{F}^d := \left\{\bigcup_{i=1}^{n} \mathcal{J}_i, n \in \mathbb{N}, \mathcal{J}_i \in \mathcal{J}^d\right\}.$$

Exercise 3.2 *For $I, J \in \mathcal{J}^d$ it holds $I \cap J \in \mathcal{J}^d$ and $I \backslash J \in \mathcal{F}^d$*

Exercise 3.3 *Let $F \in \mathcal{F}^d$. Then there exists $I_1, ..., I_n \in \mathcal{J}^d, I_i \cap I_j = \emptyset$ for $i \neq j$, such that*

$$F = \bigcup_{i=1}^{n} I_i.$$

Exercise 3.4 *$\mathcal{F}^d$ is a ring over $\mathbb{R}^d$.*

These preparations, of course, were necessary to apply the techniques obtained in Section 2. Now we will turn to discussing the corresponding volume on $\mathcal{F}^d$, which will turn out to be the geometric volume.

Definition 3.5 *Let $I \in \mathcal{J}^d$, $I = [a, b[$. We define*

$$\overline{\lambda}(I) = \begin{cases} \prod_{i=1}^{d}(b_i - a_i) & \text{if } I \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Theorem 3.6 *There exists a unique volume λ on $\mathcal{F}^d$ such that λ extends $\overline{\lambda}$ on $\mathcal{J}^d$. λ is a pre-measure.*

Proof. Following Exercise 3.2 the set $F \in \mathcal{F}^d$ may be written as $F = \bigcup_{i=1}^n I_i$ with pairwise disjoint $I_i \in \mathcal{F}^d$. Since a volume has to be additive, there is only one way to define $\lambda(F)$, namely

$$\lambda(F) = \sum_{i=1}^n \overline{\lambda}(I_i).$$

Of course, we need to check that this construction is well defined. To this end we write

$$F = \bigcup_{i=1}^n I_i = \bigcup_{j=1}^m J_j$$

where the $I_i, J_j \in \mathcal{J}^d$ and the (I_i) are pairwise disjoint as well as the (J_j).We then need to see that

$$\sum_{i=1}^n \overline{\lambda}(I_i) = \sum_{j=1}^m \overline{\lambda}(J_i).$$

First note, if $[a, b[\in \mathcal{J}^d, \tilde{a}_1$ is such that

$$a_1 < \tilde{a}_1 < b_1 \quad \text{and } \tilde{a}_i = a_i \quad \text{ for } i \geq 2, \text{ then}$$

$$[a, b[= [a, \tilde{a}[\,\dot{\cup}\, [\tilde{a}, b[,$$

as well as

$$\begin{aligned}
\overline{\lambda}([a, b[) &= \prod_{i=1}^d (b_i - a_i) \\
&= [(b_1 - \tilde{a}_1) + (\tilde{a}_1 - a_1)] \prod_{i=2}^d (b_i - a_i) \\
&= \prod_{i=1}^d (b_1 - \tilde{a}_1) + \prod_{i=1}^d (\tilde{a}_1 - a_1) \\
&= \overline{\lambda}([\tilde{a}, b[) + \overline{\lambda}([a, \tilde{a}[).
\end{aligned}$$

Induction over: gives that for $\tilde{a}_i \leq c_i \leq b_i$

$$\overline{\lambda}([a, b[) = \overline{\lambda}([a, b[) + \overline{\lambda}([c, b[).$$

Another induction gives that, if $\mathcal{J}^d \ni I = \bigcup_{i=1}^n I_i$ with $I_i \in \mathcal{J}^d$, that

$$\overline{\lambda}(I) = \sum_{i=1}^n \overline{\lambda}(\mathcal{I}).$$

So λ defined above is well defined on $\mathcal{J}^d$. Eventually let $F \in \mathcal{F}^d$ be of the form

$$F = \bigcup_{i=1}^n I_i = \bigcup_{j=1}^m J_m$$

with $I_i, J_j \in \mathcal{J}^d$ and (I_i) pairwise disjoint as well as the (J_j). Then $(I_i \cap J_j)_{\substack{i\le n\\ j\le m}}$ is a common refinement of both the $(I_i)_i$ and the (J_j) and, of course the sets $I_i \cap J_j$ are pairwise disjoint. Then applying the above

$$\begin{aligned}\sum_{i=1}^n \overline{\lambda}(I_i) &= \sum_{i=1}^n \sum_{j=1}^m \overline{\lambda}(I_i \cap J_j) \\ &= \sum_{j=1}^m \sum_{i=1}^n \overline{\lambda}(I_i \cap J_j) = \sum_{j=1}^m \overline{\lambda}(J_j).\end{aligned}$$

Hence defining

$$\lambda(F) := \sum_{i=1}^n \overline{\lambda}(I_i)$$

we obtain a well defined and finite volume on $\mathcal{F}^d$. To see that λ indeed also is a pre - measure, we only need to check that λ is $\emptyset$ - finite (this is an application of Theorem 2.5, since λ is finite on each $[a, b[$, if $a \neq -\infty, b \neq \infty$). So let $(F_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $\mathcal{F}^d$. We will show that

$$\delta := \lim_{n\to\infty} \lambda(F_n) = \inf_{n\to\infty} \lambda(F_n) > 0$$

implies that $\bigcap_{n=1}^\infty F_n \neq \emptyset$. We will use a definition of compactness that states that an intersection of a sequence of decreasing closed sets is empty if and only if one of the sets is empty. To be more precise:
Since each F_n is a finite union of disjoint elements in $\mathcal{J}^d$ we may find $G_n \in \mathcal{F}^d$ with

$$\begin{gathered} G_n \subset \overline{G}_n \subset F_n \quad \text{and} \\ |\lambda(G_n) - \lambda(F_n)| \le 2^{-n}\delta. \end{gathered}$$

Put $H_n := \bigcap_{i=1}^{n} G_i$, then $H_n \in \mathcal{F}^d$ and $H_n \supseteq H_{n+1}$ as well as $\overline{H}_n \subseteq \overline{G}_n \subseteq F_n$. F_n is bounded. Thus $\left(\overline{H}_n\right)_n$ is a sequence of bounded and hence compact subsets of $\mathbb{R}^d$ with $F_n \supseteq \overline{H}_{n+1}$. Thus $\bigcap_{n=0}^{\infty} \overline{H}_n \neq \emptyset$ (and therefore also $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$), if only $H_n \neq \emptyset$ for each n. To this end we show

$$\lambda\left(H_n\right) \geq \lambda\left(F_n\right) - \delta\left(1 - 2^{-n}\right) \geq \delta 2^{-n} \tag{3.1}$$

where only the first inequality has to be proven. This will be done by induction over n. For $n = 1$ (3.1) is true since $H_1 = G_1$ and $\lambda\left(F_1\right) - \lambda\left(G_1\right) \leq \frac{\delta}{2}$. Assuming that the hypothesis is true for n we know that

$$\lambda\left(H_n\right) \geq \lambda\left(F_n\right) - \delta\left(1 - 2^{-n}\right)$$

as well as

$$\lambda\left(G_{n+1}\right) \geq \lambda\left(F_{n+1}\right) - \delta 2^{-(n+1)}$$

and $G_{n+1} \cup H_n \subseteq F_{n+1} \cup F_n = F_n$. Putting this together yields:

$$\begin{aligned} \lambda\left(H_{n+1}\right) &\geq \lambda\left(F_{n+1}\right) - \delta 2^{-(n+1)} - \delta\left(1 - 2^{-n}\right) \\ &\geq \lambda\left(F_{n+1}\right) - \delta\left(1 - 2^{-(n+1)}\right). \end{aligned}$$

This proves $H_n \neq \emptyset$ for all n and thus the theorem. ■

Thus we know that λ is a σ -finite pre - measure on the ring $\mathcal{F}^d$. Applying Theorem 2.15 we immediately obtain

Corollary 3.7 *The pre - measure λ on $\mathcal{F}^d$ can be uniquely extended to a measure λ on $\sigma\left(\mathcal{F}^d\right)$.*

Definition 3.8 *The measure λ in Corollary 3.7 is called the Lebesgue - measure. $\sigma\left(\mathcal{F}^d\right)$ is called the Borel σ - algebra and abbreviated by $\mathcal{B}^d$. Sometimes we will also write λ^d instead of λ to emphasize its dimension dependence.*

Note that, of course, also λ^d is σ - finite. We will now first discuss the form of the σ - algebra $\mathcal{B}^d$ a bit more in detail. From a topological point of view the following result is very satisfactory:

Theorem 3.9 *Denote by $\mathcal{O}^d, \mathcal{C}^d$, and $\mathcal{K}^d$ the systems of all open, closed, and compact subsets of $\mathbb{R}^d$, respectively. Then*

$$\mathcal{B}^d = \sigma\left(\mathcal{O}^d\right) = \sigma\left(\mathcal{C}^d\right) = \sigma\left(\mathcal{K}^d\right) \tag{3.2}$$

Proof. Note that $\mathcal{K}^d \subseteq \mathcal{C}^d$ and therefore $\sigma\left(\mathcal{K}^d\right) \subseteq \sigma\left(\mathcal{C}^d\right)$. On the other hand every set $C \in \mathcal{C}^d$ is the countable union of a sequence of sets $C_n \in \mathcal{K}^d$. Indeed, if

$$K_n := \left\{x \in \mathbb{R}^d : ||x|| \leq n\right\}$$

then $C = \bigcup_{n=1}^{\infty}(C \cap K_n)$. But then $\mathcal{C}^d \subseteq \sigma\left(\mathcal{K}^d\right)$ which together with the above shows that $\sigma\left(\mathcal{C}^d\right) = \sigma\left(\mathcal{K}^d\right)$. On the other hand the complement of a closed set is an open set and thus

$$\sigma\left(\mathcal{O}^d\right) = \sigma\left(\mathcal{K}^d\right) = \sigma\left(\mathcal{C}^d\right).$$

Eventually we show that $\sigma\left(\mathcal{O}^d\right) = \mathcal{B}^d$. To this end first note that $[a, b[\in \mathcal{J}^d$ may be written as a countable intersection of $]a^{(n)}, b[$, where

$$a^{(n)} = \left(a_1 - \frac{1}{n}, .., a_d - \frac{1}{n}\right).$$

Thus $\mathcal{B}^d = \sigma\left(\mathcal{J}^d\right) \subseteq \sigma\left(\mathcal{O}^d\right)$. On the other hand $]a, b[\in \mathcal{O}^d$ is the union of $]\tilde{a}^{(n)}, b[\in \mathcal{J}^d$, where

$$\tilde{a}^{(n)} = \left(a_1 + \frac{1}{n}, .., a_d + \frac{1}{n}\right).$$

On the other hand every open set $G \in \mathcal{O}^d$ can be written as a countable union of $]a, b[\in \mathcal{O}^d$ (e.g. those with rational coordinates). This shows that $\sigma\left(\mathcal{O}^d\right) \subseteq \mathcal{B}^d$, thus $\sigma\left(\mathcal{O}^d\right) = \mathcal{B}^d$ and hence proves the theorem. ■

In some exercises we will now discuss the Lebesgue measure of some fairly simple subsets of $\mathbb{R}^d$.

Exercise 3.10 *Let H be a hypersurface in $\mathbb{R}^d$, that is perpendicular to one of the coordinate extras. Prove that $\lambda^d(H) = 0$.*

Exercise 3.11 *Prove that every countable subset of $\mathbb{R}^d$ has Lebesgue measure zero.*

The Lebesgue measure introduced above is the prototype of a Borel measure, i. e. of a measure on $\left(\mathbb{R}^d, \mathcal{B}^d\right)$. A closer look to its construction reveals that in dimension one of the starting point is the attach the measure $b - a$ to an interval $[a, b[$. This is geometrically reasonable, but in general (in particular, if we think of probability measures) not necessary. One might in general attach a measure $F(b) - F(a)$ to $[a, b[$. For that F has to be increasing (otherwise some intervals have negative measure) and left - continuous, since $x_n \uparrow x$ implies $[y, x_n[\uparrow [y, x[$ and thus $\mu([y, x_n[) \to \mu([y, x[)$ for every measure μ.

Definition 3.12 *A function F is called measure generating, if F is increasing and left continuous.*

The following theorem will not be proven in this course. Its proof is to a large extend similar to the construction of Lebesgue measure.

Theorem 3.13 *Let F be measure generating. Then there exists a unique measure μ_F on $B^1$with*

$$\mu_F\left([a,b[\right) = F\left(b\right) - F\left(a\right).$$

Moreover, if G is another measure generating function on $\mathbb{R}$ with $\mu_F = \mu_G$, the $F = G + c$ for some constant c.

In a subsequent course in probability theory a special role will be played by probability measure, i. e. measures which have total mass one. Of course, they can be obtained from the finite measures by normalization. Concentrating on $(\mathbb{R}, B^1)$ again, we see that μ_F is a probability measure on $(\mathbb{R}, B^1)$, if and only if $\lim_{x\to\infty} F\left(x\right) - \lim_{y\to-\infty} F\left(y\right) = 1$.Usually one takes $\lim_{x\to\infty} F\left(x\right) = 1$.
In order to continue the discussion of Lebesgue - measure, we need to interlude on the connection of measures and mappings. This is done in the next section.

Exercise 3.14 *(a bit of typology (!!) which we needed in this section): Let K be compact. Let$(A_n)_n$ be a sequence of closed subsets of K with $\bigcap_{i=1}^n A_i \neq \emptyset$ for all n.Then also*

$$\bigcap_{i=1}^{\infty} A_i \neq \emptyset.$$

4 Measurable mappings and image measures

Assume that we have a set Ω an σ - algebra $\mathcal{A}$ on Ω (we will call $(\Omega, \mathcal{A})$ a measurable space). Moreover assume μ is a measure on $\mathcal{A}$. In this section we will discuss how to "teleport" μ to another measurable space (Ω', A') by a mapping.

Definition 4.1 *Let $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ be measurable spaces. A mapping $T : \Omega \to \Omega'$ is called $\mathcal{A} - A'$ - measurable, if*

$$T^{-1}\left(A'\right) \in \mathcal{A} \quad \textit{for all } A' \in \mathcal{A}.' \tag{4.1}$$

Example 4.2 *Every constant mapping is measurable, since*

$$T^{-1}(A') \in \{\emptyset, \Omega) \quad \text{for all } A' \in \mathcal{A}'.$$

Exercise 4.3 *Let* $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ *be two measurable spaces. Let* $\mathcal{E}'$ *be a generator of* $\mathcal{A}'$. *Show that* $T : \Omega \to \Omega$ *is* $\mathcal{A} - \mathcal{A}'$ *- measurable if and only if*

$$T^{-1}(E') \in \mathcal{A} \quad \text{for all } E' \in \mathcal{E}'. \tag{4.2}$$

Apply this to show that every continuous function $T : \mathbb{R}^d \to \mathbb{R}^{d'}$ *is* $\mathcal{B}^d - \mathcal{B}^{d'}$ *- measurable.*

Exercise 4.4 *Let* $T_1 : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ *and* $T_2 : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ *be measurable mappings. Then* $T_2 \circ T_1$ *is* $\mathcal{A}_1 - \mathcal{A}_3$ *- measurable.*

One could in principles discuss more properties of measurable functions along the lines of set theoretic topology. We will rather refrain from that and now discuss how measures are "teleported".

Theorem 4.5 *Let* $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ *be measurable. Then for every measure* μ *on* $(\Omega, \mathcal{A})$

$$T(\mu)(A') := \mu\left(T^{-1}(A')\right), \qquad A' \in \mathcal{A}'$$

defines a measure on $(\Omega', \mathcal{A}')$.

Proof. The only thing that needs to be understood is that, if $(A'_n)_{n\in\mathbb{N}}$ is pairwise disjoint in $\mathcal{A}'$ then $(T^{-1}(A'_n))_{n\in\mathbb{N}}$ is pairwise disjoint in $\mathcal{A}$,and that

$$T^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} T^{-1}(A'_n).$$

The rest follows easily. ■

Definition 4.6 *The measure* $T(\mu)$ *in the situation of Theorem 4.5 is called the image measure of* μ *under the mapping* T.

A very important example, namely that of $\left(\mathbb{R}^d, \mathcal{B}^d\right)$ and the mappings T being translations will be discusses in Section 5 below.

5 Further properties of Lebesgue measure

In this section we will study some further properties of Lebesgue measure. A special focus will be on its behavior under certain linear mappings. Let us start with one of the easiest cases, that of a translation: A translation by $a \in \mathbb{R}^d$ is a mapping $T_a : \mathbb{R}^d \to \mathbb{R}^d$ with $T_a(x) = a + x$ for all $x \in \mathbb{R}^d$.

Proposition 5.1 *Lebesgue measure is translation - invariant, i.e.* $T_a\left(\lambda^d\right) = \lambda^d$ *for all* $a \in \mathbb{R}^d$*.*

Proof. This follows immediately since for each $I \in \mathcal{J}^d$ we have by definition

$$\lambda^d(I) = T_a\left(\lambda^d\right)(I).$$

■

As it next turns out this property is characteristic of Lebesgue measure. Let $W := [0,1[$ be the unit cube in $\mathbb{R}^d$ (0 and 1 are d - vectors). Then:

Theorem 5.2 *Let* μ *be a measure on* $\left(\mathbb{R}^d, \mathcal{B}^d\right)$ *with* $T_a(\mu) = \mu$ *for all* $a \in \mathbb{R}^d$ *and*

$$\alpha := \mu(W) < \infty. \tag{5.1}$$

Then

$$\mu = \alpha\, \lambda^d. \tag{5.2}$$

i.e. Lebesgue measure is the unique translation - invariant measure up to scaling.

Proof. The trick is that $\mu(W)$ also determines $\mu(W_n)$ for small cubes of side length $\frac{1}{n}$ and that every $I \in \mathcal{J}^d$ may be arbitrary well approximated by $W_{n's}$. More precisely let

$$W_n = [0, \frac{1}{n}[\in \mathcal{J}^d$$

Then

$$\mu(W_n) = \frac{\alpha}{n^d}. \tag{5.3}$$

Indeed W is the disjoint union of n^d copies of W_n, which by translation invariance all have the same μ - measure. This gives (5.3).
Moreover, for $[a, b[\in \mathcal{J}^d$ with $a = (a_1, ..., a_d)$, $b = (b_1, ..., b_d)$, $a_i, b_i \in \mathbb{Q}$ for all $i = 1...d$, we see in the same way

$$\mu([a, b[) = \alpha\lambda^d([a, b[). \tag{5.4}$$

Indeed, for n large enough, we may partition $[a,b[$ into little cubes of side length $\frac{1}{n}$. For each of these cubes the relation holds, thus also for their finite union. Eventually

$$\mathcal{J}^d_{rat} := \left\{[a,b[\in \mathcal{J}^d : a,b \text{ have rational coordinates}\right\} \tag{5.5}$$

is a $\cap$ - stable generator of $\mathcal{B}^d$. Hence, if μ and $\alpha\lambda^d$ agree on $\mathcal{J}^d_{rat}$ they agree on $\mathcal{B}^d$, which was our claim. ■

Exercise 5.3 *Show that $\mathcal{J}^d_{rat}$ as defined in (??) is a $\cap$ - stable generator of $\mathcal{B}^d$.*

We will now see that for Lebesgue measure to exist, finite dimensionality is essential: "Lebesgue measure in $\mathbb{R}^\infty$ does not exist".

Theorem 5.4 *In $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ there is no translation invariant measure λ, that assigns mass a positive, finite to any bounded set W, i.e. with*

$$0 < \lambda(W) < \infty.$$

Proof. Assume such a λ exists. Take a closed Ball of radius $0 < \varepsilon < \frac{\sqrt{2}}{2}$. Denote by

$$B_\varepsilon(y) := \{x \in \mathbb{R}^\infty : \|x-y\| < \varepsilon\}.$$

$B_\varepsilon(0)$ has positive measure because it contains a small cube (e.g. of side length $\frac{\varepsilon}{4}$). On the other hand the $(B_\varepsilon(e_i))_{i=1}^\infty$ are all disjoint (here e_i denotes the i'th unit vector) and

$$\left(\bigcup_{i=1}^\infty B_\varepsilon(e_i)\right) \subseteq B_3(0). \tag{5.6}$$

(5.7) implies that $\lambda\left(\bigcup_{i=1}^\infty B_\varepsilon(e_i)\right)$ is finite. On the other by translation invariance all $B_\varepsilon(e_i)$ have the same measure. But since all $B_\varepsilon(e_i)$ are disjoint $\bigcup_{i=1}^\infty B_\varepsilon(e_i)$ then has infinite measure. This is a contradiction. ■
The translation invariance of Lebesgue measure also the main obstacle that prevents all subsets of $\mathbb{R}^d$ from being measurable:

Theorem 5.5 *$\mathcal{B}^d \neq \mathcal{P}\left(\mathbb{R}^d\right)$, i.e. there is a non - measurable subset of $\mathbb{R}^d$.*

Proof. Introduce the equivalence relation $\sim$ on $\mathbb{R}$ (we will just treat the case $d = 1$, which is sufficient), via

$$x \sim y :\Longleftrightarrow x - y \in \mathbb{Q} \tag{5.7}$$

We consider the equivalence classes with respect to $\sim$. Since for any $x \in \mathbb{R}$, there is $n \in \mathbb{Q}$ with $n \leq x < n + 1$, we may choose a representative for each of the equivalence classes in $[0, 1[$. That means that there is a set $K \subset [0, 1[$ which contains exactly one element of each equivalence class. (Note that here the axiom of choice plays an important role).The set K satisfies

$$\bigcup_{y \in \mathbb{Q}} (y + k) = \mathbb{R} \tag{5.8}$$

and

$$y_1 \neq y_2 \Rightarrow (y_1 + K) \cap (y_2 + K) = \emptyset. \tag{5.9}$$

But this leads to a contradiction as follows. Assume K was measurable. Then

$$\lambda \left(\bigcup_{y \in \mathbb{Q}} (y + k) \right) = \lambda (\mathbb{R}) = \infty$$

Due to translation invariance of λ:

$$\lambda (y + K) = \lambda (K) \qquad \text{for all } y \in \mathbb{Q}$$

Since $\mathbb{Q}$ is countable $\lambda (K)$ thus cannot be zero. On the other hand

$$\bigcup_{y \in \mathbb{Q} \cap [0,1]} (y + K) \subseteq [0, 2[.$$

Thus $\lambda (K)$ cannot be non - zero, either $\mathbb{Q} \cap [0, 1]$ is infinite). This is a contradiction. Hence K is not measurable. ■
At the end of this section we will discuss the relation of Lebesgue measure with more general mappings. To this end introduce the space

$$\mathcal{W} \left(\mathbb{R}^d\right) := \left\{T : \mathbb{R}^d \to \mathbb{R}^d : d\left(T(x), T(y)\right) = d(x, y) \qquad \text{for all } x, y \in \mathbb{R}\right\}$$

Recall that $\mathcal{W} \left(\mathbb{R}^d\right)$ contains exactly those mappings that can be written as a orthogonal, linear mapping plus a shift. Under $T \in \mathcal{W} \left(\mathbb{R}^d\right)$ a geometric figure will be moved to a congruent figure. We will now see that the Lebesgue measure is invariant under $T \in \mathcal{W} \left(\mathbb{R}^d\right)$.

Theorem 5.6 *For each $T \in \mathcal{W}$ it holds $T(\lambda^d) = \lambda^d$.*

Proof. Since $T \in \mathcal{W}$ can be written as

$$T(x) = T_0(x) + a, \qquad , x \in \mathbb{R}$$

where $a \in \mathbb{R}^d$ and T_0 is orthogonal and we already know that Lebesgue measure is translation invariant, it suffices to consider the case that T us orthogonal, in particular $T(0) = 0$. For such a T, $a \in \mathbb{R}^d$, and $b = T^{-1}(a)$ it holds

$$T_a \circ T = T \circ T_b \tag{5.10}$$

where $T_c(x) = x + c$ (Exercise 5.7 below). Using that λ^d is translation invariant we thus obtain

$$T_a(T(\lambda^d)) = T(T_b(\lambda^d)) = T(\lambda^d)$$

for all $a \in \mathbb{R}^d$. But this means that $\mu := T(\lambda^d)$ is translation invariant. Hence $T(\lambda^d) = \alpha\lambda^d$, with $\alpha = \mu(W)$, and $W = [0, 1[$. Consider

$$B_1(0) := \{x \in \mathbb{R}^d : ||x|| \leq 1\}.$$

With T also T^{-1} is orthogonal and hence $T^{-1}[B_1(0)] := B_1(0)$. Therefore $T(\lambda^d) = \alpha\lambda^d$ gives

$$\lambda^d(B_1(0)) = \lambda^d(T^{-1}[B_1(0)]) = T(\lambda^d)(B_1(0)) = \alpha\lambda^d(B_1(0)).$$

Since $\lambda^d(B_1(0)) \notin \{0, \infty\}$ this implies $\alpha = 1$. ■

Exercise 5.7 *Show formula (5.10) above.*

Exercise 5.8 *Show that any hypersurface $H \subset \mathbb{R}^d$ has Lebesgue measure λ^d zero.*

Eventually we will turn to general linear mappings T with $\det T = 0$ (the reason for this restriction will become clear in (5.11) below.

Theorem 5.9 *For every $T \in GL(d, \mathbb{R}^d)$ we have*

$$T(\lambda^d) = \frac{1}{|detT|}\lambda^d. \tag{5.11}$$

Proof. Let $T \in GL(d, \mathbb{R}^d)$ be an invertible, linear mapping. The same computation as in the proof of Theorem 5.7 shows that $T(\lambda^d)$ is translation invariant. Since

$$\Phi(T) := T(\lambda^d)(W) < \infty \qquad (W := [0,1[)$$

we obtain

$$T(\lambda^d) = \Phi(T)(\lambda^d). \tag{5.12}$$

To determine $\Phi(T)$ we note that for $S \in GL(d, \mathbb{R}^d)$

$$\Phi(S \cdot T) = \Phi(S)\Phi(T) \tag{5.13}$$

(see Exercise 5.10 below). This means Φ is a homorphism from $GL(d, \mathbb{R})$ into the multiplicative group $\mathbb{R} \setminus \{0\}$. Hence we know from linear algebra that there is an homomorphism

$$\phi : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$$

such that

$$\Phi(T) = \phi(\det T).$$

Thus if $\det T = 1$ we know $\Phi(T) = 1$. For arbitrary T put $\gamma := \det T \neq 0$. For $\alpha \neq 0$ let

$$D_\alpha : (x_1, \ldots, x_d) \mapsto (x_1, \ldots x_{d-1}, \alpha x_d).$$

Obviously $\det D_\alpha = \alpha$. Putting $S := T \circ D_{1/\gamma}$ we see that $\det S = 1$ with $T = S \circ D_\gamma$. Since D_γ is translation- invariant in the first $d-1$ coordinates and the d'th coordinate is stretched by a factor γ we see that

$$D_\gamma(\lambda^d) = \frac{1}{|\gamma|}\lambda^d.$$

This is immediately checked for intervals and goes through for arbitrary sets by the standard arguments. Since Φ is a homorphism with $\Phi(S) = 1$ we obtain

$$\Phi(T) = \Phi(S \circ D_\gamma) = \Phi(S)\Phi(D_\gamma) = \Phi(D_\gamma) = \frac{1}{|\gamma|} = \frac{1}{|\det T|}.$$

■

Exercise 5.10 *Check that Φ defined in (5.12) fulfills (5.13).*

Exercise 5.11 *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a linear mapping with $\det T = 0$. Show that for all $A \in \mathcal{B}^d$ there is a measurable set $N \subset \mathbb{R}^d$ with $\lambda^d(N) = 0$ and $T(A) \subseteq N$. Hence (5.11) holds true in a certain sense.*

6 First steps towards integration

In this section we will start to construct an integral from a given measure. But why would we do so? Let us consider

Example 6.1 *An option is the right to buy an option at a certain time T_0 and for a certain price P_0. We buy this option a time 0 for price p. At this point of time the price of the stock is known to be K_0. After that it develops randomly such that at time T_0 its price can be described by a probability measure μ on $\mathbb{R}^+$. How much can we expect to gain from the option we bought? Obviously, if the price of the stock at time T_0, which we denote by K_{T_0} is less than P_0 we do not execute our option because we can get the stock at a lower price on the the stock market. Otherwise we gain $K_{T_0} - P_0$. Taking into account that we paid p for the option our total win or loss is thus given by*

$$f(K_{T_0}) = (K_{T_0} - P_0)^+ - p$$

where $(x)^+ := \max(x, 0)$. Since K_{T_0} is random our average win will be given by "summing f with weights $\mu(\cdot)$". Now $\mathbb{R}$ is uncountable so we would rather want to compute

$$\int f(K_{T_0}) d\mu(K_{T_0}),$$

if we just knew what this was. An integration theory will thus also be a first step into the theory of option pricing, which we will treat in much greater detail, when we also know, how to describe the measure μ appearing in this example.

For good reasons (e.e. since the supremum of a series of $\mathbb{R}$-valued functions make also take the value infinity) we aim at integrating functions with values in $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. The Borel σ-algebra there is defined as

$$\overline{\mathcal{B}}^1 := \{B, B \cup \{\infty\}, B \cup \{-\infty\}, B \cup \{\infty\} \cup \{-\infty\} | B \in \mathcal{B}^1\}.$$

Is is obvious that $\overline{\mathcal{B}}^1$ is a σ-algebra with

$$\mathbb{R} \cap \overline{\mathcal{B}}^1 := \mathcal{B}^1.$$

A function $f : \Omega \to \overline{\mathbb{R}}$ will then be called a numeric function if it is $\mathcal{A} - \overline{\mathcal{B}}^1$-measurable.

Example 6.2 *The most relevant class of numeric functions for the following is the class of indicators*

$$1_A(\omega) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

for $A \in \mathcal{A}$.

Exercise 6.3 *Check that the following rules for computations with indicators hold true*

- $A \subseteq B \Rightarrow 1_A \leq 1_B \quad A, B \in \mathcal{A}$.
- $1_{\bigcup_{i=1}^{\infty} A_i} = \sup_i 1_{A_i} \quad A_i \in \mathcal{A}$.
- $1_{\bigcap_{i=1}^{\infty} A_i} = \prod_{i=1}^{\infty} 1_{A_i} \quad A_i \in \mathcal{A}$.

Let us first establish a criterion for numeric functions.

Proposition 6.4 *Let* $(\Omega, \mathcal{A})$ *be a measurable space. Then* $f : \Omega \to \overline{\mathbb{R}}$ *is* $\mathcal{A} - \overline{\mathcal{B}}^1$*-measurable, if and only if*

$$\{f \geq \alpha\} \in \mathcal{A} \qquad \text{for all } \alpha \in \mathbb{R} \tag{6.1}$$

Proof. We only need to show that $\overline{\mathcal{J}} := \{[\alpha, \infty], \alpha \in \mathbb{R}\}$ generates $\overline{\mathcal{B}}^1$. First note that $[a, b[= [a, \infty] \setminus [b, \infty]$. This implies $\sigma(\mathcal{J}) = \mathcal{B}^1 \subseteq \overline{\mathcal{J}}$. Moreover $\overline{\mathcal{J}}$ contains the points $\{\infty\} := \bigcap_{n=1}^{\infty} [n, \infty]$ and $\{-\infty\} := \overline{\mathbb{R}} \setminus \bigcup_{n=1}^{\infty} [-n, \infty]$. This does the job. ■

Exercise 6.5 *Show that the following are equivalent*

1. f *is* $\mathcal{A} - \overline{\mathcal{B}}^1$*-measurable.*
2. $\{f \geq \alpha\} \in \mathcal{A} \qquad$ *for all* $\alpha \in \mathbb{R}$.
3. $\{f > \alpha\} \in \mathcal{A} \qquad$ *for all* $\alpha \in \mathbb{R}$.
4. $\{f < \alpha\} \in \mathcal{A} \qquad$ *for all* $\alpha \in \mathbb{R}$.
5. $\{f \leq \alpha\} \in \mathcal{A} \qquad$ *for all* $\alpha \in \mathbb{R}$.

Proposition 6.6 *Let* f, g *be numeric functions. Then* $\{f \leq g\}$, $\{f < g\}$, $\{f = g\}$, *and* $\{f \neq g\}$ *are measurable.*

Proof. Since $\mathbb{Q}$ is countable the assertion follows from

$$\{f < g\} = \bigcap_{q \in \mathbb{Q}} \{f < q\} \cap \{q < g\}$$

and

$$\{f \le g\} = \{g < f\}^c, \{f = g\} = \{g \le f\} \cap \{f \le g\}, \{f \ne g\} = \{f = g\}^c.$$

■

Proposition 6.7 *Let f, g be numeric functions. Then also $f \pm g$ and $f \cdot g$ – if well defined – are measurable.*

Proof. First note that with g also $-g$ is measurable, since $\{-g \le \alpha\} = \{g \ge -\alpha\}$. Thus with $f + g$ also $f - g$ will be measurable. Moreover with g also $g + t$, $t \in \mathbb{R}$ is measurable, since $\{g + t \le a\} = \{g \le \alpha - t\}$. Now for real functions f, g

$$\{f + g \le \alpha\} = \{f \le -g + \alpha\}$$

which according to the above and Proposition 6.6 proves measurability of $f \pm g$. Moreover

$$f \cdot g = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right].$$

Now with f also f^2 is measurable, since

$$\{f^2 \le \alpha\} = \{f \le \sqrt{\alpha}\} \cap \{f \ge -\sqrt{\alpha}\}.$$

This shows the assertion for real functions. The extension to numeric functions is immediate, since, if well defined, the sets $\{f \pm g = \pm\infty\}$ and $\{f \cdot g = \pm\infty\}$ are measurable. ■

Theorem 6.8 *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of numeric functions on a measurable set $(\Omega, \mathcal{A})$. Then the following functions are measurable with respect to $\mathcal{A}$:*

$$\sup_n f_n, \inf_n f_n, \limsup f_n, \liminf_n f_n$$

Proof. $\sup f_n$ is measurable, because

$$\{\sup_n f_n \le \alpha\} = \bigcap_{n=1}^{\infty} \{f_n \le \alpha\}.$$

Moreover $\inf_n f_n = -\sup_n -f_n$, $\limsup f_n = \inf_n \sup_{m \ge n} f_m$, and $\liminf_n f_n = -\limsup -f_n$. Thus all the functions are measurable. ■

Corollary 6.9 *Let $f, (f_n)_{n\in\mathbb{N}}$ be measurable functions on a measurable set $(\Omega, \mathcal{A})$. Then for all n*

$$\inf_{m=1\ldots n} f_m \quad \text{and} \quad \sup_{m=1\ldots n} f_m \tag{6.2}$$

and (if existent)

$$\lim_n f_n \tag{6.3}$$

and $|f|$ are measurable.

Proof. (6.2) follows by setting $f_k = f_n$ for all $k \geq n$ in the above theorem. For (6.3) recall that, if $\lim f_n$ exists, then $\lim f_n = \limsup f_n$. Eventually

$$|f| = f^+ - f^-$$

where $f^+ = \max(f, 0)$ and $f^- = (-f)^+$ are measurable. ■

Exercise 6.10 *Show that the measurability of $|f|$ in general does not imply the measurability of f.*

7 The construction of the integral

We will start to construct the integral inspired by the idea which also to the Riemann integral: If $(\Omega, \mathcal{A})$ is a measurable space with measure μ, and $A \in \mathcal{A}$, then the indicator function 1_A geometrically describes a "rectangle" in $\Omega \times \mathbb{R}$ with sidelengths $\mu(A)$ and 1. Hence

$$\int 1_A d\mu$$

should be $\mu(A)$. Yet we will see that therer is an importante difference between the integral we are going to cunstruct and the Riemann integral. This basically can be described by the fact that the above idea also works, if A is just measurable but not an interval as in the case of the Riemann integral. Let us start with the defintion of the first functions we want to integrate:

Definition 7.1 *Let $(\Omega, \mathcal{A})$ be a measurable space. A step function is a function of the form*

$$f(\omega) = \sum_{i=1}^{n} \alpha_i 1_{A_i}(\omega).$$

Here $A_i \in \mathcal{A}$, $i = 1, \dots, n$ are pairwise disjoint with $\bigcup A_i = \Omega$ and $\alpha_i \geq 0$. The set of step functions on $(\Omega, \mathcal{A})$ will be abbreviated by

$$E := E(\Omega, \mathcal{A}). \tag{7.1}$$

Definition 7.1 and the considerations of Section 6 imply

Lemma 7.2 *Let $u, v \in E$, $\alpha \in \mathbb{R}^+$, then $\alpha u, u+v, \max\{u, v\}, \min\{u, v\} \in E$.*

Proof. Obvious from Section 6. ■

We will now prepare the integration of step functions

Lemma 7.3 *Let $u \in E$:*

$$u = \sum_{i=1}^{m} \alpha_i 1_{A_i} = \sum_{i=1}^{n} \beta_i 1_{B_i}.$$

Then

$$\sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \beta_i \mu(B_i).$$

Proof. Note that

$$\bigcup B_i = \bigcup A_i = \Omega$$

and the (A_i) (and the (B_i) respectively) are pairwise disjoint. Then $A_i = \bigcup_{j=1}^{n} (A_i \cap B_j)$ as well as $B_j = \bigcup_{i=1}^{m} (A_i \cap B_j)$. Additivity of μ yields

$$\mu(A_i) = \sum_{j=1}^{n} \mu(A_i \cap B_j) \quad \text{ and } \mu(B_j) = \sum_{i=1}^{m} \mu(A_i \cap B_j).$$

Taking on consideration that on $(A_i \cap B_j)$ we have that $\alpha_i = \beta_j$ we obtain

$$\begin{aligned} \sum_{i=1}^{m} \alpha_i \mu(A_i) &= \sum_{i,j} \alpha_i \mu(A_i \cap B_j) = \sum_{i,j} \beta_i \mu(A_i \cap B_j) \\ &= \sum_{i=1}^{n} \beta_i \mu(B_i) \end{aligned}$$

■

Definition 7.4 *Let $u \in E$ with representation*

$$u = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{7.2}$$

Then

$$\int u d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

is called the (μ)-integral of u. Here μ is a measure on $(\Omega, \mathcal{A})$. $\int u d\mu$ is independent of the special representation chosen in (7.2).

Proposition 7.5 *Let $u, v \in E, \alpha \in \mathbb{R}^+$, and $A \in \mathcal{A}$. Then*

$$\int 1_A d\mu = \mu(A), \tag{7.3}$$

$$\int \alpha u d\mu = \alpha \int u d\mu, \tag{7.4}$$

$$\int u + v d\mu = \int u d\mu + \int v d\mu, \tag{7.5}$$

$$u \leq v \Rightarrow \int u d\mu \leq \int v d\mu \tag{7.6}$$

Proof. (7.3) and (7.4) are obvious.
For (7.5) consider two representations

$$u = \sum_{i=1}^{m} \alpha_i 1_{A_i} \qquad v = \sum_{j=1}^{n} \beta_j 1_{B_j}. \tag{7.7}$$

Then $u + v$ is a step function again with

$$u + v = \sum_{i,j} (\alpha_i + \beta_j) 1_{A_i \cap B_j}.$$

Since $A_i = \bigcup_{j=1}^n (A_i \cap B_j)$ as well as $B_j = \bigcup_{i=1}^m (A_i \cap B_j)$ and $(A_i)_i$ and $(B_j)_j$ form a partition we obtain

$$\begin{aligned}
\int u + v d\mu &= \sum_{i,j} \alpha_i + \beta_j \mu(A_i \cap B_j) \\
&= \sum_{i,j} \alpha_i \mu(A_i \cap B_j) + \sum_{i,j} \beta_j \mu(A_i \cap B_j) \\
&= \sum_{i} \alpha_i \mu(A_i) + \sum_{j} \beta_j \mu(A_i B_j) \\
&= \int u d\mu + \int v d\mu.
\end{aligned}$$

For (7.6) we start again with the decomposition (7.7) and write

$$u = \sum_{i,j} \alpha_i 1_{A_i \cap B_j} \text{ and } v = \sum_{i,j} \beta_j 1_{A_i \cap B_j}.$$

Then $u \leq v$ implies that $\alpha_i \leq \beta_j$ on $A_i \cap B_j \neq \emptyset$. Thus

$$\int u d\mu = \sum_{i,j} \alpha_i \mu(A_i \cap B_j) \leq \sum_{i,j} \beta_j \mu(A_i \cap B_j) = \int v d\mu.$$

■

Eventually we see that if $u \in E$ and $A_i \in \mathcal{A}$, $\alpha_i \in \mathbb{R}^+$ are such that

$$u = \sum_{i=1}^n \alpha_i 1_{A_i}$$

(note that we do not require the $(A_i)_i$ to be a partition anymore, then (7.3) - (7.5) imply that still $\int u d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$.

Exercise 7.6 *Is Dirichlet's function defined by*

$$1_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \textit{otherwise} \end{cases}$$

a step function. If so what is its integral?

We will now turn to defining the integral of a numeric function. This will first be done for positive functions, since afterwards we will decompose an arbitrary function into positive and negative part. The key idea will be to approximate a given function by step functions. Of course, we also need to check afterwards that also the corresponding integrals of the step functions converge to a limit. In this case we can define the integral of the given function as the limit of the integrals of the approximating step functions. The major difference between this construction, which is due to Lebesgue and the Riemann integral is that we start with a measure on $\mathcal{A}$. Hence we do not need indicators on nice sets, such as intervals to approximate a given function.

The following lemma plays a key role for our further steps

Lemma 7.7 *Let $u, (u_n)_n \in E$ and let (u_n) be increasing. Then*

$$u \leq \sup_n u_n \Rightarrow \int u d\mu \leq \sup_n \int u_n d\mu.$$

Proof. Let u be given by

$$u = \sum_{i=1}^m \alpha_i 1_{A_i}$$

with $A_i \in \mathcal{A}$ and $\alpha_i \in \mathbb{R}^+$. Let $\alpha \in (0,1)$. As u_n is measurable

$$B_n := \{u_n \geq \alpha u\} \in \mathcal{A}.$$

By definition of B_n: $u_n \geq \alpha u 1_{B_n}$, hence

$$\int u_n d\mu \geq \alpha \int u 1_{B_n} d\mu.$$

Now (u_n) in increasing and $u \leq \sup_n u_n$. Thus $B_n \uparrow \Omega$ and also $A_i \cap B_n \uparrow A_i$ for all i. Hence continuity from below of μ gives

$$\begin{aligned} \int u d\mu &= \sum_{i=1}^m \alpha_i \mu(A_i) = \lim_{n\to\infty} \sum_{i=1}^m \alpha_i \mu(A_i \cap B_n) \\ &= \lim_{n\to\infty} \int u 1_{B_n} d\mu. \end{aligned}$$

This yields

$$\begin{aligned}\sup_n \int u_n d\mu &\geq \alpha \int u 1_{B_n} d\mu = \alpha \lim_{n\to\infty} \int u 1_{B_n} d\mu \\ &= \alpha \int u d\mu.\end{aligned}$$

Since $\alpha \in (0,1)$ was arbitrary this proves the lemma. ■

Corollary 7.8 *For two increasing sequences (u_n) and (v_n) in E it holds that*

$$\sup u_n = \sup v_n \Rightarrow \sup_n \int u_n d\mu = \sup_n \int v_n d\mu \tag{7.8}$$

Proof. We have that $u_n \leq \sup v_m$ and $v_m \leq \sup u_n$ for all n, m. Thus the result is implied by the previous lemma. ■

Definition 7.9 *Let $E^*(\Omega, \mathcal{A}) = E^*$ denote the set of all numeric functions, such that there is an increasing sequence (u_n) in E such that*

$$f = \sup u_n.$$

The point is that Corollary 7.8 tells us that $\sup_n \int u_n d\mu$ is independent of the particular choice of the sequence (u_n) by which we approximate f. We may thus define

Definition 7.10 *Let $f \in E^*$. We define the integral of f with respect to μ by*

$$\int f d\mu = \sup_n \int u_n d\mu$$

where (u_n) is an increasing sequence in E such that

$$f = \sup u_n.$$

Remark that $E \subseteq E^*$.

Exercise 7.11 *Check that the following holds for $f, g \in E^*, \alpha \in \mathbb{R}^+$.*

$$\alpha f, f+g, f \cdot g, \min f, g, \max f, g \in E^*$$

$$\int \alpha f d\mu = \alpha \int f d\mu,$$

$$\int f + g d\mu = \int f d\mu + \int g d\mu,$$

$$f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$$

Exercise 7.11 tell that the integral is a positive,increasing linear form on the space of integrable functions. An approach due to Daniell starts with such forms and interprets them as integrals We cannot follow this track here.
It might not be too surprising that Lemma 7.7 is inherited by functions $f \in E^*$. However the following that goes Back to B. Levi is an essential improvement over theorems that try to combine limits of functions with Riemann integration.

Theorem 7.12 *(monotone convergence):*
Let $(f_n)_n$ be an increasing sequence in E^*. Then $\sup f_n \in E^*$ and

$$\int \sup f_n d\mu = \sup \int f_n d\mu. \tag{7.9}$$

Proof. Put $f := \sup f_n$. For $f_n \in E^*$ there is a sequence $(u_{m,n})_m$ in E that increases with limit f_n. Lemma 7.2 shows that

$$v_m := \max (u_{m,1}, ...u_{m,m}) \in E.$$

Since the $(u_{m,n})$ are increasing, so is $(v_m)_m$. Apparently $v_m \leq f_m$, thus $v_m \leq$. On the other hand for $m \leq n$ we have $u_{m,n} \leq v_m$ and hence

$$\sup_{m \in \mathbb{N}} u_{mn} = f_n \leq \sup_{m \in \mathbb{N}} v_m.$$

Therefore $\sup v_m = f$, which means that $f \in E^*$. But then $\int f d\mu = \sup \int v_n d\mu$ by definition. Now $v_n \leq f_n$ implies $\int v_n d\mu \leq \int f_n d\mu$. This shows that

$$\int f d\mu \leq \int f_n d\mu.$$

But the converse inequality $\int f d\mu \geq \int f_n d\mu$ is obvious from $f_n \leq f$ for all $n \in \mathbb{N}$. ■

Corollary 7.13 *Let* $(f_n)_n$ *be a sequence in* E^**, then*

$$\sum_{n=1}^{\infty} f_n \in E^*$$

and

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Proof. Apply Theorem 7.12 to $(f_1 + ... + f_n)_n$. ■

Exercise 7.14 *Show that for the Dirac-measure* δ_ω *it holds*

$$\int f d\delta_\omega = f(\omega)$$

for all $f \in E^*$*.*

Example 7.15 *Let* $\Omega = \mathbb{N}$ *and* $\mathcal{A} = \mathcal{P}(\Omega)$*. Because of* σ*- additivity a measure* μ *on* $(\mathbb{N}, P(\mathbb{N}))$ *is uniquely determined by* $\alpha_n := \mu(\{n\})$*. In this case* $f \in E^*$*, whenever it is a positive numeric function. Indeed put* $f_n = f \cdot 1_{\{n\}}$*; then* $f_n \in E^*$ *and* $f = \sum_{n=1}^{\infty} f_n$*. So Corollary 7.13 shows that* $f \in E^*$ *and*

$$\int f d\mu = \sum_{n=1}^{\infty} f(n)\, \alpha_n.$$

The above example raises the question how large a subset of the numeric functions the set E^* is. The answer is somewhat surprising, though easy to verify.

Theorem 7.16 $f \in E^*$ *if and only if* f *is a positive, numeric function.*

Proof. One direction is obvious. For the other let

$$A_{in} = \begin{cases} \left\{f \geq \frac{i}{2^n}\right\} \cap \left\{f < \frac{i+1}{2^n}\right\} & i = 0, ..., n2^n - 1 \\ \{f \geq n\} & i = n2^n \end{cases}$$

and

$$u_n = \sum_{i=0}^{n2^n} \frac{i}{2^n} 1_{A_{in}}.$$

Obviously $u_n \in E$ and $(u_n)_n$ is increasing. Eventually $\sup u_n = f$, since either $f(u) = \infty$,then $u_n(\omega) = n$ for all n, or $|u_n(\omega) - f(\omega)| < \frac{1}{2^n}$, otherwise, for n large enough. ■

Exercise 7.17 *Prove that every bounded numeric function on* $(\Omega, \mathcal{A})$ *is the uniform limit of an increasing sequence of stepfunctions.*

In a final step we can now define what we mean with the integral of an arbitrary numeric function. Recall that for a numeric function f we may introduce

$$f^+ := \max(f, 0) \quad \text{and} \quad f^- := \max(-f, 0).$$

Then $f = f^+ - f^-$. The additivity requirement for integrals immediately leads to the following

Definition 7.18 *A numeric function is called integrable, if* $\int f^+ d\mu$ *and* $\int f^- d\mu$*.are real numbers. Then*

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu. \tag{7.10}$$

Remark 7.19 1. *(7.10) also makes sense, if one of* $\int f^+ d\mu$ *and* $\int f^- d\mu$ *are infinite (but not both). In this case one talks about quasi - integrability.*

2. *If* $\Omega = \mathbb{R}^d$*,*$\mathcal{A} = B^d$*, and* $\mu = \lambda^d$ *one also talks about the Lebesgue - integral.*

We conclude this section by discussing integrability properties of functions:

Theorem 7.20 *The following are equivalent for a numeric function* f*.*

1. f *is integrable*
2. *There are integrable functions* $u \geq 0$ *and* $v \geq 0$*, such that* $f = u - v$*.*
3. *There is a integrable function* g*, such that* $|f| \leq g$
4. $|f|$ *is integrable.*

Proof. : "$1 \Longrightarrow 2$": If f is integrable, so are f^+ and f^-, hence $u = f^+$ and $v = f^-$ does the job.

"$2 \Longrightarrow 3$": Note that $f = u - v \leq u + v$ and $-f = v - u \leq u + v$. Thus $|f| \leq g := u + v$ and g is integrable.

"$3 \Longrightarrow 4$": This follows, since the integral is an increasing function: $\int |f| \, d\mu \leq \int g d\mu < \infty$.

"$4 \Longrightarrow 1$": Since $f^+ \leq |f|$ and $f^- \leq |f|$, the integrability of $f^+$and f^- follows again from monotonicity. ■

Exercise 7.21 *Let f,g be integrable, $\alpha \in \mathbb{R}$. Show that αf,$f+g$, $\max(f,g)$, $\min(f,g)$ are integrable and that*

$$\int \alpha f d\mu = \alpha \int f d\mu \quad \textit{and} \quad \int f + g d\mu = \int f d\mu + \int g d\mu.$$

Proposition 7.22 *Let f, g be integrable. Then*

$$f \leq g \Longrightarrow \int f d\mu \leq \int g d\mu \tag{7.11}$$

$$\left| \int f d\mu \right| \leq \int |f| \, d\mu \tag{7.12}$$

(7.12) is called the triangle - inequality.

Proof. $f \leq g$ implies that $f^+ \leq g^+$ and $f^- \geq g^-$. Hence (7.11) follows from the monotonicity of the integral on E^*.
(7.12) is a special case of (7.11), since $f \leq |f|$ as well as $-f \leq |f|$. ■

Definition 7.23 *Let $(\Omega, \mathcal{A})$ be a measurable space and μ a measure on it. Then*

$$\mathcal{L}^1(\mu) := \{f : \Omega \to \mathbb{R}, f \textit{ is integrable}\}$$

is defined to be the space of all integrable functions.

Exercise 7.24 *Prove that $\mathcal{L}^1(\mu)$ together with (pointwise) addition and multiplication with real numbers is a real vector space.*

Example 7.25

1. *Consider the situation of Example 7.15. From what was shown there it evident that*

$$\mathcal{L}^1(\mu) := \left\{ f : \mathbb{N} \to \mathbb{R} : \sum_{n=1}^{\infty} |f(n)| \, \alpha_n < \infty \right\}.$$

2. *Let $(\Omega, \mathcal{A}, \mu)$ be such that μ is finite, i. e. $\mu(\Omega) < \infty$. Then*

$$\{f : f \equiv const\} \subseteq \mathcal{L}^1/\mu.$$

Hence due to Theorem 7.20 also

$$\{f : \Omega \to \mathbb{R}, f \textit{ is bounded } \} \subseteq \mathcal{L}^1(\mu).$$

3. *Let μ, ν be measure on $(\Omega, \mathcal{A})$. $f : \Omega \to \bar{\mathbb{R}}$ (a numeric function) is $(\mu + \nu)$ - integrable, if and only if it's μ - integrable and ν - integrable. Then*
$$\int f d(\mu + \nu) = \int f d\mu + \int f d\nu. \tag{7.13}$$
Indeed, if f is a step function, this is obvious. If $f \in E^*$ the usual approximation step yields (7.13). Eventually, if f is arbitrary we decompose $f = f^+ - f^-$ as usual. In particular
$$\mathcal{L}(\mu + \nu) = \mathcal{L}(\mu) \cap \mathcal{L}(\nu).$$

So far we have just considered integration over the whole space Ω. Integrals over measurable sets $A \subseteq \Omega$ are indeed easy to define:

Definition 7.26 *Let $f \in E^* \cup \mathcal{L}^1(\mu)$. Then for $A \in \mathcal{A}$ we define the integral over A as*
$$\int_A f d\mu := \int f 1_A d\mu.$$
In particular $\int f d\mu = \int_\Omega f d\mu$.

Exercise 7.27 *Check the following:*
$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu - \int_{A \cap B} f d\mu \tag{7.14}$$
and
$$f|A \leq g|A \Longrightarrow \int_A f d\mu \leq \int_A g d\mu. \tag{7.15}$$

It can be shown and is intuitively clear that $\int_A f d\mu$ can also be defined by restricting $(\Omega, \mathcal{A}, \mu)$ to A, i.e. by considering A furnished with the σ - algebra $\mathcal{A} \cap A := \{B \cap A : B \in \mathcal{A}\}$ and the measure $\mu \mid A(B) := \mu(B \cap A), B \in \mathcal{A} \cap A$. We will refrain from giving a proof for this obvious fact here.

Exercise 7.28 *Let $(\Omega, \mathcal{A})$ be a measurable space and μ a finite measure on $(\Omega, \mathcal{A})$. Show that if f is the uniform limit of a sequence $(f_n)_{n \in \mathbb{N}}$ in $\mathcal{L}^1(\mu)$, then $f \in \mathcal{L}^1(\mu)$. Why is finiteness of μ necessary?[Hint: Construct $g_n \in \mathcal{L}^1(\mu)$ with $0 \leq g_n \leq 1$ and $\int g_n d\mu \geq n^2$ and consider $f_n := \sum_{i=1}^n i^{-2} g_i$].*

8 Almost - everywhere existing properties

A closer look at the last section reveals that nothing really changes, if we modify e.g. a given function on a set of measure 0. This will be formalized and discussed in the present section.

Definition 8.1 *We will say that a property is true μ - almost everywhere on a measurable space $(\Omega, \mathcal{A})$ with measure μ, if there is a set N with $\mu(N) = 0$, such that the property holds true on $\Omega \backslash N$.*

Example 8.2 *The Dirichlet function $1_{\mathbb{Q}}$ is zero λ^1 - almost everywhere, we also write*

$$1_{\mathbb{Q}} = 0 \qquad \lambda^1 - a.e.$$

(this is true since $\lambda^1[\mathbb{Q}] = 0$).

The following theorem in a modified form is well known for Riemann integration

Theorem 8.3 *Let $f \in E^*(\Omega, \mathcal{A})$. Then*

$$\int f d\mu = 0 \Leftrightarrow f = 0 \qquad \mu - a.e.$$

Proof. Since f is measurable

$$N := \{f \neq 0\} = \{f > 0\} \in \mathcal{A}.$$

We show

$$\int f d\mu = 0 \Leftrightarrow \mu(N) = 0.$$

First let $\int f d\mu = 0$. Now $A_n := \left\{f \geq \frac{1}{n}\right\} \in \mathcal{A}$ for all $n \in \mathbb{N}$. Furthermore $A_n \uparrow N$. Obviously $f \geq \frac{1}{n} 1_{A_n}$ and thus

$$0 = \int f d\mu \geq \int \frac{1}{n} 1_{A_n} d\mu = \frac{1}{n} \mu(A_n).$$

Therefore $\mu(A_n) = 0$ for all n which shows that $\mu(N) = 0$.
On the other hand, if $\mu(N) = 0$ then $u_n := n 1_N \in E$ for all $n \in \mathbb{N}$ and $\int u_n d\mu = 0$. Put $g := \sup u_n$, then (by definition) $g \in E^*, u_n \uparrow g$ and $\int g d\mu = \sup \int u_n d\mu = 0$. But $f \leq g$ which shows that $\int f d\mu = 0$. ∎

Exercise 8.4 *Let f be $\mathcal{A}$-measurable and $N \in \mathcal{A}$ be such that $\mu(N) = 0$. Show that*

$$\int_N f d\mu = 0.$$

Theorem 8.5 *Let f, g be measurable numeric function such that $f = g$ $\mu - a.e$ on Ω. Then*
a) If $f \geq 0$ and $g \geq 0$ then $\int f d\mu = \int g d\mu$
b) If f is integrable, then so is g and $\int f d\mu = \int g d\mu$.

Proof. a) $f = g$ $\mu - a.e.$ implies that $f - g = 0$ $\mu - a.e.$. Hence from Theorem 8.3 we obtain

$$\int (f - g)\, d\mu = 0 \Longrightarrow \int f d\mu = \int g d\mu.$$

b) It follows from $f = g$ $\mu - a.e.$ hence from a) we get $\int f^+ d\mu = \int g^+ d\mu$ and $\int f^- d\mu = \int g^- d\mu$. This proves the result. ∎

Corollary 8.6 *Let f, g be measurable, numeric functions with $|f| \leq g$ $\mu - a.e.$ Then with g also f is μ - integrable.*

Proof. Consider $g' = \max(g, |f|)$. Then $g = g'$ $\mu - a.e.$ Thus also g' is μ - integrable. But $|f| \leq g'$ everywhere. This proves the result. ∎
It seems evident that an integrable function cannot become too large (since otherwise the integral becomes infinite). This is formalized by

Theorem 8.7 *Let f be μ - integrable. Then $|f| < \infty$ $\mu - a.e.$ and $\{f \neq 0\}$ has σ - finite measure.*

Proof. Denote by $N := \{|f| = \infty\} \in \mathcal{A}$. Then for all $\alpha \in \mathbb{R}$ we have $\alpha 1_N \leq |f|$ and hence $\alpha\mu(N) \leq \int |f|\, d\mu < \infty$. This gives $\mu(N) = 0$. For part two we may assume that $f \geq 0$ (since $|f|$ and f have the same zeros). Then

$$\{f \neq 0\} = \{f > 0\} = \bigcup_{n=1}^{\infty} \left\{ f \geq \frac{1}{n} \right\} =: \bigcup_{n=1}^{\infty} A_n.$$

But $1_{A_n} \leq nf$ and hence. Hence

$$\mu(A_n) \leq \int f d\mu < \infty.$$

This proves the theorem, since $\bigcup_{n=1}^{\infty} A_n = \{f \neq 0\}$. ■

Theorem 8.7 has the following consequence: Assume that $f : \Omega' \to \overline{\mathbb{R}}$ is a measurable function on $\Omega' \subseteq \Omega$, such that $\Omega \backslash \Omega' \subseteq N$ with $\mu(N) = 0$. If we then consider an extension $\overline{f}$ of f to all of Ω then either any extension is μ - integrable, or none of them is. This justifies:

Definition 8.8 *Let f be a μ - almost everywhere defined, measurable, numeric function. We will call f μ - integrable, if there is an extension $\overline{f} : \Omega \to \overline{\mathbb{R}}$ of f that is μ - integrable. We define*

$$\int f d\mu := \int \overline{f} d\mu.$$

Exercise 8.9 *For two functions f, g defined on $(\Omega, \mathcal{A}, \mu)$ assume that $f = g$ $\mu -$ a.e.. Show that generally measurability of f does not imply that also g is measurable.*

9 The spaces $\mathcal{L}^p(\mu)$

In Section 7 we already met the space of all integrable functions $\mathcal{L}^1(\mu)$. We saw that it is a vectorspace. To make it also a field $\mathcal{L}^1(\mu)$ needed also to be closed under products, i.e. with $f, g \in \mathcal{L}^1(\mu)$, we would also need to have $f \cdot g \in \mathcal{L}^1(\mu)$. This is not completely impossible, since we already saw, that the product of two measurable functions is measurable again. The following example show that this is not inherited by integrable functions.

Example 9.1 *Let $(\Omega, \mathcal{A}) = (\mathbb{N}, \mathcal{P}(N))$ and $\mu(\{u\}) = \alpha_n = n^{-p-1}$. Define $f(n) = n$.*
For $1 < p < \infty$, $f(n)$ is integrable but $f^p(n)$ is not, e.g. for $p = 2$, f^2 is not integrable.

We will now start to investigate, when $|f|^p$ is integrable. In the sequel $p \geq 1$. Note that for any measurable, numeric function $f : \Omega \to \bar{\mathbb{R}}$, also $|f|^p$ is measurable, since $\{|f|^p \geq \alpha\}$ is either Ω or equal to $\left\{|f| \geq \alpha^{\frac{1}{p}}\right\}$. Thus for every such f the quantity

$$N_p(f) := \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} \tag{9.1}$$

is defined with $0 \leq N_p(f) \leq +\infty$. Evidently,

$$N_p(\alpha f) = |\alpha| N_p(f).$$

Two inequalities are central for $N_p(\cdot)$.

Theorem 9.2 *Let $p > 1$ and q be defined by*

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{9.2}$$

Then for any two measurable, numeric functions f, g on Ω it holds

$$N_1(fg) \leq N_p(f) N_q(g) \tag{9.3}$$

This is called the Hölder inequality.

Proof. Without loss of generality $f \geq 0$ and $g \geq 0$. Put $\sigma := N_p(f)$ and $\tau := N_q(g)$. We may further assume that $\sigma > 0$ and $\tau > 0$. (If, namely $\sigma = 0$ or $\tau = 0$ then $f = 0$ or $g = 0$ μ - almost - surely and thus and thus also $f \cdot g = 0$ $\mu - a.s.$ and hence $N_1(fg) = 0$.) On the other hand, we may also assume $\sigma < +\infty$ and $\tau < +\infty$.
In Exercise 9.3 we will see that Bernoulli's inequality

$$(1+\eta)^{\frac{1}{p}} \leq \frac{\eta}{p} + 1 \quad \text{for all } \eta \in \mathbb{R}^+ \tag{9.4}$$

holds true. Putting $\xi := 1 + \eta$ this yields

$$\xi^{\frac{1}{p}} \leq \frac{\xi}{p} + \frac{1}{q} \quad \text{for all } \xi \geq 1. \tag{9.5}$$

For any two numbers $x, y \in \mathbb{R}^+$ either $x \leq y$ or $y > x$, thus either $xy^{-1} \geq 1$ or $x^{-1}y \geq 1$. Hence, if we put $\xi := \max\{xy^{-1}, x^{-1}y\}$, we know $\xi \geq 1$. Inserting this into (9.5) yields

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{1}{p}x + \frac{1}{q}y \quad \text{for all } x, y \in \mathbb{R}^+. \tag{9.6}$$

Choose $x := (f(\omega)/\sigma)^p$ and $y := (g(\omega)/\tau)^q$ for $\omega \in \Omega$ with $f(\omega) < \infty$ and $g(\omega) < \infty$, then

$$\frac{1}{\sigma\tau} fg \leq \frac{1}{\sigma^p p} f^p + \frac{1}{\tau^q q} g^q. \tag{9.7}$$

(9.7) is obvious on $\{f = +\infty\} \cup \{g = +\infty\}$. Integrating (9.7) yields

$$\int fg d\mu \leq \sigma\tau$$

which is (9.3). ■

Exercise 9.3 *Show that the Bernoulli inequality (9.4) holds true.*

Theorem 9.4 *Let f, g be measurable, numeric functions such that $f + g$ is defined on the entire set Ω. Then for $1 \leq p < \infty$:*

$$N_p(f+g) \leq N_p(f) + N_p(g)\,. \tag{9.8}$$

Proof. Since $|f+g| \leq N_p\,|f| + |g|$ we have

$$N_p(f+g) \leq N_p(|f| + |g|)$$

in thus is suffices to consider $f \geq 0$ and $g \geq 0$. First note that for $p = 1$ (9.8) holds true with equality. So we may assume that $1 < p < +\infty$. Again we choose q such that $\frac{1}{p} + \frac{1}{q} = 1$. Eventually we may assume that $N_p(f)$ and $N_p(g)$ are finite. But then

$$(f+g)^p \leq (2\max(f,g))^p = 2^p \max(f^p, g^p) \leq 2^p (f^p + g^p)\,,$$

which implies that $(f+g)^p$ is integrable and therefore $N_p(f+g) < \infty$. But now:

$$\int (f+g)^p\, d\mu = \int f\,(f+g)^{p-1}\, d\mu + \int (f+g)^{p-1}\, g d\mu. \tag{9.9}$$

Applying Hölder's inequality to both summands on the right hand side of (9.9) yields

$$\begin{aligned}
\int (f+g)^p\, d\mu &\leq N_p(f)\, Nq\left((f+g)^{p-1}\right) \\
&\quad N_p(g)\, Nq\left((f+g)^{p-1}\right) \\
&= (N_p(f) + N_p(g))\, Nq\left((f+g)^{p-1}\right)
\end{aligned}$$

Since $q\,(p-1) = p$ this gives

$$(N_p(f+g))^p \leq [N_p(f) + N_p(g)]\,(N_p(f+g))^{p-1}$$

As $N_p(f+g) < \infty$ this gives (9.8). ■

What we have proven so far are results about p - times integrable functions. They will now be given a name:

Definition 9.5 *A function $f : \Omega \to \overline{\mathbb{R}}$ is called p - times integrable, if it is measurable and $|f|^p$ is integrable. The space of all p - times integrable functions is called $\mathcal{L}^p(\mu)$. In case $p = 2$ we also speak about square - integrability.*

Exercise 9.6 *Let f, g be p - times integrable. Then so is αf for $\alpha \in \mathbb{R}$ and $f + g$, as well as $\max(f, g)$ and $\min(f, g)$. Eventually f is p - times integrable if and only if f^+ and f^- are p - times - integrable.*

Hölder's inequality immediately yields

Theorem 9.7 *The product of a p - times integrable function and a q - times integrable function is integrable if $\frac{1}{p} + \frac{1}{q} = 1$.*

Corollary 9.8 *Let $\mu(\Omega) < \infty$. Then every p - times integrable function is 1 - times integrable, $1 < p < \infty$.*

Proof. Since $\mu(\Omega) < \infty$ the constant function, 1 is q - times integrable. Hence $f = f \cdot 1$ is integrable as a consequence of Theorem 9.7. ■

Exercise 9.9 *Show that the finiteness assumption on μ in Corrollary 9.8 is essential.*

As a consequence of Theorem 9.7 we also show:

Theorem 9.10 *Let $f : \Omega \to \overline{\mathbb{R}}$ p - times integrable $(1 \leq p < \infty)$ and $g : \Omega \to \overline{\mathbb{R}}$ be bounded by some $\alpha \in \mathbb{R}_+$. Then $f \cdot g$ is p -times integrable.*

Proof. : We know $|g| \leq \alpha$. But then $|gf| \leq \alpha |f|$, and $\alpha |f|$ is p - times integrable. This shows the assertion. ■
As a last step we turn to the case $p = 1$. Define

$$\mathcal{L}^\infty(\mu) := \{f : \Omega \to \mathbb{R},\ f \text{ is measure and } \mu - a.e. \text{ bounded}\}$$

Trivially $\mathcal{L}^\infty(\mu)$ is a vector space over $\mathbb{R}$.

10 Convergence Theorems

Here we will discuss the spaces $\mathcal{L}^p(\mu)$ introduced in the last section. The central observation here is that $N_p(f)$ defines a semi - norm on $\mathcal{L}^p(\mu)$ is defined by the properties:

$$\begin{aligned} N_p &: \mathcal{L}^p(\mu) \to \mathbb{R}^+ \\ N_p(\alpha f) &= |\alpha| N_p(f) \quad (10.1) \\ N_p(f+g) &= N_p(f) + N_p(g) \quad (10.2) \end{aligned}$$

(10.2) implies the triangle inequality for

$$d_p(f,g) := N_p(f-g) \quad f, g \in \mathcal{L}^p(\mu).$$

The reason for that $N_p(\cdot)$ is only a semi-norm and $d_p(\cdot,\cdot)$ is only a pseudo-metrics is that $N_p(f) = 0$ does not imply $f \equiv 0$, but only $f = 0\ \mu - a.e.$ (correspondingly $d_p(f,g) = 0$ only implies $f = g\ \mu - a.e.$). We can consider convergence with respect to this semi - norm by saying that f_n converges to f in $\mathcal{L}^p$ (in symbols $f_n \to^{\mathcal{L}^p} f$), if $d_p(f_n, f) \to 0$. Indeed, there is a slight difficulty with this definition, since the limit is not unique - as mentioned above. This is, of course, a nuissance for any sort of convergence. So formally we will work on the quotient space $\frac{\mathcal{L}^p(\mu)}{\simeq_\mu}$ where $f \simeq_\mu g :\Longleftrightarrow f = g$ $\mu - a.e.$ Here the limit with respect to $d_p(f_n, f)$ is unique, since $N_p(\cdot)$is a norm on $\frac{\mathcal{L}^p(\mu)}{\simeq_\mu}$.
As a first step we establish a lemma that is central for all of integration theory and probability theory.

Lemma 10.1 *(Fatou)* *Let $(\Omega, \mathcal{A})$ be a measurable space with measure μ. Then for all sequences $(f_n)_n$ of measurable, positive ($f_n \geq 0$) functions it holds:*

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu. \quad (10.3)$$

Proof. As was shown previously $f := \liminf_{n\to\infty} f_n$ and $g_n := \inf_{m\geq n} f_n$ are in E^*. By definition of the $\liminf g_n \uparrow f$ and therefore

$$\int f d\mu = \sup_{n\in\mathbb{N}} \int g_n d\mu = \lim_{n\to\infty} \int g_n d\mu. \quad (10.4)$$

Eventually, $f_m \geq g_n$ for all $m \geq n$ and hence

$$\int g_n d\mu \leq \inf_{m \geq n} \int f_m d\mu \tag{10.5}$$

(10.5) together with (10.4) gives (10.3). ■
Choosing as $f_n = 1_{A_n}, A_n \in \mathcal{A}, \liminf_{n\to\infty} 1_{A_n}$ is given by

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m. \tag{10.6}$$

Hence $\liminf A_n$ is the set of all $\omega \in \Omega$ that are in all but a finite number of A_m's. Similarly one defines

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m, \tag{10.7}$$

the set of all $\omega \in \Omega$ that are contained in infinitely many of the An's. Easy calculation shows that

$$(\limsup A_n)^c = \liminf (A_n^c).$$

Exercise 10.2 *Derive the following from Fatou's lemma:*
a) $\mu(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mu(A_n)$
b) *If μ is finite, then*

$$\mu\left(\limsup_{n\to\infty} A_n\right) \geq \limsup_{n\to\infty} \mu(A_n).$$

It is surprisingly easy to derive from Fabou's lemma our first convergence result, that relates almost sure convergence to convergence in $\mathcal{L}^p(\mu)$.

Theorem 10.3 *(Riesz) Let $(f_n)_n$ be a sequence in $\mathcal{L}^p(\mu)$ with $f_n \to f \in \mathcal{L}^p(\mu)$ $\quad \mu - a.e.$ Then f_n converges in $\mathcal{L}^p(\mu)$ to f, if and only if*

$$\lim_{n\to\infty} \int |f_n|^p d\mu = \int |f|^p d\mu \tag{10.8}$$

Proof. Note that

$$N_p(f) = N_p(f-g+g) \le N_p(f-g) + N_p(g)$$

and

$$N_p(-g) = N_p(g)$$

imply

$$\mid N_p(f) - N_p(g) \mid \le N_p(f \pm g). \tag{10.9}$$

Hence, if $f_n \to^{\mathcal{L}^p(\mu)} f$ then $N_p(f_n - f) \to 0$ and so (10.8) follows from (10.9). For the converse note that for $\alpha, \beta \in \mathbb{R}^+$ one has $|\alpha - \beta|^p \le (\alpha+\beta)^p \le 2^p(\alpha^p + \beta^p)$. Hence

$$g_n := 2^p(|f_n|^p + |f|^p) - |f_n - f|^p\,, \quad n \in \mathbb{N}$$

defines a sequence of non - negative functions in $\mathcal{L}^1(\mu)$. Since $f_n \to f \ \ \mu{-}a.e.$ we know $g_n \to 2^{p+1}\,|f|^p \ \mu - a.e.$. This is also the $\liminf$ of the g_n's. Thus (10.8) together with Fatou's lemma yields

$$\begin{aligned} \int 2^{p+1}\,|f|^p\,d\mu \quad &= \quad \int \liminf_{n\to\infty} g_n d\mu \le \liminf_{n\to\infty} \int g_n d\mu \\ &=^{(10.8)} 2^{p+1} \quad \int |f|^p - \limsup \int |f_n - f|^p\,d\mu. \end{aligned}$$

Therefore

$$\limsup_{n\to\infty} \int |f_n - f|^p\,d\mu \le 0.$$

■

Our next major goal is to prove a second criterion for when the convergence of f_n to f implies the $\mathcal{L}^p$ - convergence of f_n to f: This will reveal that the Lebesgue - integral is a major improvement over the Riemann - integral. In a preparatory step we generalize the Minkowski - inequality to series of functions.

Lemma 10.4 *Let $(f_n)_n$ be a sequence in $E^*(\Omega, \mathcal{A})$. Then*

$$N_p(\sum_{i=1}^{\infty} f_n) \le \sum_{n=1}^{\infty} N_p\,(f_n) \qquad (1 \le p < \infty)\,. \tag{10.10}$$

Proof. Put $s_n := f_1 + .. + f_n$. Then by Minkowski's inequality:

$$N_p(s_n) \leq \sum_{i=n}^{n} N_p(f_i) \leq \sum_{i=1}^{\infty} N_p(f_i).$$

s_n is increasing and converges to $\sum_{n=1}^{\infty} f_n$ and so the p'th powers. By monotone convergence:

$$N_p(\sum_{n=1}^{\infty} f_n) = \sup_n N_p(s_n)$$

and by what we have seen before thus

$$N_p(\sum_{n=1}^{\infty} f_n) \leq \sum_{n=1}^{\infty} N_p(f_n).$$

■

The following theorem goes back to H. Lebesgue and often is called Lebesgue's convergence theorem or the dominated convergence Theorem.

Theorem 10.5 *Let (f_n) be a sequence in $\mathcal{L}^p(\mu), 1 \leq p < +\infty$, that converges $\mu - a.e.$ on Ω. Let $g : \Omega \to \bar{\mathbb{R}}^+$ be in $\mathcal{L}^p(\mu)$ with*

$$|f_n| \leq g, \qquad n \in \mathbb{N}. \tag{10.11}$$

Then there is $f : \Omega \to \mathbb{R}$ with $f_n \to f \quad \mu. - a.e.$ Every such f is in $\mathcal{L}^p(\mu)$ and $f_n \to^{\mathcal{L}^p} f$.

Proof. First we kick out those sets which just spoil our calculations: there are nullsets M_1, M_2 such that $\lim f_n \in \overline{\mathbb{R}}$ exists on M_1^c and such that $g < \infty$ on M_2^c (for that note that g^p is integrable). Define

$$f(\omega) = \begin{cases} \lim_{n\to\infty} f_n(\omega) & \omega \in (M_1 \cup M_2)^c \\ 0 & \omega \in M_1 \cup M_2 \end{cases}$$

Then $f : \Omega \to \mathbb{R}$ and f is $\mathcal{A}$ - measurable. $f_n \to f \, \mu - a.e.$ Since $|f| \leq g$ $\mu - a.e.$ $\quad |f|^p \in \mathcal{L}^1(\mu)$. We define

$$g_n := |f_n - f|^p$$

and aim to show $\int g_n d\mu = 0$. By definition

$$0 \leq g_n \leq (|f_n| + |f|)^p \leq (|f| + g)^p .$$

Thus with $h := (|f_n| + g)^p$ also g_n is integrable. Applying Fatou's lemma to $(h - g_n)_n$ yields.

$$\begin{aligned} \int \liminf\,(h - g_n)\,d\mu &\leq \liminf \int (h - g_n)\,d\mu \\ &= \int h d\mu - \limsup \int g_n d\mu. \qquad (10.12) \end{aligned}$$

Now $f_n \to f \quad \mu. - a.e.$ and therefore $g_n \to 0 \mu - a.e.$ and thus

$$\int \liminf\,(h - g_n)\,d\mu \to \int h d\mu.$$

Hence (10.12) gives $\limsup \int g_n d\mu \leq 0$. This proves the theorem. ■

Theorem 7.12 and Theorem 10.5 are a major improvement over the case of the Riemann integral. There the only case where one might conclude from $f_n \to f$ that also $\int f_n dx \to \int f dx$ is, if f_n converges to f uniformly. Here the conditions are essentially relaxed. The way we achieved this improvement was to take a partitioning of Ω (which in the case of Lebesgue integration is $\mathbb{R}$) that is more adapted to the given function f than the partitioning by intervals in the Riemann case.
The last theorem told us, that, if f_n converges pointwise to a limit f and f is bounded by a p-integrable function (and f_n, f are p - integrable), then the convergence is in $\mathcal{L}^p$. A natural question to ask is, when there is such a limit f, or, in other words, when does a Cauchy sequence in $\mathcal{L}^p(\mu)$ converge? The answer is given by the following theorem, that for $p = 2$ goes back to F. Riesz and E. Fischer:

Theorem 10.6 *Every Cauchy sequence $(f_n)_n$ in $\mathcal{L}^p(\mu)$ converges in $\mathcal{L}^p$ to some limit $f \in \mathcal{L}^p$. A subsequence of $(f_n)_n$ converges to $f \mu - a.e.$*

Proof. Since (f_n) is a Cauchy sequence, we can find a subsequence $(f_{n_k})_k$ such that

$$N_p(f_{n_{k+1}} - f_{n_k}) \leq 2^{-k}.$$

Put

$$g_k := f_{n_{k+1}} - f_{n_k}$$

and

$$g := \sum_{n=1}^{\infty} |g_k|$$

Then Lemma 10.4 implies

$$N_p(g) \leq \sum_{n=1}^{\infty} N_p(g_k) \leq \sum_{n=1}^{\infty} 2^{-k} = 1.$$

Thus the $\mathcal{A}$ - measurable function g is p - integrable and thus $\mu - a.e.$ finite, i.e. $\sum g_n$ converges $\mu - a.e.$ absolutely. But $\sum_{k=1}^{m} g_k = f_{n_{m+1}} - f_{n_1}$. Thus $(f_{n_k})_k$ converges $\mu - a.e.$ on Ω. Moreover

$$\left|f_{n_{k+1}}\right| = |g_1 + ... + g_n + f_{n_1}| \leq g + |f_{n_1}|\,.$$

$g + |f_{n_1}|$ is p - times integrable. Therefore $(f_{n_k)_k}$ satisfies the conditions of the dominated convergence theorem. Thus there exists $f \in \mathcal{L}^p(\mu)$ such that $f_{n_k} \to^{\mathcal{L}^p} f$. But now $(f_n)_n$ is a Cauchy - sequence. Hence, if the subsequence $(f_{n_k})_k$ converges in $\mathcal{L}^p$ to f, so does $(f_n)_n$. Indeed given $\varepsilon > 0$, there is N_0, such that

$$N_p(f_n - f_m) \leq \varepsilon \quad \text{for all } n, m \geq N_0.$$

On the other hand since $f_{n_k} \to^{\mathcal{L}^p} f$ there is N_1 such that

$$N_p(f_{n_k} - f) \leq \varepsilon \quad \text{for all } n_k \geq N_1$$

Now for $N = \max(N_0, N_1)$ we obtain if $n, n_k \geq N$ that

$$N_p(f_n - f) \leq N_p(f_n - f_{n_k}) + N_p(f_{n_k} - f) \leq 2\varepsilon$$

■

Let us now illustrate that in general a sequence that converges in $\mathcal{L}^p$ does not need to converge $\mu - a.e.$

Example 10.7 *Let $\Omega = [0, 1[$ and $\mu := \lambda^1$ on $\mathcal{A} := B^1 \cap \Omega$. For $n \in \mathbb{N}$ choose $k, h \in \mathbb{N}$ such that $n = 2^h + k < 2^{h+1}$. This choice is unique. Put*

$$A_n : [k2^{-h}, (k+1)2^{-h}[\text{ and } f_n := 1_{A_n}.$$

Then

$$\int f_n^p d\mu = \int f_n d\mu = \mu(A_n) = 2^{-h}$$

Since $h \to \infty$ as $n \to \infty$ the sequence f_n converges to zero in $\mathcal{L}^p(\lambda^1)$ for all p. On the other hand: for no $\omega \in \Omega$ is (f_n) convergent. Indeed, for $\omega \in \Omega$ and $h = 0, 1, 2, ...$ there is a unique k with $\omega \in A_{2^h+k}$. If $k < 2^h - 1$ we have $\omega \notin A_{2^h+k+1}$, if $h = 2^h - 1$ and $h \geq 1$ we have $\omega \notin A_{2^{h+1}}$

On the other hand almost everywhere convergence does not imply $\mathcal{L}^p$ - convergence either:

Example 10.8 *Let* $\Omega = [0, 1[, \mathcal{A} = \mathcal{B}^1 \cap \Omega$ *and* $\mu = \lambda^1 \mid \Omega$. *Choose* $An = [0, \frac{1}{n}[$ *and* $f_n = n^2 1_{[0,\frac{1}{n}[} = n^2 1_{A_n}$. *Then* $f_n \to 0 \quad \lambda^1 -$ *a.e., but for every* $1 \leq p < \infty$.

$$\int |f_n|^p \, d\mu = n^p \to_{n\to\infty} \infty.$$

Thus $N_p(f_n)$ *does not converge to zero.*

The following exercise follows from Theorem 10.6:

Exercise 10.9 *Let* (f_n) *be a Cauchy sequence in* $\mathcal{L}^p(\mu)$ *that converges* $\mu-$*a.e. to a real,* $\mathcal{A}$ *- measurable function* $f : \Omega \to \mathbb{R}$. *Then* $f \in \mathcal{L}^p(\mu)$ *and* $f_n \to f$ *in* $\mathcal{L}^p(\mu)$.

We already learned that the spaces $\mathcal{L}^p$ are not closed under products. But if f, g are in some $\mathcal{L}^p$ and $\mathcal{L}^q$ then one might well say something about $f \cdot g$. This is also true for convergence results:

Theorem 10.10 *Let* $(f_n)_n$ *be a sequence in* $\mathcal{L}^p(\mu)$ *converging in* $\mathcal{L}^p(\mu)$ *to some* $f \in \mathcal{L}^p(\mu)$, *and let* $(g_n)_n$ *be a sequence in* $\mathcal{L}^q(\mu)$ *converging in* $\mathcal{L}^q(\mu)$ *to some* $g \in \mathcal{L}^q(\mu)$. *Here* $\frac{1}{p} + \frac{1}{q} = 1$ *then* $(f_n g_n)_n$ *converges to* $f \cdot g$ *in* $\mathcal{L}^1(\mu)$.

Proof. From the triangle inequality one gets:

$$|f_n g_n - fg| \leq |f_n - f| \, |g_n| + |f| \, |g_n - g|$$

Applying Hölder's inequality yields:

$$N_1(f_n g_n - fg) \leq N_p(f_n - f) N_q(g_n) + N_p(f) N_q(g_n - g).$$

This gives the assertion of the theorem. ■
The following is of particular importance in probability theory.

Exercise 10.11 *Let* μ *be finite. Show that for* $1 \leq p' \leq p < \infty$ *and a measurable, numeric function*

$$N_{p'}(f) \leq N_p(f) \mu^{\frac{1}{p'} - \frac{1}{p}}(\Omega)$$

and

$$\mathcal{L}^p(\mu) \subseteq \mathcal{L}^{p'}(\mu).$$

Moreover show that convergence in $\mathcal{L}^p(\mu)$ *implies convergence in* $\mathcal{L}^{p'}(\mu)$.

Exercise 10.12 *Show that Theorem 10.10 also holds true for $p = 1$ and $q = \infty$.*

A last remark concerning the spaces $\mathcal{L}^p(\mu)$. Consider

$$L^p(\mu) = \frac{\mathcal{L}^p(\mu)}{\simeq \mu}$$

where $\simeq \mu$ is the equivalence modulo $\mu - a.e.$ equality, i.e. $f \simeq_\mu g :\Leftrightarrow f = g$ $\mu - a.e.$ Then it is seen that $L^p(\mu)$ is a vector space with norm

$$\left\|\tilde{f}\right\|_p = \int |\tilde{f}|^p d\mu$$

where $\tilde{f} \in L^p(\mu)$. $L^2(\mu)$ even is a Hilbert space with inner product

$$< \tilde{f}, \tilde{g} >:= \int \tilde{f}\tilde{g} d\mu.$$

This goes beyond the scope of this course. As a consequence of the above we will now discuss the realtion between Riemann - and Lebesgue - integration in greater details. We will first treat the case of integrals in compact intervals $[a, b]$.

Theorem 10.13 *Let $f : [a, b] \to \mathbb{R}$ be a Borel - measurable function. If f is Riemann - integrable, then it is also Lebesgue - integrable and the two integrals coincide.*

Proof. Let $\tau : a = \alpha_0 \leq \alpha_1 \leq \alpha_2 \leq .. \leq \alpha_n = b$ be a partition of $[a, b]$. Riemann - integration theory requires to consider

$$L_\tau := \sum_{i=1}^{n} \phi_i(\alpha_i - \alpha_{i-1}) \quad \text{ and } \mathcal{U}_\tau := \sum_{i=1}^{n} \Phi_i(\alpha_i - \alpha_{i-1})$$

where

$$\phi_i = \inf\{f(x), x \in [\alpha_{i-1}, \alpha_i]\} \text{ and}$$
$$\Phi_i := \sup\{f(x), x \in [\alpha_{i-1}, \alpha_i]\}.$$

Now note that for $\mu = \lambda^1$ the functions

$$l_\tau := \sum_{i=1}^{n} l_i 1_{A_i} \text{ and } u_\tau = \sum_{i=1}^{n} \Phi_i 1_{A_i}$$

where $A_i = [\alpha_{i-1}, \alpha_i])$ are μ - integrable with

$$L\tau = \int l_\tau d\mu \quad \text{and} \quad \mathcal{U}_\tau = \int u_\tau d\mu.$$

Since f is Rieman integrable, there is a sequence of nested partitions of $[a, b]$, such that $(L_{\tau_n})_n$ and $(\mathcal{U}_{\tau_n})_n$ have the same limit in $\mathbb{R}$. By definition the corresponding sequences $(\mathcal{U}_{\tau_n})$ and (l_{τ_n}) are decreasing and increasing, respectively and the function

$$q := \lim_{n\to\infty} (u_{\tau_n} - l_{\tau_n})$$

exists. Noting that $u_{\tau_n} \geq l_{\tau_n}$ we may apply Fatou's lemma to obtain

$$0 \leq \int q d\mu \leq \lim_{n\to\infty} (\mathcal{U}_{\tau_n} - L_{\tau_n}) = 0$$

This shows that $q = 0 \quad \lambda - a.e.$ and since moreover $l_{\tau_n} \leq f \leq n_{\tau_n}$ this shows that $\lim_{n\to\infty} l_{\tau_n} = f \quad \lambda^1 - a.e.$
Now Riemann - integrability of f implies that f is bounded and hence Lebesgue - integrable, since $\lambda([a, b])$ is finite. Thus $(|\ell_{\tau_n}|)$ may be dominated by a λ - integrable function. Thus by dominated convergence

$$\int f d\lambda = \int \lim \ell_{\tau_n} d\lambda = \lim \int l_{\tau_n} d\lambda = \lim_{n\to\infty} L_{\tau_n} = \int f dx.$$

■

Remark 10.14 *Recall that e.g. Dirichlet's function $1_{\mathbb{Q}}(x)$ is not Riemann - integrable, but, as was shown in the exercises, well Lebesgue - integrable with integral zero. So the Lebesgue - integral on $[a, b]$ is indeed an extension of the Riemann - integration. The case of unbounded intervals is a quick consequence:*

Corollary 10.15 *Let $f \geq 0, f : \mathbb{R} \to \mathbb{R}$ be B^1 - measurable. Let f be Riemann - integrable on each interval $[a, b], a < b \in \mathbb{R}$. f is Lebesgue - integrable, if and only if the limiting Riemann - integrals*

$$\rho := \lim_{n\to\infty} \int_{-n}^{n} f(x)dx$$

exists. In this case ρ equals the Lebesgue - integral.

Proof. Let ρ_n be the Riemann - integral of f over $A_n = [-n, n]$. Theorem 10.13 tells us that

$$\rho_n = \int_{A_n} f d\lambda^1 = \int_{\mathbb{R}} 1_{A_n} f d\lambda^1.$$

Moreover $f 1_{A_n} \uparrow f$ and hence

$$\sup \rho_n = \int f d\lambda.$$

f is Riemann integrable if and only if $\sup \rho_n < \infty$ and then $\rho = \sup \rho_n$. This proves the assertion. ■

Applying Corollary 10.15 to f^+ and f^- (defined in the usual sense) shows that Riemann - integral and Lebesgue - integral of arbitrary Borel - measurable functions agree if $|f|$ is Riemann - integrable on $\mathbb{R}$. Of course this is also true, if $I \subseteq \overline{\mathbb{R}}$ is an arbitrary interval. On the other hand the existence of the Riemann integral on $\mathbb{R}$ does not imply Lebesgue - integrability in general, this is shown in the following extended exercise:

Exercise 10.16 *Consider the function $f : \mathbb{R}^+ \to \mathbb{R}$ given by $f(x) := \frac{1}{x}\sin(x)$.*
a) Show that f can be extended to a continuous function in 0.
b) Show that f is Riemann - integrable on $\mathbb{R}^+_{\{0\}}$, i.e. shows that

$$\lim_{a\to\infty} \int_0^a \frac{\sin x}{x} dx \text{ exists.}$$

c) Show that f is not Lebesgue - integrable. Hint:

$$\frac{|\sin x|}{x} \geq \frac{1}{(k+1)\pi} |\sin x| .$$

11 Stochastic Convergence

In the previous section we already studied two different convergence criteria: $\mu - a.e.$ convergence and convergence in $\mathcal{L}^p(\mu)$. In this section we introduce a third concept, called stochastic convergence. Stochastic convergence is partially motivated by the Weak Law of Large Numbers, hence by probability theory, but also is interesting in it's own rights. As usual we consider a measure space $(\Omega, \mathcal{A}, \mu)$

Definition 11.1 *A sequence $(f_n)_n$ of measurable, real functions is called stochastically convergent (or convergent in measure) to a measurable function $f : \Omega \to \mathbb{R}$, if for each $\varepsilon > 0$ and every $A \in \mathcal{A}$ with $\mu(A) < \infty$ it holds*

$$\lim_{n\to\infty} \mu(\{|f_n - f| \geq \varepsilon right\} \cap A) = 0. \tag{11.1}$$

In particular, if μ is finite (11.1) is equivalent with

$$\lim_{n\to\infty} \mu\left(\{|f_n - f| \geq \varepsilon\}\right) = 0 \tag{11.2}$$

for all $\varepsilon > 0$.

For non - finite (e.g. σ - finite) measures (11.1) and (11.2) are in general not equivalent. This is illustrated by

Example 11.2 *Let $\Omega = \mathbb{N}$ and $\mathcal{A} = \mathcal{P}(\mathbb{N})$. μ is uniquely defined by $\mu(\{n\}) = \frac{1}{n}$. Obviously μ is σ -finite. Consider $A_n = \{n, n+1, ...\}$ and $f_n = 1_{A_n}$. f_n converges stochastically to zero, since for all $0 < \varepsilon < 1$ the set $\{f_n \geq \varepsilon\}$ equals A_n. But $A_n \downarrow \emptyset$, which implies $\mu(A_n \cap A) = 0$ for all $A \in \mathcal{P}(\mathbb{N})$ of finite measure. On the other hand $\mu(A_n) = \infty$ for all $n \in \mathbb{N}$ since the harmonic series diverges.*

Now we will ask ourselves the question whether stochastic limits are uniquely defined.

Theorem 11.3 *a) Let $(f_n)_n$ be stochastically convergent to f and $f^* = f$ $\mu -$ a.e. Then $(f_n)_n$ also converges stochastically to f^*.*
b) Any two limits of a stochastically convergent sequence are $\mu -$ a.e. equal, if μ is σ - finite.

Proof. a) Is obvious since $\{f_n - f\} \cap A$ and $\{f_n - f^*\} \cap A$ only differ by an n-independent nulset.
b) Let f and f^* be stochastic limits of a sequence (f_n), then the triangle inequality yields for $\varepsilon > 0$:

$$\{|f - f^*| \geq \varepsilon\} \leq \left\{|f_n - f| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|f_n - f^*| \geq \frac{\varepsilon}{2}\right\}.$$

Thus

$$\mu\left(\{|f - f^*| \geq \varepsilon\} \cap A\right) \leq \mu\left(\left\{|f_n - f| \geq \frac{\varepsilon}{2}\right\} \cap A\right) + \mu\left(\left\{|f_n - f^*| \geq \frac{\varepsilon}{2}\right\} \cap A\right)$$

Since the two summands on the right hand side tend to zero as $n \to \infty$, we obtain

$$\mu\left(\{|f - f^*| \geq \varepsilon\} \cap A\right) = 0$$

for all $\varepsilon > 0$ and $A \in \mathcal{A}$ with $\mu(A) < \infty$. But then $f = f^* \quad \mu - a.e.$ on A, since

$$\{f = f^*\} \cap A = \bigcup_{k=1}^{\infty} \left\{|f - f^*| \geq \frac{1}{k}\right\} \cap A$$

has μ - measure zero. Choosing a sequence $A_n \in \mathcal{A}$ with $A_n \uparrow \Omega$ and $\mu(A_n) < \infty$ gives the desired result. ■

Let us see that σ - finiteness in Theorem 11.4 b) is indeed necessary:

Example 11.4 *Let $(\Omega, \mathcal{P}(\Omega), \mu)$, where $\Omega = \{w_0, w_1\}$ and $\mu(\{w_0\}) = 0$ and $\mu(\{w_1\}) = +\infty$. Then the constant sequence $f_n = 0$ for all n converges to any arbitrary function $f : \Omega \to \mathbb{R}$ stochastically.*

The following inequality is a key tool in probability theory and establishes the link between stochastic convergence and convergence in $\mathcal{L}^p$.

Lemma 11.5 *Let $f : \Omega \to \mathbb{R}$ be measurable. Moreover let $g : \mathbb{R} \to \mathbb{R}^+$ be strictly increasing. Then for all $\varepsilon > 0$ the following inequality (Chebyskev - Markov inequality) holds:*

$$\mu(f \geq \varepsilon) \leq \frac{1}{g(\varepsilon)} \int g(f) d\mu. \tag{11.3}$$

Here we assume that $\int g(f)d\mu$ exists.

Proof. Define $A_\varepsilon := \{f \geq \varepsilon\} \in \mathcal{A}$. Then

$$\int g(f) d\mu \geq \int_{A_\varepsilon} g(f) d\mu \geq \int_{A_\varepsilon} g(\varepsilon) = g(\varepsilon)\mu(A_\varepsilon).$$

■

An immediate consequence is

Theorem 11.6 *Let $(f_n)_n$ be a sequence in $\mathcal{L}^p(\mu)$. If (f_n) converges to $f \in \mathcal{L}^p(\mu)$ in $\mathcal{L}^p$ then it also converges μ - stochastically.*

Proof. Applying the Chebyshev - Markov inequality to $|f_n - f|$ and with $g(x) = x^p$ (which is increasing and positive on $\mathbb{R}^+$) we get

$$\mu\left(\{|f_n - f| \geq \varepsilon\} \cap A\right) \leq \mu\left(\{|f_n - f| \geq \varepsilon\}\right) \leq \varepsilon^{-p} \int |f_n - f|^p \, d\mu$$

The last expression converges to zero by the $\mathcal{L}^p(\mu)$ - convergence of f_n to f. ■

Theorem 11.6 shows that stochastic convergence is weaker than $\mathcal{L}^p$ - convergence. The next theorem reveals that it also is weaker than almost - everywhere convergence.

Theorem 11.7 *Let (f_n) be a sequence of functions on Ω, that are measurable and converge $\mu - a.e.$ to a measurable, real function f on Ω. Then (f_n) also converges μ - stochastically to f.*

Proof. Since

$$\{|f_n - f| \geq \varepsilon\} \subseteq \left\{\sup_{m \geq n} |f_m - f| \geq \varepsilon\right\}$$

and therefore

$$\mu\left(\{|f_n - f| \geq \varepsilon\} \cap A\right) \leq \mu\left(\left\{\sup_{m \geq n} |f_m - f| \geq \varepsilon\right\} \cap A\right)$$

for all $\varepsilon > 0$ and $A \in \mathcal{A}$. The assertion therefore is a consequence of the following lemma, since for A with $\mu(A) < \infty, \mu\,|_{A \subset \mathcal{A}}$has finite mass. ■

Lemma 11.8 *Let μ be finite. The $(f_n)_n$ (a sequence of measurable, real functions on Ω) converges $\mu - a.e.$ to zero, if one of the following (equivalent) conditions is satisfied:*

$$\lim_{n \to \infty} \mu(\{\sup_{m \geq n} |f_m| \geq \varepsilon\}) = 0 \quad \textit{for all } \varepsilon > 0 \tag{11.4}$$

$$\lim_{n \to \infty} \mu(\{\sup_{m \geq n} |f_m| > \varepsilon\}) = 0 \quad \textit{for all } \varepsilon > 0 \tag{11.5}$$

$$\mu(\{\limsup_{n \to \infty} \{|f_n| > \varepsilon\}) = 0 \quad \textit{for all } \varepsilon > 0. \tag{11.6}$$

Proof. Let us first see that (11.4) is equivalent with the $\mu-a.e.$ convergence. For $\varepsilon > 0$ and $n \in \mathbb{N}$ put

$$A_n^\varepsilon := \left\{ \sup_{m \geq n} |f_m| \geq \varepsilon \right\}.$$

Obviously $n \mapsto A_n^\varepsilon$ and $\varepsilon \mapsto A_n^\varepsilon$ are decreasing. Thus $k \mapsto A_n^{\frac{1}{k}}$ is increasing. If we define eventually

$$A := \{\omega : \lim_{n\to\infty} f_n(\omega) = 0\} = \{\omega : \lim_{n\to\infty} \sup f_n(\omega) = 0$$

then since $\limsup f_n$ is measurable $A \in \mathcal{A}$. Obviously

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \left(A_n^{\frac{1}{k}} \right)^c = \bigcap_{\alpha > 0} \bigcup_{n=1}^{\infty} (A_n^\alpha)^c$$

and thus

$$A^c = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}} = \bigcup_{\alpha>0} \bigcap_{n=1}^{\infty} A_n^\alpha.$$

Therefore

$$\bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}} \to A^c \text{ and } A_n^{\frac{1}{k}} \downarrow \bigcap_{m=1}^{\infty} A_m^{\frac{1}{k}}.$$

Consequently

$$\mu\left(A^c\right) = \sup_k \mu \left(\bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}} \right) = \sup_k \inf \mu(A_n^{\frac{1}{k}}) \tag{11.7}$$

since μ as a finite measure is continuous from above and below. Thus f_n converges to zero $\mu - a.e.$ if and only if the number in (11.7) is zero. But this is the case, if and only if

$$\inf_n \mu \left(A_n^{\frac{1}{k}} \right) = \lim_{n\to\infty} \mu \left(A_n^{\frac{1}{k}} \right) = 0 \quad \text{ for all } k \in \mathbb{N}.$$

This shows the equivalence of (11.4) with $\mu - a.e.$ convergence.

(11.4) and (11.5) are obviously equivalent. The equivalence of (11.5) with (11.6) follows from

$$\lim_{n\to\infty} \mu \left(\left\{ \sup_{m\geq n} |f_m| > \varepsilon \right\} \right) = \mu\left(\limsup\{|f_n| > \varepsilon\right) \quad \text{ for all } \varepsilon > 0. \tag{11.8}$$

To see (11.8) put

$$B_n := \bigcup_{m=n}^{\infty} \{|f_m| > \varepsilon\} \text{ and } B := \{\limsup_{n\to\infty} \{|f_n| > \varepsilon.$$

Then on the one hand $B_n \downarrow B$ and thus $\lim \mu(B_n) = \mu(B)$, on the other hand

$$B_n = \bigcup_{m=n}^{\infty} \{|f_m| > \varepsilon\} = \{\sup_{m\geq n} |f_m| > \varepsilon.$$

This implies (11.8) and therefore shows the lemma. ■

The following two examples show that there are indeed situations, where (f_n) converges stochastically but not almost everywhere or in $\mathcal{L}^p$.

Example 11.9 *Consider the situation of Example 10.7. There we constructed an example where (f_n) converges to 0 in $\mathcal{L}^p$, for all $1 \leq p < \infty$ but not $\mu - \alpha.e.$ As a consequence of Theorem 11.7 then(f_n) also converges also μ -stochastically but still not $\mu - a.e.$*

Example 11.10 *Consider the situation of Example 10.8. There we constructed a situation, where (f_n) converges to 0 $\mu - a.e.$ and hence μ - stochastically (as a consequence of Theorem 11.8) but not in $\mathcal{L}^p(\mu)$, $1 \leq p < \infty$.*

To motivate the following theorem one should have a closer look at Example 10.10. There we have the situation of intervals of width 2^{-h_n},$h_n \to \infty$, wandering around in the unit interval $[0, 1]$. Since there width 2^{-h_n} converges to zero, the sequence of indicators on these intervals goes λ^1- stochastically to zero. On the other hand, because of the intervals wandering around, the convergence is not $\lambda^1 - a.e.$ But, of course, if we concentrate on the subsequence of indicators on intervals containing zero, this subsequence does converge $\lambda^1 - a.e.$ The following theorem states that there is a general principle behind this observation:

Theorem 11.11 *For any sequence (f_n) of measurable real functions that converges μ - stochastically to a measurable, real limit f and any $A \in \mathcal{A}$ with $\mu(\mathcal{A}) < \infty$ there is a subsequence of (f_n) that converges $\mu - a.e.$ on A.*

Proof. Without loss of generality $\mu(\Omega) < \infty$ and $A = \Omega$. By the triangle inequality

$$\{|f_m - f_n| \geq \varepsilon\} \leq \left\{|f_n - f| \geq \frac{\varepsilon}{2}\right\} + \left\{|f_m - f| \geq \frac{\varepsilon}{2}\right\}$$

and since the right hand side goes to zero the left hand side can be made arbitrary small, e.g. smaller than ε. Thus, if (η_k) is a sequence of positive numbers with

$$\sum_{n=1}^{\infty} \eta_k < +\infty$$

then for each $k \in \mathbb{N}$ there is n_k with

$$\mu\left(\left\{\left|f_m - f_{n_k}\right| \geq \eta_k\right\}\right) \leq \eta_k$$

for all $m \geq n_k$ and all $k \in \mathbb{N}$. Putting

$$A_k := \left\{\left|f_{n_{k+1}} - f_{n_k}\right| \geq \eta_k\right\}$$

we arrive at

$$\sum_{k=1}^{\infty} \mu\left(A_k\right) \leq \sum_{k=1}^{\infty} \eta_k < \infty$$

which implies

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} \mu\left(A_k\right) = 0.$$

For $A := \limsup A_n$ this means $\mu(A) = 0$, since $A \subset \bigcup_{k=n}^{\infty} A_k$ and therefore $\mu(A) \leq \sum_{k=n}^{\infty} \mu(A_k)$. But on $\Omega \setminus A$ (which is a set of full measure) $\left|f_{n_{k+1}} - f_{n_k}\right| \geq \eta_k$ can only happen finitely often. Now, since $\sum \eta_k < \infty$, the series

$$\sum_{k=1}^{\infty} \left|f_{n_{k+1}}(\omega) - f_{n_k}(\omega)\right|$$

converges absolutely, but this means that (f_{n_k}) converges to a measurable function $f^* : \Omega \to \mathbb{R}$. This converge is on $\frac{\Omega}{A}$ hence $\mu - a.e.$ But then (f_{n_k}) also converges stochastically to f^*. Hence $f^* = f$. ■

The following exercise shows that stochastic convergence may even be characterized by a subsequence principle:

Exercise 11.12 *Let (f_n) be a sequence of measurable real functions. Show that (f_n) converges to a measurable, real limit f stochastically, if for all $A \in \mathcal{A}$ with $\mu(A) < \infty$ and all subsequences (f_{n_k}) of (f_n) there is a subsubsequence (f_{n_k}) that converges $\mu - a.e.$ on A to f.*

Exercise 11.13 *Let* (f_n) *and* (g_n) *be stochastically convergent to* f *and* g*, respectively. Show that for* $\alpha, \beta \in \mathbb{R}$ *the sequence* $\alpha f_n + \beta g_n$ *converges stochastically to* $\alpha f + \beta g$*. What about* $\max(f_n, g_n)$ *and* $\min(f_n, g_n)$*?*

Exercise 11.14 *Is the following relaxation of Exercise 11.12 also true:* (f_n) *converges* μ*-stochastically to a limit* f*, if it contains a* $\mu-$*a.e.to* f *convergent subsequence?*

The following is a very useful consequence of Exercise 11.12:

Exercise 11.15 *Let* $(f_n)_n$ *be a sequence of real, measurable functions that converges stochastically to a real, measurable limit* f*. Let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be continuous. Then* $(\varphi \circ f_n)_n$ *converges stochastically to* $\varphi \circ f$*.*

12 The Radon-Nikodym Theorem

Already in a first course in probability one encounters measures of a very special form, absolutely continuous probability distributions. For such a probability μ it holds that there is a function $h : \mathbb{R} \to \mathbb{R}^+$ (the density) that is λ^1 - integrable (and such that $\int_{\mathbb{R}} h(x) d\lambda(x) = 1$) such that for each $A \in \mathcal{B}^1$

$$\mu(A) = \int_A h(x) dx.$$

The advantage of studying such measures is directly visible: they are comparable with Lebesgue - measure. In particular they are continuous (therefore their name) with respect to Lebesgue measure, i.e. for $\varepsilon > 0$ there exists $\delta > 0$, such that if

$$\lambda(A) \leq \delta \quad \text{then} \quad \mu(A) \leq \varepsilon. \tag{12.1}$$

The most prominent example of an absolutely continuous measure is the normal distribution in one dimension with density

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

In this section we will ask ourselves the question, when for two measures μ and ν (on the same measurable space $(\Omega, \mathcal{A})$) there is a $\mathcal{A}$-measurable function h, such that for all $A \in \mathcal{A}$

$$\mu(A) = \int_A h(x) d\nu(x). \tag{12.2}$$

It turns out that the answer to this question is intrinsically related to a continuity property as in (12.1).

Definition 12.1 *Let $h \geq 0$ be a measurable function $h : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}^+$. The measure μ defined by (12.2) is called the measure with density h with respect to ν and is also abbreviated by*

$$\mu = h\nu. \tag{12.3}$$

Exercise 12.2 *Show that $\mu = h\nu$ in the situation of Definition 12.1 indeed defines a measure.*

Before discussing Definition 12.1 we will also baptist property (12.1).

Definition 12.3 *A measure μ on $\mathcal{A}$ is called continuous with respect to a measure ν on $\mathcal{A}$ if every $N \in A$ with $\nu(N) = 0$ also has μ - measure zero: $\mu(N) = 0$.*

Let us now study consequences of these two definitions.

Theorem 12.4 *Let $\nu = f\mu$ with $f \in E^*$. Then for all $\varphi \in E^*$ it holds*

$$\int \varphi d\nu = \int \varphi f d\mu \tag{12.4}$$

Moreover a function $\varphi : \Omega \to \overline{\mathbb{R}}$ is ν - integrable if and only if $\varphi \cdot f \cdot$ is μ - integrable and in this case (12.4) holds as well.

Proof. We first check the assertion for step functions

$$\varphi = \sum_{i=1}^{n} \alpha_i 1_{A_i}, A_i \in \mathcal{A}.$$

Then

$$\int \varphi d\nu = \sum_{i=1}^{n} \alpha_i \nu(A_i) = \sum_{i=1}^{n} \alpha_i \int 1_{A_i} f d\mu = \int \varphi d\mu. \tag{12.5}$$

An arbitrary $\phi \in E^*$ is approximated by a sequence (u_n) in E with $u_n \uparrow \varphi$. Then also $u_n f \uparrow \varphi f$ which together with (12.5) implies (12.4) by monotone convergence. For an arbitrary integrable φ (12.4) follows by decomposing φ into φ^+ and φ^- and applying additivity of the integral. ■

Exercise 12.5 *Show that, if* $\nu = f\mu$ *and* $\varphi = g\nu$ *with* $f, g \in E^*$*, then* $\varphi = (gf)\mu = g(f\mu)$.

Next we discuss the question whether densities are unique:

Theorem 12.6 *For* $f, g \in E^*$ *it holds:*

$$f = g\,\mu - a.e. \Longrightarrow f\mu = g\mu \tag{12.6}$$

If f *or* g *are* μ *- integrable the converse of (12.6) also holds true.*

Proof. $f = g\,\mu - a.e.$ implies $f1_A = g1_A\mu - a.e.$ for all $A \in \mathcal{A}$. Thus

$$\int_A f d\mu = \int_A g d\mu \quad \text{for all } A \in \mathcal{A}$$

i.e. $f\mu = g\mu$. Now assume that f is μ - integrable and $f\mu = g\mu$. Then also g is μ - integrable. Consider $N := \{f > g\} \in \mathcal{A}$, and

$$h := 1_N f - 1_N g.$$

Since $1_N f \leq f, 1_N g \leq g$, the functions $1_N f$ and $1_N g$ are μ - integrable and because of $f\mu = g\mu$ they have the same μ integral. Thus

$$\int h d\mu = \int_N f d\mu - \int_N g d\mu = 0$$

This shows $\mu(N) = 0$. Interchanging f and g gives $\mu\,(f \neq g) = 0$. ■
Let us now study the second property. Continuity of two measure

Theorem 12.7 *Let* μ, ν *be two measures on* $(\Omega, \mathcal{A})$*, and assume that* ν *is finite.* ν *is* μ *- continuous if and only if for each* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that*

$$\mu\,(A) \leq \delta \Longrightarrow \nu\,(A) \leq \varepsilon \quad \textit{for all } A \in \mathcal{A}. \tag{12.7}$$

Proof. One direction is easy: If (12.7) holds, $\nu\,(A) \leq \varepsilon$ for all A with $\mu(A) = 0$ and all $\varepsilon > 0$. Thus $\nu\,(A) = 0$ for all A with $\mu\,(A) = 0$, i.e. ν is μ - continuous.

On the other hand assume (12.7) was wrong. Hence there is $\varepsilon > 0$ and a sequence $(A_n)_n$ in $\mathcal{A}$ with

$$\mu\,(A_n) \leq 2^{-n} \text{ and } \nu\,(A_n) > \varepsilon, \quad n \in \mathbb{N}.$$

Define

$$A := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \in \mathcal{A},$$

then on the one hand

$$\mu(A) \leq \mu\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mu(A_m) = 2^{-n+1}, \quad n \in \mathbb{N}$$

hence $\mu(A) = 0$, and on the other hand

$$\nu(A) \geq \limsup \nu(A_n) \geq \varepsilon > 0.$$

Here for the first inequality we used that ν is finite. Thus ν is not μ - continuous. ■

We now turn to the central question of this section: What is the relation of Definition 12.1 and Definition 12.3? To prepare for the answer of this question we first prove an important theorem from Hilbert space theory (Recall that a Hilbert space is a normed, complete, linear space with a norm coming from an inner product. The prototype of a Hilbert space, is the space $L^2(\mu)$ with inner product $< f, g >= \int fg d\mu$.).

Theorem 12.8 *(Riesz representation theorem): Let $\lambda : H \to \mathbb{R}$ be a linear, continuous functional on a Hilbert space H. Then there exists a $\lambda \in H$ with $\lambda(x) =< x, \lambda >$ for all $x \in H$.*

Proof. Let $H_\lambda := \{x \in H : \lambda(x) = 0\}$ (if $\lambda \not\equiv 0$, in which case the assertion is trivial). H_λ is closed, since λ is continuous. Let $a \in H \backslash H_1$ and $a_0 \in H_\lambda$ be its orthogonal projection onto H_λ. Obviously $0 \neq a - a_0 \in H_\lambda^+$ (the orthogonal complement of H_λ). Put $a_1 := \frac{a-a_o}{\|a-a_0\|} \in H\frac{1}{\lambda}$. Then

$$\lambda(a_1) = \frac{1}{\|a - a_0\|}\lambda(a) \neq 0$$

and thus $x - \frac{\lambda(x)}{\lambda(a_1)}a_1 \in H_\lambda$ is well - defined and

$$< x - \frac{\lambda(x)}{\lambda(a_1)}a_1, a_1 >= 0.$$

Solving this for $\lambda(x)$ gives

$$\lambda(x_1) = \lambda(a_1) < x, a_1 > .$$

Defining $a_\lambda := \lambda(a_1)\, a_1$ yields $\lambda(x) =< x, a_\lambda >$. ■
The following theorem is absolutely central in the entire field of measure theory and probability theory. It has also coined the name Radon - Nihodym density for the density of a μ - continuous measure ν. One also writes $\nu \ll \mu$, if ν is μ - continuous and $\frac{d\nu}{d\mu}$ for the density of ν with respect to μ.

Theorem 12.9 *(Radon - Nikodym): Let μ, ν be measures on $(\Omega, \mathcal{A})$. If μ is σ - finite, the following are equivalent:*
(i) $\quad$ *ν has a density with respect to μ*
(ii) $\quad$ *$\nu \ll \mu$.*

Proof. : $\quad$ $(i) \Longrightarrow (ii)$ has already been proven.
$(ii) \Longrightarrow (i)$: We start with the case that μ, ν are finite. Put $\lambda := \mu + \nu$. Since λ is finite $L^2(\lambda) \subseteq L^2(\nu) \subseteq L^1(\nu)$. For $f \in L^2(\lambda)$ put

$$\Lambda(f) := \int f d\nu$$

Then

$$|\Lambda(f)| \leq \nu(\Omega)^{\frac{1}{2}} \left(\int \left|f^2\right| d\nu \right)^{\frac{1}{2}} \leq \nu(\Omega)^{\frac{1}{2}} \left(\int \left|f^2\right| d\lambda \right)^{\frac{1}{2}} = \nu(\Omega)^{\frac{1}{2}} \, ||f||_2.$$

Therefore $\Lambda : L^2(\lambda) \to \mathbb{R}$ is a linear, bounded and hence continuous function. Following Theorem 12.8 we see that there is a $f_0 \in L^2(\lambda)$ with

$$\Lambda(f) = \int f d\nu = \int f \cdot f_0 d\lambda =< f, f_0 > \tag{12.8}$$

for all $f \in L^2(\lambda)$. In particular, for $f = 1_E, E \in \mathcal{A}$:

$$\nu(E) = \int_E f_0 d\lambda \geq 0.$$

Thus $f_0 \geq 0, \nu - a.e.$ On the other hand:

$$\int_E (1 - f_0) d\lambda = \lambda(E) - \nu(E) = \mu(E) \geq 0,$$

for all $E \in \mathcal{A}$, hence $f_0 \leq 1$. We choose such a $0 \leq f_0 \leq 1$ with (12.8). Define $\Omega_1 = \{f_0 = 1\}$,$\Omega_2 := \{0 < f_0 < 1\}$, and $\Omega_3 := \{f_0 = 0\}$. Then for all $E \in \mathcal{A}$

$$\int_E (1 - f_0)\, d\nu = \nu(E) - \int_E f_0 d\nu = \int_E f_0 d\lambda - \int_E f_0 d\nu = \int_E f_0 d\mu.$$

For $E = \Omega_1$ this implies $\mu(\Omega_1) = 0$ and hence (since $\nu \ll \mu$) that $\nu(\Omega_1) = 0$. Using the usual approximation techniques we obtain for all $f : \Omega_2 \to \mathbb{R}^+$

$$\int_{\Omega_2} f\,(1 - f_0)\, d\nu = \int_{\Omega_2} f f_0 d\mu.$$

Applying this to $f = \frac{1_E}{(1-f_0)}, E \subseteq \Omega_2, E \in \mathcal{A}$ we get

$$\nu(E) = \int_E \frac{f_0}{1 - f_0} d\mu.$$

Taking into account that $\nu(\Omega_3) = 0$ we obtain

$$\nu(E) = \nu(E \cap \Omega_2) = \int_{E \cap \Omega_2} \frac{f_0}{1 - f_0} d\mu.$$

This shows that by defining

$$\frac{d\nu}{d\mu}(\omega) := f(\nu) := \left\{ \begin{array}{ll} \frac{f_0}{1-f_0} & \omega \in \Omega_2 \\ 0 \text{ otherwise} & \end{array} \right\}$$

f is a density for ν with respect to μ.

In a second step we assume that $\mu(\Omega) < \nu(\Omega) = \infty$. We will present a partition of Ω into pairwise disjoint sets $\Omega_0, \Omega_1, ... \in \mathcal{A}$,with $\Omega = \bigcup_{i=0}^{\infty} \Omega_i$ and
a) For $A \in \Omega_0 \cap \mathcal{A}$ either $\mu(A) = \nu(A) = 0$ or $\mu(A) > 0$ and $\nu(A) = +\infty$ holds
b) $\nu(\Omega_n) < +\infty$,for all $n \in \mathbb{N}$ To see this let

$$\begin{aligned} \mathcal{Q} &: = \{Q \in \mathcal{A} : \nu(Q) < +\infty\} \text{ and} \\ \alpha &: = \sup_{Q \in \mathcal{Q}} \mu(Q). \end{aligned}$$

By definition there is a sequence (Q_m) in $\mathcal{Q}$ with $\alpha = \lim \mu(Q_m)$. This sequence may be chosen to be increasing. Then $Q_0 = \bigcup_{n=1}^{\infty} Q_n \in \mathcal{A}$ has μ -

measure $\mu(Q_0) = \alpha$. Our candidate for Ω_0 is Q_0^c. Indeed, take $A \in Q_0^c \cap \mathcal{A}$ with $\nu(A) < +\infty$. Then $Q_m \cup A \in \mathcal{Q}$ for all m, hence $\mu(Q_m \cup A) \leq \alpha$ for all and thus

$$\mu(Q_0 \cup A) = \lim \mu(Q_m \cup A) \leq \alpha.$$

Now $A \cap Q_0 = \emptyset$ and therefore $\mu(Q_0 \cup A) = \mu(Q_0) + \mu(A) = \alpha + \mu(A) \leq \alpha$. This implies $\mu(A) = 0$. Since eventually $\nu \ll \mu$ we also have $\nu(A) = 0$. This show that a) is fulfilled for $\Omega_0 = Q_0^c$. b) is satisfied for $\Omega_m := Q_m \backslash Q_{m-1}$ and

$$\Omega_1 = Q_1.$$

Defining $\mu_n := \mu \mid \Omega_n \cap \mathcal{A}$ and $\nu_n \mid \Omega_n \cap \mathcal{A}$ we know $\nu_n \ll \mu_n$, for $n = 0, 1, 2....$ For $n \geq 1$ the measure ν_n and μ_n are finite. By the previous considerations we know that there is a measurable function $f_n \geq 0$ on Ω_n with $\nu_n = f_n \mu_n$. On Ω_o we may put $f_0 \equiv +\infty$ to obtain $\nu_0 = f_0 \mu_o$ (due to proberty). Concetenating the f_i's to

$$f(\omega) = f_i(\omega) 1_{\Omega_i}(\varpi)$$

we see that $\nu = f\mu$.

In the final last step we also assume that μ is σ - finite. Then there exits a μ - integrable function h on Ω with $0 < h(\omega) < +\infty$ for all $\omega \in \Omega$. Indeed, there is a sequence $(A_n)_n$ in $\mathcal{A}$ with $A_n \uparrow \Omega$ and $\mu(A_n) < +\infty$. Then we can find $0 < \eta_n \leq 2^{-n}$ with $\eta_n \mu(A_n) \leq 2^{-n}$.Therefore

$$h = \sum_{n=1}^{\infty} \eta_n 1_{A_n}$$

does the job.But the measure $h\mu$ is finite and has the same nulsets as μ. Hence $\nu \ll h\mu$. By what we have seen above there is a function $f \geq 0$ such that $\nu = f(h\mu) = (fh)\mu$. Hence (fh) is a density for ν with respect to μ. This proves the theorem. ■

Exercise 12.10 *Show that the Dirac measure δ_x on $B^1(\delta_x(a) = 1$ if $x \in A$ and 0 otherwise) does not have a density with respect to λ^1.*

13 Uniform integrability

A very useful application of stochastic convergence is an extension of the dominated convergence theorem:

Theorem 10.5 states that, if f_n converges $\mu - a.e.$ to a limit f and this is bounded by a p - integrable function then the convergence is also in $\mathcal{L}^p$. Examples 10.8 and 11.10 show that this condition is sufficient but not at all necessary: There do exist examples of sequences that converge in $\mathcal{L}^p$, but not almost everywhere. The following definition helps to turn Theorem 10.5 into an equivalence statement.

Definition 13.1 *A family $(f_i)_{i\in I}$ of measurable numeric function is called uniformly integrable, if for each $\varepsilon > 0$ there is an integrable $g : \Omega \to \mathbb{R}^+$ with*

$$\int_{\{|f_i|\geq g\}} |f_i| \, d\mu \leq \varepsilon \tag{13.1}$$

for all f_i.

Example 13.2 *Every finite set $\{f_1, ..., f_n\}$ of μ - integrable functions is uniformly integrable. For g we may take $2 \max \{|f_1|, ..., |f_n|\}$.*

Example 13.3 *Let $\Omega = \mathbb{N}, \mathcal{A} = \mathcal{P}(\mathbb{N})$ and μ be defined by $\mu(\{n\}) = 2^{-n}$. Then μ is finite. hence all constants are integrable. Define $f_n = 1_{\{n\}} 2^n n^{-1}$. Then (f_n) are uniformly integrable. Indeed for the constant, integrable function $g \equiv \alpha$ it holds*

$$\int_{\{f_n\geq\alpha\}} f_n d\mu \leq \frac{1}{n}.$$

This shows uniform integrability since for $\varepsilon > 0$ we may choose $g \equiv \alpha := 2^{\frac{1}{\varepsilon}}$. Note, however that the smallest function g majorizing all the f_n is $g(n) = \frac{2^n}{n}$, which is not integrable.

The following characterization of uniform integrability turns out to very useful:

Theorem 13.4 *A family of functions $(f_i)_{i\in I}$ is uniformly μ - integrable, if and only if it satisfies the following two conditions:*

$$sup_i \int |f_i| \, d\mu < \infty \tag{13.2}$$

For every $\varepsilon > 0$, there is a μ - integrable $h \geq 0$ and a $\delta > 0$ such that for all $A \in \mathcal{A}$

$$\int_A h d\mu \leq \delta \Rightarrow \int_A |f_i| \, d\mu \leq \varepsilon, \quad i \in I \tag{13.3}$$

Proof. For all $A \in \mathcal{A}, f : \Omega \to \bar{\mathbb{R}}$ measurable and $g : \Omega \to \mathbb{R}^+$ integrable it holds

$$\int_A |f|\, d\mu = \int_{A\cap\{|f|\geq g\}} |f|\, d\mu + \int_{A\cap\{|f|<g\}} |f|\, d\mu \leq \int_{\{|f|\geq g\}} |f|\, d\mu + \int_A g d\mu.$$

For $A = \Omega$ we obtain:

$$\int |f|\, d\mu \leq \int_{\{|f|\geq g\}} |f|\, d\mu + \int g d\mu.$$

Now assume $(f_i)_{i\in I}$ is uniformly integrable and choose g as a $\frac{\varepsilon}{2}$ - bound in (13.1). Choosing $h := g$ and $\delta := \frac{\varepsilon}{2}$ we obtain (13.2) and (13.3). If, on the other hand, (13.2) and (13.3) are satisfied then choose h and $\delta \geq 0$ as in (11.11). Observe that

$$\int |f_i|\, d\mu \geq \int_{\{|f|\geq \alpha h\}} \alpha h d\mu$$

which gives

$$\int_{\{|f|\geq \alpha h\}} h d\mu \leq \frac{1}{\alpha} \int |f_i|\, d\mu.$$

Since this is true for all $i \in I$ and $\sup_i \int |f_i|\, d\mu < \infty$ we may choose α so large that

$$\int_{\{|f_i|\geq \alpha h\}} h d\mu \leq \delta \Longrightarrow \int_{\{|f_i|\geq \alpha h\}} |f_i|\, d\mu \leq \varepsilon:$$

This shows that αh is an ε - bound in (13.1). ■

We are now ready to prove a first generalization of Theorem 10.5:

Theorem 13.5 *Let $(f_n)_n$ be a sequence in $\mathcal{L}^p(\mu)$. Then the following are equivalent:*

(i) $\quad$ *(f_n) converges in $\mathcal{L}^p(\mu)$*

ii) $\quad$ *(f_n) converges stochastically and $(|f_n|^p)$ are uniformly μ - integrable.*

Proof. $(i) \Longrightarrow (ii)$. Since converge in $\mathcal{L}^p(\mu)$ implies stochastic convergence, we just need to show the uniform integrability part. Since $\int |f_n|^p\, d\mu \to \int |f|^p\, d\mu$ condition (13.2) is satisfied. Now for all $A \in \mathcal{A}$

$$\begin{aligned} \left(\int_A |f_n|^p\, d\mu\right)\frac{1}{p} &\leq \left(\int_A |f_n - f|^p\, d\mu\right)\frac{1}{p} + \left(\int_A |f|^p\, d\mu\right) \\ &\leq Np\,(f_n - f) + \left(\int_A |f|^p\, d\mu\right)^{\frac{1}{p}} \end{aligned}$$

For $\varepsilon > 0$ there is n_0, such that for all $n \geq n_0$

$$Np(f_n - f) < \frac{1}{2}\varepsilon\frac{1}{p}$$

(since $f_n \to f$ in $\mathcal{L}^p$). Thus choosing $\delta := 2^{-p}\varepsilon$ and $h = \max(|f_1|^p, ..., |f_{n_o}|^p, |f|^p$ we obtain also (13.3). Hence Theorem 13.4 proves uniform integrability.
$(ii) \Longrightarrow (i)$: This is a bit more involved. Since we already know that $\mathcal{L}^p$ is complete we just need to show that $(f_n)_n$ is a $\mathcal{L}^p$ - Cauchy - sequence, i.e. that for $f_{mn} := f_m - f_n$

$$\lim_{m,n\to\infty} \int |f_{mn}| p d\mu = 0.$$

Exercise 11.22 below shows that $(|f_{mn}|^p)$ is uniformly integrable. Choosing go as an ε - bound for this sequence in (13.1), we see that $g : g \circ \frac{1}{p}$ is p - times integrable and that

$$\int_{\{|f_{mn}|\geq g\}} |f_{mn}|^p \, d\mu \leq \varepsilon \qquad m, n \in \mathbb{N}. \tag{13.4}$$

Splitting the integral

$$\int |f_{mn}|^p \, d\mu = \int_{\{|f_{mn}|\geq g\}} |f_{mn}|^p \, d\mu + \int_{\{|f_{mn}|<g\}} |f_{mn}|^p \, d\mu$$

we see that we only need to make the second summand small. Now $g^p\mu$ is a finite measure on $\mathcal{A}$ and hence continuous from above. As

$$\bigcap_{n>0} \{g < \eta\} = \{g = 0\}$$

we can therefore find $\eta > 0$ such that

$$\int_{\{g<\eta\}} g^p d\mu \leq \varepsilon.$$

But then also

$$\int_{\{|f_{mn}|<g\}\cap\{g<\eta\}} |f_{mn}|^p \, d\mu \leq \int_{\{g<\eta\}} g d\mu \leq \varepsilon. \tag{13.5}$$

Applying the Chebyshev-Markov inequality with $x \mapsto x^p$, we see that

$$\mu(\{g \geq \eta\}) \leq \frac{\int g^p d\mu}{\eta^p} < \infty.$$

Now (f_m) is a stochastic Cauchy sequence, so

$$\lim_{m,n\to\infty} \mu(\{|f_m - f_n| \geq\} \cap A) = 0$$

for all $A \in \mathcal{A}$ with $\mu(A) < \infty$ and all $\alpha > 0$. Combining this with the last observation we obtain for

$$A_{mn} := \{|f_{mn}| \geq \alpha\} \cap \{g \geq \eta\}$$

that

$$\lim_{m,n\to\infty} (A_{mn}) = 0.$$

Now we choose $\alpha > 0$ so small that

$$\left(\frac{\alpha}{\eta}\right)^p \int g^p d\mu \leq \varepsilon$$

$g^p \mu$ is obviously μ - continuous. By Theorem 12.7 we can then find no such that

$$\int_{A_{mn}} g^p d\mu \leq \varepsilon$$

for all $m, n \geq n_0$. But then also

$$\int_{\{|f_{mn}|<g\}\cap A_{mn}} |f_{mn}|^p \, d\mu \leq \int_{A_{mn}} g^p d\mu \leq \varepsilon. \tag{13.6}$$

Eventually another application of the Chebysher - Markov - inequality gives for

$$A_{mn} := \{|f_{mn}| < \alpha\} \cap \{g \geq \eta\},$$

that

$$\int_{\{|f_{mn}|<g\}\cap A'_{mn}} |f_{mn}|^p \, d\mu \leq \alpha^p \mu(\{g \geq \eta\}) \leq \left(\frac{\alpha}{\eta}\right) \int g^p d\mu \leq \varepsilon. \tag{13.7}$$

Combining (13.4)-(13.7) we see that

$$\int |f_{mn}|^p \, d\mu \leq 4?\varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this proves the theorem. ■
To conclude this section with one more equivalence relation between $\mathcal{L}^p$ - convergence and other conditions we need to prove one more lemma:

Lemma 13.6 *Let $(f_n), f_n \geq 0$ be sequence in $\mathcal{L}^1(\mu)$. Assume f_n converges μ - stochastically to $f \geq 0$, $f \in \mathcal{L}^1(\mu)$. If moreover*

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$$

then f_n converges to f in $\mathcal{L}^1(\mu)$.

Proof. : Consider the sequence $(f \wedge f_n)_n$ (where $f_1 g = \min(f, g)$). Since $0 \leq f \wedge f_n \leq f$ the sequence $(f_1 f_n)$ is uniformly integrable (it suffies to find an ε - bound in (13.1) for f). On the other hand

$$0 \leq f - (f \wedge f_n) \leq |f_n - f|$$

which implies that $(f \wedge f_n) \to f$. According to Theorem 13.5 $(f \wedge f_n)$ then converges to f in $\mathcal{L}^1(\mu)$, i.e.

$$\lim_{n\to\infty} \int f \wedge f_n d\mu = \int f d\mu. \tag{13.8}$$

Note that $f + f_n = f \vee f_n + f \wedge f_n$ (where $f \vee g = \max(f, g)$). Thus (13.8) implies

$$\lim_{n\to\infty} \int f \vee f_n d\mu = \int f d\mu. \tag{13.9}$$

(13.8) and (13.9) together imply the assertion since

$$|f_n - f| = f \vee f_n - f \wedge f_n$$

■
Now we can present a final improvement of Theorem 10.5 and theorem 13.5:

Theorem 13.7 *For each sequence (f_n) in $\mathcal{L}^p(\mu)$ that converges μ - stochastically to a function $f \in \mathcal{L}^p(\mu)$ the following are equivalent:*
(i) *(f_n) converges to f in $\mathcal{L}^p(\mu)$*
(ii) *$(|f_n|^p)$ is uniformly integrable.*
(iii) *We have*

$$\lim_{n\to\infty} \int |f_n|^p d\mu = \int |f|^p d\mu.$$

Proof. Theorem 13.5 tells that (i) and (ii) are equivalent.
$(i) \Rightarrow (iii)$: As we already saw in Section 10:

$$|N_p(f_n) - N_p(f)| \leq N_p(f_n - f) \to_{n\to\infty} 0.$$

$(iii) \Rightarrow (ii)$: Since (f_n) converges to f μ - stochastically Exercise 11.16 implies that also $(|f_n|^p)$ converges stochastically to $|f|^p$. Then lemma 13.6 shows that $(|f_n|^p)$ converges to $|f|^p$ in $\mathcal{L}^1$. Then following Theorem 13.5 - with $p = 1$ - this implies uniform integrability. ■

Exercise 13.8 *Show that a family $(f_i, i \in I)$ of measurable, numeric functions is uniformly integrable, if for all $\varepsilon > 0$ there is an integrable $h \geq 0$ such that*

$$\int (|f_i| - h)^+ \, d\mu \leq \varepsilon$$

for all $i \in I$.

Exercise 13.9 *A sequence (f_n) in $\mathcal{L}^p(\mu)$ converges to $f\mu - a.e.$ Show that $f \in \mathcal{L}^p(\mu)$ and the convergence is in $\mathcal{L}^p(\mu)$, if and only if $(|f_n|^p)$ is uniformly integrable.*

14 Product measures and Fubini's theorem

So far we have considered the case of n single measurable space $(\Omega, \mathcal{A})$ with measure μ on it. We will learn in a probability theory class that, if $\mu(\Omega) = 1$, this can describe the outcome of a random experiment. But actually in probability theory one is interested in the situation where such a random experiment is performed a large number of trials independently. How can this be modelled? Already in a first introductory course in probability one learns, that the independence assumption corresponds with multiplying the probabilities. As measures such probabilities are product measures where, if restricted to one of the components of the underlying product space, there measures ought to be the same (since we perform one and the same experiment a large number of times). But how does one define such measure properly, how does one integrate with respect to them? These questions are tackled in this section. The generic example is d -dimensional Lebesgue measure λ^d. As a matter of fact, the generating sets for λ^d are the d -dimensional

intervals. Those are the products of one - dimensional intervals. Their measure is the products of the measures of the one - dimensional intervals.
The general case is treated in very much the same way: One defines the measure first on those subsets of the product space which have a "natural" product measure. Then we extend it to the generated σ - algebra. We need to start with defining products of spaces and σ - algebras.
In this section we will always be given measurable spaces $(\Omega_i, \mathcal{A}_i), i = 1, ...n$. We define

$$\Omega := \prod_{i=1}^{n} \Omega_i = \Omega_1 \times ... \times \Omega_n,$$

and the projection

$$p_i : \Omega \to \Omega_i$$

which maps $(\omega_1, ...\omega_n)$ to ω_i. Define the product σ - algegbra

$$\otimes_{i=1}^{n} \mathcal{A}_i := \mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n := \sigma(p_1, ..., p_n)$$

which is generated by the projections (the smallest σ - algebra on Ω, such that all p_i are measurable). The following way to generate $\mathcal{A}$ will be central

Theorem 14.1 *Let $\mathcal{E}_i$ be a generator of $\mathcal{A}_i$ for $i = 1...n$, such that for each i there are sequences $(E_{ik})_k$ in $\mathcal{E}_i$ with $E_{ik} \uparrow \Omega_i$ as $k \to \infty$. Then $\mathcal{A}_1 \otimes .. \otimes \mathcal{A}_n$ is generated by*

$$\{E_1 \times \cdots \times E_n, \quad E_i \in \mathcal{E}_i\}.$$

Proof. *Let $\mathcal{A}$ be any σ - algebra on Ω. p_i is $\mathcal{A} - \mathcal{A}_i$ - measurable, if and only if each $p_i^{-1}(E_i) \in \mathcal{A}$ for all $E_i \in \mathcal{E}_i$. But since this is true for all i also*

$$E_1 \times ... \times E_n = \bigcap_{i=1}^{n} p_i^{-1}(E_i) \in \mathcal{A}.$$

If, on the other hand $E_1 \times ... \times E_n \in \mathcal{A}$ for all $E_i \in \mathcal{E}_i$ then also

$$F_k := E_{1k} \times E_{2k} \times \cdots \times E_{(i-1)k} \times E_i \times E_{(i+1)k} \times ... \times E_{nk \in} \mathcal{A}$$

for all k. But (F_k) converges to

$$\Omega_1 \times E_i \times \Omega_{i+1} \times .. \times \Omega_k = p_i^{-1}(E_i),$$

hence also $p_i^{-1}(E_i) \in \mathcal{A}$. ■

Exercise 14.2 *Prove that the condition of $E_{ik} \uparrow \Omega_i$ is actually needed in the previous theorem.*

Example 14.3 *Choosing $\Omega_i = \mathbb{R}$ for all i, $\mathcal{A}_i = \mathcal{B}^1$ for all i and $\mathcal{E}_i = \mathcal{J}^1$, then obviously*

$$\{E_1 \times .. \times E_n, E_i \in \mathcal{E}_i\} = \mathcal{J}^d.$$

With Theorem 14.1 this gives $\mathcal{B}^d = \mathcal{B}^1 \times .. \times \mathcal{B}^k$. As we saw in Section 5 λ^n is the unique measure on $\mathcal{B}^n$ with

$$\lambda^n(I_1 \times .. \times I_n) = \lambda^1(I_1)....\lambda^1(I_n)$$

for all $I_i \in \mathcal{J}^1$.

The above example immediately raises the following question: Given measure μ_i on $(\Omega_i, \mathcal{A}_i)$. Under which conditions is there a measure π on $(\Omega, \mathcal{A})$ such that

$$\pi\left(E_1 \times .. \times E_n\right) = \mu(E_1)...\mu\left(E_n\right) \tag{14.1}$$

for all $E_i \in \mathcal{E}_i$ (and the E_i generators of $\mathcal{A}_i$)? When is such a π unique? The second question can be answered at once:

Theorem 14.4 *If each generator $\mathcal{E}_i$ of $\mathcal{A}_i$ is $\bigcap$ - stable and contains a sequence (E_{ik}) with $E_{ik} \uparrow \Omega_i$ and $\mu\left(E_{ik}\right)$, there is at most one π as given in (14.1)*

Proof. Define

$$\mathcal{E} := \{E_1.. \times E_n, E_i \in \mathcal{E}_i\}.$$

Theorem 14.1 shows that $\mathcal{E}$ generators $\mathcal{A}_1 \otimes .. \otimes \mathcal{A}_n$. Since

$$\left(\prod_{i=1}^n E_i\right) \cap \left(\prod_{i=1}^n F_i\right) = \prod_{i=1}^n \left(E_i \cap F_i\right)$$

with $\mathcal{E}_i$ also $\mathcal{E}$ is $\cap$ - stable. Moreover

$$E_k := E_{1k} \times .. \times E_{nk} \uparrow \Omega.$$

The assertion follows thus from Theorem 2.13, since

$$\pi\left(E_k\right) = \mu_1\left(E_{1k}\right)....\mu_n\left(E_{nk}\right) < \infty.$$

■

We will now turn to answering the first question as well, i. e. we will construct the product of two measure spaces $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$. The generalization to arbitrary n is then a fairly standard induction argument. For $Q \subseteq \Omega_1 \times \Omega_2$ and $\omega_i \in \Omega_i, i = 1, 2$, we will first define

$$Q_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in Q\}$$

and

$$Q_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in Q\}$$

Then we obtain

Lemma 14.5 *14.5: Let $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Then for arbitrary $\omega_1 \in \Omega$, and $\omega_2 \in \Omega_2$, $Q_{\omega_1 \in} \mathcal{A}_2$ and $Q_{\omega_1} \in \mathcal{A}_1$.*

Proof. For $Q, Q_1, Q_2, ... \leq \Omega_1 \times \Omega_2$ and arbitrary $\omega_1 \in \Omega$, it holds

$$(\Omega \backslash Q)_{\omega_1} = \Omega_2 \backslash Q_{\omega_1}$$

as well as

$$\left(\bigcup_{i=1}^{\infty} Q_i \right)_{\omega_1} = \bigcup_{i=1}^{\infty} (Q_i)_{\omega_1.}$$

Moreover $\Omega_{\omega_1} = \Omega_2$ and more generally

$$(A_1 \times A_2)_{\omega_1} = 1_{\omega_1}(A_1) \cdot A_2$$

$(A_i \subset \Omega_i)$. Thus for all $\omega_1 \in \Omega$,

$$\mathcal{A}' := \{Q \subset \Omega : Q_{\omega_1} \in \mathcal{A}_2\}$$

is a σ - algebra over the set Ω. $\mathcal{A}'$ contains all $A_1 \times A_2, A_i \in \mathcal{A}_i$. But from Theorem 14.1 we have that $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the smallest such σ - algebra. This proves the lemma for Q_{ω_1}. The proof for Q_{w_2} is analogous. ■

According to the previous lemma we may measure Q_{ω_1} with μ_2 and Q_{ω_2} with μ_1. Moreover we can show

Lemma 14.6 *Assume that μ_1, μ_2 are σ - finite. Then for all $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ the functions*

$$\omega_1 \longmapsto \mu_2(Q_{\omega_1}) \text{ and } \omega_2 \longmapsto \mu_2(Q_{\omega_2})$$

(defined on Ω_1 and Ω_2 respectively) are $\mathcal{A}_1$ - measurable and $\mathcal{A}_2$ - measurable, respectively.

Proof. Define $\varsigma_Q(\omega_1) = \mu_2(Q_{\omega_1})$. First assume $\mu_2(\Omega_2) < \infty$. Define

$$\mathcal{D} := \{D \in \mathcal{A}_1 \otimes \mathcal{A}_2, \varsigma_D \text{ is } \mathcal{A}_1 \text{ - measurable}\}. \tag{14.2}$$

Then $\mathcal{D}$ is a Dynkin - system that contains all sets of the form $A_1 \times A_2, A_i \in \mathcal{A}_\omega i = 1,2$ (Exercise 14.7). The system $\mathcal{E}$ of all sets $A_1 \times A_2, A_i \in \mathcal{A}_i, i = 1,2$ is $\cap$ - stable and generates $\mathcal{A}_1 \otimes \mathcal{A}_2$. Thus $\mathcal{D}(\mathcal{E}) = \mathcal{A}_1 \otimes \mathcal{A}_2$ and because of $\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$ we obtain $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$.

If μ_2 is only σ - finite, there is a sequence (B_n) in $\mathcal{A}_2$ with $B_n \uparrow \Omega_2$ and $\mu_2(B_n) < \infty$. For all n the measure $\mu_{2,n}(A_2) := \mu_2(A_2 \cap B_n)$ is finite. Thus $\omega_1 \longmapsto \mu_{2,n}(Q_{\omega_1})$ is measurable with respect to $\mathcal{A}_1$ for $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$. But

$$\sup_n \mu_{2,n}(Q_{\omega_1}) = \mu_2(Q_{\omega_1}),$$

since μ_2 as a measure is continuous from below. As the supremum of measurable functions is measurable this shows the assertion for $\omega_1 \mapsto \mu_2(\omega_{\omega_1})$. The other assertion is proved analogously. ■

Exercise 14.7 *Show that $\mathcal{D}$ as defined in (14.2) is a Dynkin - system that contains $\mathcal{E}$ (as in Lemma 14.6).*

Now the existence of a product measure follows easily.

Theorem 14.8 *Let $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1,2$ be σ - finite measure spaces. Then there is a unique measure π on $\mathcal{A}_1 \otimes \mathcal{A}_2$ with*

$$\pi(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \tag{14.3}$$

for all $A_i \in \mathcal{A}_i, i = 1,2$. For each $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ it holds

$$\pi(Q) = \int \mu_2(Q_{\omega_1})\mu_1(d_{\omega_1}) = \int \mu_1(Q\omega_2)\,\mu_2(d_{\omega_1}). \tag{14.4}$$

Obviously, if μ_1, μ_2 are σ - finite, so is π.

Proof. Again let $\varsigma_Q(\omega_1) = \mu_2(Q_{\omega_1})$. Define $\pi(Q) := \int \varsigma_Q d\mu_1$. For each sequence (Q_n) of pairwise disjoint sets in $\mathcal{A}_1 \otimes \mathcal{A}_2$ it holds $\varsigma_{\cup Q_n} = \sum \varsigma_{Q_n}$ and thus by monotone convergence

$$\pi\left(\bigcup_{n=1}^{\infty} Q_n\right) = \sum_{n=1}^{\infty} \pi(Q_n).$$

Because of $\varsigma_\emptyset = 0$ also $\pi(\emptyset) = 0$ and thus π is a measure on $\mathcal{A}_1 \otimes \mathcal{A}_2$. π has property (14.3) since

$$\varsigma_{A_1 \times A_2} = \mu_2(A_2) 1_{A_i}$$

and therefore

$$\pi(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2).$$

In the same way we can define a measure on $\mathcal{A}_1 \otimes \mathcal{A}_2$ by

$$\pi^{'}(Q) = \int \mu_1(Q_{\omega_2}) \mu_2(d_{\omega_2})$$

Applying Theorem 14.4 to $\mathcal{E}_1 = \mathcal{A}_1$ and $\mathcal{E}_2 = \mathcal{A}_2$ gives that there is at most one such measure, hence $\pi^{'} = \pi$ and the second equality in (14.4) follows. ■

Definition 14.9 *In the situation of Theorem 14.8 the unique measure π with property (14.3) is called the product measure of μ_1 and μ_2 and demoted by $\mu_1 \otimes \mu_2$.*

The most prominent example of a product measure is, of course, Lebesgue measure, where we have

$$\lambda^2 = \lambda^1 \otimes \lambda^1 \text{ or, more generally, } \lambda^{m+n} = \lambda^m \otimes \lambda^n.$$

We will now turn to integration with respect to product measures. To simplify notation let us write

$$\omega_2 \longmapsto f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2)$$

and

$$\omega_1 \longmapsto f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2)$$

for a function $f : \Omega_1 \times \Omega_2 \to \Omega^{'}$ (for a set $\Omega^{'}$) and $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$. For $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ we obviously have

$$(1_Q)_{\omega_1} = 1_{Q_{\omega_1}} \text{ and } (1_Q)_{\omega_2} = 1_{Q_{\omega_1}}. \tag{14.5}$$

The following is immediate from Lemma 14.5

Exercise 14.10 *For each $\left(\Omega^{'}, \mathcal{A}^{'}\right)$ and each measurable mapping*

$$f : \Omega_1 \times \Omega_2 \to \Omega^{'}$$

the mappings f_{ω_1} and f_{ω_2}, respectively, are $\mathcal{A}_2$ - $\mathcal{A}^{'}$ - measurable and $\mathcal{A}_1$ - $\mathcal{A}^{'}$ - measurable, respectively. Prove this.

Already formula (14.4) gives an idea of how integration with respect to $\mu_1 \otimes \mu_2$ should look like. The following two theorems, that are due to Tonelli and Fubini, respectively generalize (14.4) to $\mu_1 \otimes \mu_2$ - integrable functions.

Theorem 14.11 *(Tonelli)* *Let* $(\Omega_i, \mathcal{A}_i, \mu_i)$ *,* $i = 1, 2$ *two* σ *- finite measurable spaces and*

$$f : \Omega_1 \otimes \Omega_2 \to \bar{\mathbb{R}}^+$$

be $\mathcal{A}_1 \otimes \mathcal{A}_2$ *- measurable. Then*

$$\omega_2 \longmapsto \int f_{\omega_2} d\mu_1 \text{ and } \omega_1 \longmapsto \int f_{\omega_2} d\mu_2$$

are $\mathcal{A}_1$ *- measurable and* $\mathcal{A}_2$ *- measurable, respectively. Eventually:*

$$\begin{aligned} \int f d(\mu_1 \otimes \mu_2) &= \int \left(\int f_{\omega_2} d\mu_1 \right) \mu_2(d_{\omega_2}) \\ &= \int \left(\int f_{\omega_1} d\mu_2 \right) \mu_1(d\omega_1). \end{aligned} \tag{14.6}$$

Proof. : Put $\Omega := \Omega_1 \times \Omega_2, \mathcal{A} := \mathcal{A}_1 \otimes \mathcal{A}_2$, and $\pi := \mu_1 \otimes \mu_2$. Starting with step functions

$$f = \sum_{i=1}^{n} \alpha_i 1_{Q^i} \qquad , \alpha_i \geq 0, Q^i \in \mathcal{A},$$

then $f_{\omega_2} = \sum_{\alpha_i} 1_{Q^i_{\omega_2}}$ and due to (14.5)

$$\int f_{\omega_2} d\mu_1 = \sum_{i=1}^{n} \alpha_i \mu_1(Q^i_{\omega_2}).$$

Hence $\omega_2 \longmapsto \int f_{\omega_2} d\mu$ is $\mathcal{A}_2$ - measurable. Using (14.4) we get

$$\int \left(\int f_{\omega_2} d\mu_1 \right) \mu_2(d_{\omega_2}) = \sum_{i=1}^{n} \alpha_i \pi(Q^i) = \int f d\pi$$

hence the first equality in 814.6).
arbitrary $\mathcal{A}$ - measurable $f \geq 0$ let (u^n) be a sequence of step - functions with $u^n \uparrow f$. Then (using the first part of this proof)$\left(u^n_{\omega_2}\right)$ is a sequence of

step - functions (with respect to $\mathcal{A}_2$) with $u^n_{\omega_2} \uparrow f_{\omega_2}$. From what we have already proved we obtain that

$$\omega_2 \mapsto \varphi^n(\omega_2) := \int u^n_{\omega_2} d\mu_1$$

is increasing with supremum

$$(14.7) \qquad \omega_2 \longmapsto \int f\omega_2 d\mu_1.$$

Hence also the function defined in (14.7) is measurable and hence by monotone convergence

$$\begin{aligned} \int \left(\int f_{\omega_2} d\mu_2 \right) \mu_2(d\omega_2) &= \sup_n \int \varphi^n d\mu_2 \\ &= \sup_n \int u^n d\pi, \end{aligned}$$

where for the second equality we used the first step of this proof. By the choice of (u^n) we thus have

$$\int \varphi^n d\mu_2 \uparrow \int f d\pi$$

which gives

$$\int (\int f\omega_2 d\mu_1) \mu_2(d\omega_2) = \int f d\bar{\pi}.$$

Analogous considerations conclude the proof of the theorem. ■

Theorem 14.12 *(Fubini) Let $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1,2$ be two σ - finite measure spaces and*

$$f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$$

be a measurable, integrable (with respect to $\mu_1 \otimes \mu_2$) function. Then f_{ω_1} is μ_2 - integrable $\mu_1 -$ a.e. in ω_1 and $f\omega_2$ is μ_1 - integrable $\mu_2 -$ a.e. Hence the following function are defined a.e.:

$$\omega_1 \mapsto \int f\omega_1 d\mu_2 \text{ and } \omega_2 \mapsto \int f\omega_2 d\mu_1.$$

Both these functions are integrable (with respect to μ_1 and μ_2, respectively) and (14.6) holds.

Proof. : Obviously

$$|f_{\omega_i}| = |f|_{\omega_i}, (f_{\omega_i}^+) = (f^+)_{\omega_i}, (f_{\omega_i}^-) = (f^-)_{\omega_i}.$$

Applying (14.6) to $|f|$ and $\mu_1 \otimes \mu_2$ we obtain

$$\begin{aligned}\int \left(\int |f_{\omega_1}| \, d\mu_2 \right) d\mu_1 &= \int \left(\int |f_{\omega_2}| \, d\mu_1 \right) d\mu_2 \\ &= \int |f| \, d\mu_1 \otimes \mu_2 < \infty.\end{aligned}$$

Thus $\omega_1 \mapsto \int |f_{\omega_1}| \, d\mu_2$ is finite $\mu_1 - a.e.$, thus $\mu_1 - a.e.$ also μ_2 - integrable. Therefore

$$\omega_1 \mapsto \int f_{\omega_1} d\mu_2 = \int f_{\omega_1}^+ d\mu_2 - \int f_{\omega_i}^- d\mu_2$$

is $\mu_1 - a.e.$ defined and $\mathcal{A}_1$ - measurable. Applying Theorem 14.11 to f^+ and f^- we arrive at:

$$\begin{aligned}\int \left(\int f_{\omega_1} d\mu_2 \right) d\mu_1 &= \int \left(\int f_{\omega_1}^+ d\mu_2 \right) d\mu_1 - \int \left(\int f_{\omega_1}^- d\mu_2 \right) d\mu_1 \\ &= \int f^+ d\pi - \int f^- d\pi = \int f d\pi.\end{aligned}$$

Interchanging the roles of ω_1 and ω_2 concludes the proof. ■

The generalizations to an arbitrary (but finite) number of factors will now be left to the reader. The proofs work by induction.

Exercise 14.13 *Let $\mu_1, ..., \mu_n$ be σ - finite measures on $(\Omega_1, \mathcal{A}_1), ..., (\Omega_n, \mathcal{A}_n)$. Then there is a unique measure π on $\mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n$ with*

$$\pi(A_1 \times .. \times A_n) = \mu_1(A_1)...\mu(A_n)$$

for all $A_i \in \mathcal{A}_i$. Prove this.

Definition 14.14 *The measure π in Exercise 14.13 is called the product measure of $\mu_1, ..., \mu_n$ and denoted by*

$$\otimes_{i=1}^n \mu_i = \mu, \otimes ... \otimes \mu_n.$$

In very much the same way one extends Fubini's theorem to the case of several factors.

Exercise 14.15 *In the situation of Exercise 14.13 let* $f : \Omega, \times .. \Omega_n \to \bar{\mathbb{R}}$ *be a* $\otimes_{i=1}^{n} \mu_i$ *- integrable function. Then for any permutation* $i_1, ..., i_n$ *of the indices* $1, ...n$*:*

$$\int f d\left(\mu_1 \otimes .. \otimes \mu_n\right) = \int (.. \left(f\left(\omega_1 ..., \omega_n\right) \mu_{i_1} d_{\omega_i}\right) \mu_{i_n} d_{\omega_i}$$

Prove this.

Exercise 14.16 *Let* $\mathcal{B}_r(x_0)$ *be the closed ball of radius* r *centered in* x_0 *in* $\mathbb{R}^d$*. Put*

$$\alpha_d := \lambda^d(k, (0)).$$